

Molekulargenetische Faktoren der Suszeptibilität für Karotis-Plaques

Dissertation
zur Erlangung des akademischen Grades
Dr. rer. nat.

an der Medizinischen Fakultät
der Universität Leipzig

eingereicht von:
M. Sc. Janne Pott

Geburtsdatum / Geburtsort:
22.04.1990 in Erlangen

angefertigt am:
Universität Leipzig, Institut für Medizinische Informatik, Statistik und Epidemiologie

Betreuer:
Prof. Dr. Markus Scholz

Beschluss über die Verleihung des Doktorgrades vom: 12.02.2019

Inhaltsverzeichnis

1	Einführung	1
1.1	Biologischer Hintergrund.....	2
1.1.1	Pathogenese der Atherosklerose	2
1.1.2	PCSK9 als wichtiger Einflussfaktor auf den Lipidstoffwechsel	4
1.1.3	Grundlagen der Genetik	6
1.2	Fragestellung der Arbeit.....	8
1.3	Material und Methoden	9
1.3.1	Beschreibung der LIFE-Studien.....	9
1.3.2	Betrachtete Phänotypen der Atherosklerose.....	10
1.3.3	Grundlegende Methoden der Genotypisierung.....	13
1.3.4	Grundlagen der genomweiten Assoziationsanalyse	15
1.3.5	Enrichment Analyse.....	17
1.3.6	Mendelsche Randomisierung (MR).....	19
2	Ergebnisse.....	22
2.1	„Genome-wide analysis identifies novel loci of plaque burden in carotid artery“	22
2.2	„Genetic regulation of PCSK9 plasma levels and its impact on atherosclerotic vascular disease phenotypes“	33
2.3	Zusätzliche Ergebnisse.....	45
2.3.1	Vergleich LIFE-Heart – LIFE-Adult.....	45
2.3.2	Genetische Analyse des Plaque-Score eingeschränkt auf Fälle	50
2.3.3	Haplotypen-Analyse des <i>PCSK9</i> Locus.....	53
3	Schlussfolgerungen und Ausblick	56
4	Zusammenfassung der Arbeit	59
5	Literaturverzeichnis.....	63
6	Anlagen.....	67
	Darstellung des eigenen Beitrags.....	71
	Selbstständigkeitserklärung	75
	Lebenslauf	77
	Wissenschaftliche Erfahrung.....	79
	Danksagung	81

1 Einführung

*„Krankheiten befallen uns nicht wie aus heiterem Himmel,
sondern entwickeln sich aus täglichen Sünden wider die Natur.
Wenn sich diese gehäuft haben, brechen sie unversehens hervor.“*
Hippokrates [1]

Hippokrates ging davon aus, dass Krankheiten durch ein Ungleichgewicht von Körpersäften entstehen. Diese Ansicht mag überholt sein, allerdings kann man die Aussage des Zitats leicht auf die moderne Medizin ummünzen: die „täglichen Sünden“ entsprechen bekannten beeinflussbaren Risikofaktoren einer Krankheit. Die Akkumulation solcher Faktoren erhöht das Risiko für die Krankheitsentstehung.

Eine weit verbreitete Krankheit mit vielen Risikofaktoren ist die Atherosklerose, eine progressive Erkrankung der großen Arterien. Hier sind sowohl beeinflussbare als auch nicht beeinflussbare Risikofaktoren bekannt [2]. Zu den beeinflussbaren Faktoren zählen unter anderem Störung des Lipidstoffwechsels, Übergewicht, Tabakrauchen, arterielle Hypertonie, Diabetes mellitus und Bewegungsmangel. Diese können mittels Medikamentengabe oder Änderungen im Lebensstil positiv beeinflusst werden. Dem gegenüber stehen die nicht zu beeinflussenden Faktoren, wie zum Beispiel Alter und genetische Prädisposition.

Eines der bekanntesten Beispiele der klinischen Form der Atherosklerose ist die koronare Herzerkrankung (KHK), bei der eine Verengung der großen Herzgefäße vorliegt. Zur Detektion relevanter genetischer Faktoren wurden genetische Assoziationsstudien (GWAS) bzw. genetische Meta-Studien durchgeführt. Insgesamt wurden bislang 58 genetische Loci entdeckt, die einen signifikanten Einfluss auf das KHK-Risiko haben (Nikpay et al. [3], $n \sim 185,000$). Während die Funktionsbestimmung vieler Loci noch aussteht, sind einige klar im Lipidmetabolismus verankert, z. B. *APOB*, *APOE-APOC1*, *LDLR*, und *PCSK9*.

Neben der klinischen gibt es auch subklinische Formen der Atherosklerose. Zu diesen Formen zählen Gefäßveränderungen an der Karotis, zum Beispiel Karotis-Plaques oder eine Änderung der Intima-Media-Dicke in der Karotis (cIMT). Diese subklinische Form gilt als Vorläufer eines Gefäßverschlusses in der Karotis und stellt einen weiteren unabhängigen Risikofaktor der KHK dar [4]. Während bereits viele GWAS zur KHK durchgeführt wurden, gibt es nur wenige genetische Studien zu Karotis-Plaques oder cIMT. Bis et al. [5] führte eine Meta-GWAS zu cIMT durch ($n=31,211$), wobei nur drei genomweit signifikante ($p < 5 \times 10^{-8}$) Loci gefunden. Zur Prävalenz von Karotis-Plaques konnten zwei Loci detektiert werden. Neben dem Powerproblem durch kleinere Fallzahlen ist eine weitere mögliche Ursache für diese relativ kleine Zahl an Loci die uneinheitlichen Messmethoden von cIMT und Karotis-Plaques in den verwendeten Studien. Das erschwert z.B. die Detektion von genetischen Varianten mit niedriger Allelfrequenz.

Dies führt zu der Frage, ob man in einer großen Studie mit einheitlichen Messungen weitere Loci der Atherosklerose in Karotisgefäßen detektieren kann und ob man mit einem Phänotyp, der statt der Prävalenz die Anzahl an Plaques darstellt, bessere Ergebnisse erzielt. Da Karotis-Plaques eine subklinische Form der Atherosklerose darstellen, ist auch von Interesse, ob die bereits identifizierten KHK-Loci bei dem subklinischen Phänotyp angereichert sind. Relevant ist ebenfalls die Bestimmung der Genfunktion, bzw. die Analyse von kausalen Beziehungen zwischen den Genprodukten und Karotis-Plaques.

Daher beschäftigt sich der erste Teil dieser kumulativen Dissertation mit einer kurzen Einführung in den biologischen Hintergrund. Anschließend werden die Rationale der Arbeit erläutert und die verwendeten Studien und Methoden erklärt. Die Ergebnisse meiner Dissertation sind im zweiten Teil zusammengefasst und beinhalten drei wesentliche Aspekte: als erstes wird meine Publikation zur Detektion weiterer genetischer Einflussfaktoren auf die Anzahl von Karotis-Plaques und deren Überschneidung mit KHK-Loci vorgestellt (s. Kapitel 2.1). Darauf folgt die Zusammenfassung meiner zweiten Publikation, die sich mit der Genetik von *PCSK9*, einem Risiko-Locus der KHK, befasst, und den kausalen Einfluss des Proteins PCSK9 auf das Risiko einer Gefäßerkrankung analysiert (s. Kapitel 2.2). Abschließend werden einige weitere, bisher unveröffentlichte Ergebnisse beschrieben (s. Kapitel 2.3). Dies umfasst einen Vergleich des genetischen Risikos der verwendeten Studien, eine genetische Analyse der Karotis-Plaques ohne Kontrollen und eine Haplotypen-Analyse des *PCSK9* Locus. Im dritten Kapitel werden die Ergebnisse kurz in Zusammenhang gesetzt und ein Ausblick zu weiteren geplanten Analysen gegeben. Zum Abschluss sind in Kapitel 4 die wichtigsten Ergebnisse dieser Promotionsarbeit zusammengefasst.

1.1 Biologischer Hintergrund

Im folgenden Kapitel wird eine kurze Einführung in die drei wichtigsten biologischen Aspekte dieser Arbeit gegeben. Zuerst wird die Entstehung von atherosklerotischen Plaques beschrieben, anschließend die Rolle des Proteins PCSK9 im Lipidmetabolismus, dessen Gen-Locus mit KHK assoziiert ist. Abschließend werden Punktmutationen im Genom vorgestellt.

1.1.1 Pathogenese der Atherosklerose

Die Atherosklerose, im englischen auch als *arteriosclerotic vascular disease* (ASVD) bekannt, ist eine progressive Erkrankung der großen Arterien, die durch die Einlagerung von Cholesterin und anderen Fetten in die innere Wandschicht arterieller Blutgefäße (Tunica Intima) charakterisiert ist [2].

Besonders anfällig für ASVD sind Krümmungen oder Bifurkationen der Arterien. Aufgrund der anatomischen Gefäßveränderung ändert sich der hämodynamische Scherstress, der auf die Endothelzellen einwirkt [6]. Dies setzt die Permeabilität der Endothelzellen gegen Lipidmoleküle wie Low-density Lipoproteine (LDL, deutsch: Lipoproteine niedrigerer Dichte)

herab und es kommt zur initialen Läsion der Tunica Intima [7]. In den nächsten Schritten verdickt sich die Gefäßwand durch Lipideinlagerung und Bildung von Schaumzellen, bis schließlich ein Plaque mit lipidreichem Kern entsteht (s. Abbildung 1). Während die ersten Stufen der Atherosklerose klinisch unauffällig verlaufen, wird die Erkrankung durch die Plaquebildung zunehmend klinisch wahrnehmbar. Aufgrund der Verengung der Gefäße kommt es zu einer Stenose, die eine Mangelversorgung verursacht. Koronare Herzerkrankung und Ischämie sind Beispiele für die klinische Manifestation dieser Mangelversorgung [2]. Im Falle einer Plaqueruptur kommt es zu einer Gerinnungskaskade, ausgelöst durch das freiwerdende Thromboplastin des Plaques [7]. Es bildet sich ein Thrombus, der die Arterie weiter verengt oder aber vollständig verschließt. Im Falle betroffener Herzgefäße würde dies zu einem akuten Myokardinfarkt führen, wohingegen bei Karotis-Gefäßen ein Schlaganfall das Resultat wäre.

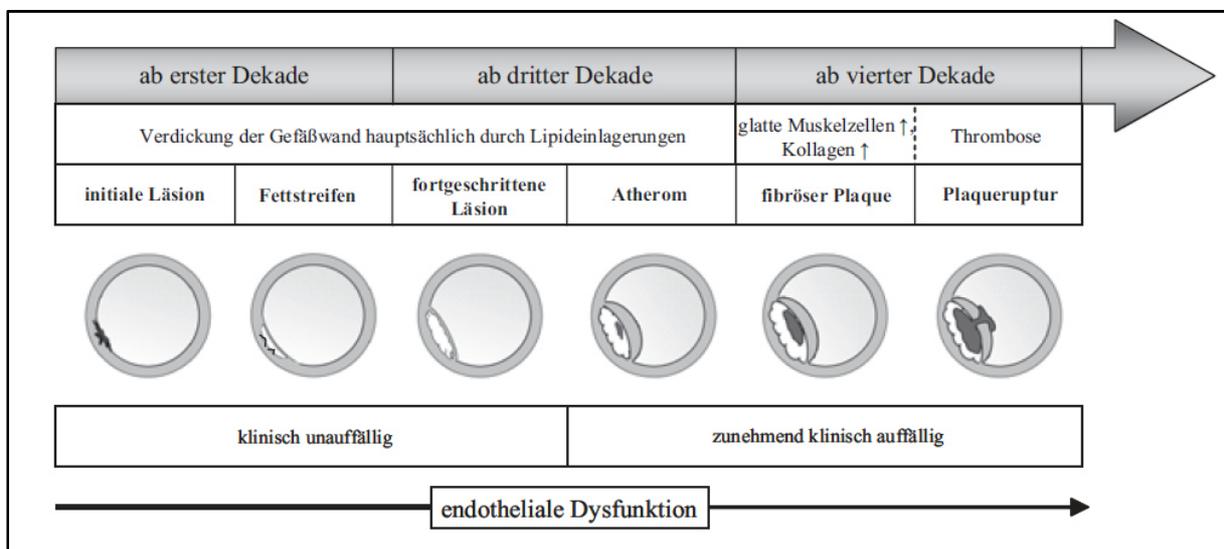


Abbildung 1: Entwicklungsstufen der Atherosklerose (aus Kuhlmann et al. [2]). Im Verlauf einer Gefäßerkrankung wächst eine initiale Läsion der Gefäßwand zu einem fibrösen Plaque, der das Gefäß verengt. Dieser Plaque kann sich lösen und einen Gefäßverschluss verursachen.

Es gibt eine Vielzahl von Risikofaktoren, die die Entstehung und Entwicklung von Atherosklerose begünstigen [2]. Zu den beeinflussbaren Risikofaktoren zählen der Cholesterinspiegel (Gesamt-Cholesterin, LDL-Cholesterin [LDL-C], High-density Lipoprotein-Cholesterin [HDL-C]), arterieller Bluthochdruck, Übergewicht, geringe körperliche Aktivität, Rauchen und Diabetes mellitus. All diese Faktoren sind mittels geeigneter Intervention (Diät, Sport, Medikation o.ä.) so einstellbar, dass das zusätzliche Risiko minimiert wird. Des Weiteren gibt es nicht beeinflussbare Risikofaktoren, wie Alter, Geschlecht und genetische Prädisposition. Bislang wurden etwa 60 genetische Risikoloci der KHK beschrieben [8]. Das genetische Risiko ist zwar nicht änderbar, aber in Kombination mit einem gesunden Lebensstil lässt sich das Gesamtrisiko eines kardiovaskulären Ereignisses minimieren. Khera et al. [9] konnten mit Hilfe eines genetischen Risikoscores (GRS) zeigen, dass das Inzidenzrisiko für Personen mit niedrigem GRS und ungesundem Lebensstil vergleichbar ist mit dem Risiko von Personen mit hohem GRS und sehr gesundem Lebensstil.

1.1.2 PCSK9 als wichtiger Einflussfaktor auf den Lipidstoffwechsel

Da sich hauptsächlich LDL-Moleküle in die Gefäßwand einlagern und somit den oben beschriebenen Prozess in Gang setzen, stellen hohe LDL-C Werte einen unabhängigen, aber modifizierbaren Risikofaktor für ASVD dar. Die Senkung von LDL-C im Plasma ist daher von hohem therapeutischen Interesse.

Die LDL-C Aufnahme vom Plasma in die Zelle wird durch den in der Zellmembran verankerten LDL-Rezeptor (LDL-R) vermittelt (eine Übersicht ist in Abb. 2 gegeben). Dieser bindet spezifisch an ApoB-100, was den Protein-Anteil von LDL darstellt. Nach der Bindung wird der LDL-C – LDL-R Komplex mittels Endozytose in die Zelle eingeschleust. Durch einen Abfall des pH-Werts im Endosom löst sich der Komplex wieder auf. Das LDL-C haltige Endosom-Innere fusioniert mit Lysosomen, während der LDL-Rezeptor wieder zurück an die Zellmembran transportiert wird [10].

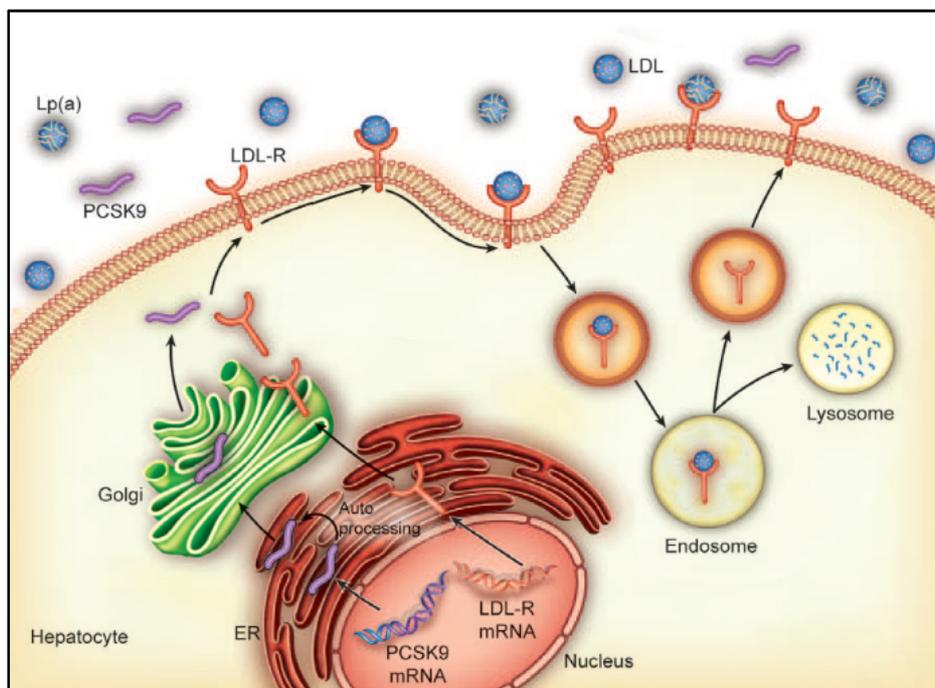


Abbildung 2: Aufnahme von LDL-Cholesterol (LDL) in eine Leberzelle (modifiziert aus Bergeron et al. [12]: die Abbildungen 1 und 2 wurden hier zusammengefasst). LDL bindet an den LDL-Rezeptor (LDL-R) und wird in die Zelle aufgenommen. Da im Endosom der pH-Wert sinkt, löst sich LDL vom Rezeptor. Dieser wird zurück an die Zellmembran geführt, während das LDL für den Stoffwechsel der Zelle genutzt werden kann.

Die Medikation mit Statinen ist das Mittel erster Wahl zur Senkung des LDL-C-Spiegels [11]. Zusätzlich wirken Statine primär- und sekundärprophylaktisch, sie senken also das Risiko von kardiovaskulären Ereignissen [12]. Die Wirkung von Statin beruht auf der kompetitiven Inhibition der HMG-CoA Reduktase, wodurch ein Abfall des intrazellulären Cholesterins verursacht wird. In Folge dessen wird die Expression von LDL-R verstärkt, was zu einer höheren Aufnahme von LDL-C in die Zellen und somit zu einer LDL-C Reduktion im Plasma führt [13]. Die maximale LDL-C Reduktion durch Statine liegt zwischen 25 und 60% (in Abhängigkeit von dem pharmazeutischen Derivat und der Dosis). Ein Grund für diese Limitation ist die Gegenregulation der Zellen. Die Aktivierung der LDL-R Expression erfolgt durch den

Transkriptionsfaktor SREBP2 (sterol regulatory element binding protein-2). Dieser Faktor aktiviert jedoch auch PCSK9 (Proproteinkonvertase Subtilisin/Kexin Typ 9), welches an LDL-R binden kann. Die LDL-C-Bindung und die Endozytose werden dadurch nicht beeinträchtigt, aber im Endosom kommt es nicht mehr zur Trennung von LDL-R, LDL-C und PCSK9. Daher werden alle Proteine im Lysosom abgebaut. PCSK9 reduziert also die Anzahl verfügbarer LDL-Rezeptoren, wodurch die LDL-C Aufnahme gehemmt ist [13] (s. Abb. 3).

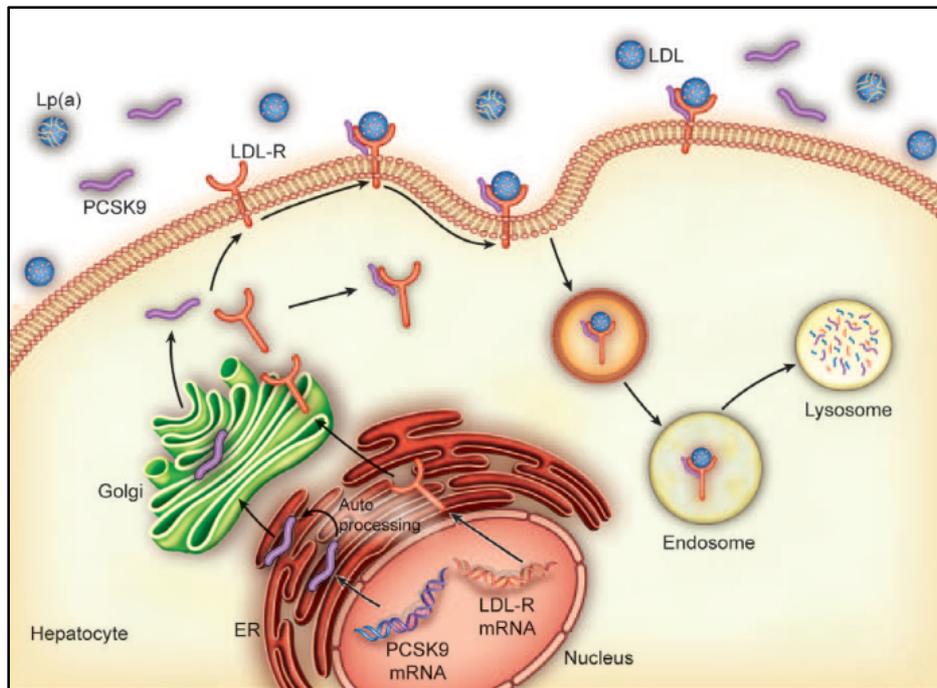


Abbildung 3: Aufnahme von LDL-Cholesterol (LDL) in eine Leberzelle, die durch PCSK9 beeinflusst wird (aus Bergeron et al. [12]). Im Unterschied zu Abb. 2 bindet hier PCSK9 an den LDL-Rezeptor (LDL-R). Der Komplex aus LDL, LDL-R und PCSK9 wird gemeinsam in die Zelle aufgenommen. Allerdings löst sich hier nicht mehr das LDL vom Rezeptor, und der Rezeptor wird zusammen mit dem LDL lysiert. Daher führt PCSK9 zu einer Reduktion von LDL-R in Leberzellen und damit zu einer verminderten LDL-Aufnahme aus dem Blut.

PCSK9 ist damit ein wichtiger Regulator des Lipidmetabolismus und ebenfalls von therapeutischem Interesse. Hier gibt es mehrere Ansatzpunkte für PCSK9-Inhibitoren [13], zum Beispiel *small interfering RNA*, die die Degradation von PCSK9 mRNA unterstützen, oder kleine Peptide, die durch strukturelle Ähnlichkeit mit dem LDL-Rezeptor an PCSK9 binden können und damit eine kompetitive Hemmung bewirken. Bereits in klinischen Studien erfolgreich waren zwei monoklonale Antikörper, Alirocumab [14] und Evolocumab [15], die an freies PCSK9 im Plasma binden und dadurch eine Bindung von PCSK9 an LDL-R verhindern. Im Fall von Evolocumab konnte nicht nur eine Reduzierung von LDL-C gezeigt werden, sondern auch eine signifikante Reduktion kardiovaskulärer Ereignisse [15].

1.1.3 Grundlagen der Genetik

Die genetische Information ist in Form der DNA-Sequenz gespeichert. Ein DNA-Strang besteht aus einem Rückgrat aus Phosphatdesoxyribosen und den vier Basen Adenin (A), Guanin (G), Thymin (T) und Cytosin (C). Im Normalfall liegt die DNA als Doppelstrang vor, das heißt, sie ist verbunden mit ihrem Gegenstrang in der komplementären Basenreihenfolge. Hierbei bindet A über Wasserstoffbrücken an T bzw. C an G.

Die kodierenden Bereiche der DNA werden mittels Transkription in mRNA umgeschrieben; nach der Translation und möglichen Modifikationen entstehen fertige Proteine. Das menschliche Genom besteht aus einem diploiden Chromosomensatz mit 23 Chromosomen (22 autosomale + X), die insgesamt etwa $3,27 \times 10^9$ Basenpaaren (bp) umfassen.

Mutationen sind spontan auftretende, aber dauerhaft bleibende Veränderungen des Erbguts [16]. Diese Arbeit befasst sich ausschließlich mit Punktmutationen, daher werden Chromosom- oder Genommutationen hier nicht behandelt. Eine Punktmutation betrifft eine einzelne Base in der DNA, die entweder durch eine andere ausgetauscht wird (Bsp: A \rightarrow C), durch Deletion entfernt wird (A \rightarrow -), oder durch Insertion neu dazu kommt (- \rightarrow A; Indel = Insertion oder Deletion). Als SNPs (single nucleotide polymorphism, Einzelnukleotid-Polymorphismus) werden die „erfolgreichen“ Punktmutationen bezeichnet, die sich in einer Population zu einem gewissen Grad durchgesetzt haben, beispielsweise wenn das seltene Allel häufiger als 1% vorhanden ist (MAF, minor allele frequency). Hier sind daher an einer Position zwei (oder mehr) Basen möglich (s. Abb. 4). SNPs können überall im Genom auftreten und repräsentieren einen großen Teil der individuellen genetischen Variation innerhalb einer Population. Der Genotyp eines SNPs ist in diesem Zusammenhang die individuelle Kombination der beiden Allele.

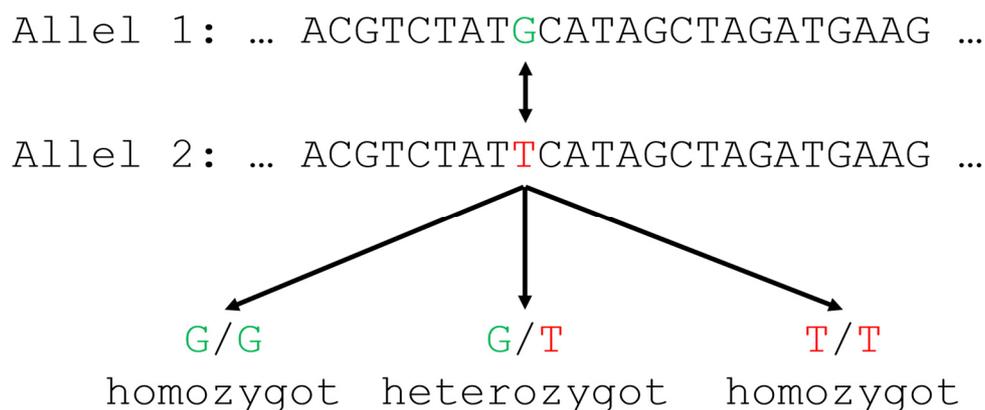


Abbildung 4: Ein Beispiel für einen G/T-SNP. An der neunten Basenposition dieses Beispiels kann entweder die Base G (grün) oder T (rot) eingebaut sein. Je nach Kombination der Allele ist der resultierende Genotyp hier homozygot GG, heterozygot GT oder homozygot TT.

Die Folgen eines SNPs können sehr unterschiedlich sein, je nach Lokalisation und Funktion. Beispielsweise kann ein SNP im nicht-kodierenden Bereich die Bindungsstärke eines Transkriptionsfaktors beeinflussen und somit einen Effekt auf die Expressionsrate eines Gens

ausüben, der sich in der Konzentration des Proteins in der Zelle oder im Blut widerspiegelt. Weitere Beispiele sind SNPs im kodierenden Bereich eines Genes. Hier kann man grob unterscheiden zwischen synonymen Mutationen, deren Basenaustausch keine Konsequenzen auf die Aminosäuren hat, Missense-Mutationen, deren Basenaustausch die Aminosäure ändert, und Nonsense-Mutationen, die zu einem verfrühten Stop Codon führen. Missense- und Nonsense-Mutationen können Auswirkungen auf die Funktionalität des Proteins haben. Ein Beispiel dafür ist die Blutgruppe 0, die durch eine Nonsense-Mutation in einem Glycosyltransferase-Gen entsteht.

Ob SNPs mit Krankheitsphänotypen assoziiert sind, kann mittels genomweiter Assoziationsanalysen (GWAS) getestet werden (s. auch Kapitel 1.3.4). Ein Allel wird dazu als Referenz festgelegt und auf 0 kodiert, das andere bezeichnet man als Effekt-Allel und wird auf 1 codiert. Im additiven Modell nimmt man an, dass sich die Wirkung pro Effekt-Allel addiert. Daher kann die Gendosis als Summe beider Allele berechnet werden. Im Beispiel von Abbildung 4 hätte also eine Person mit Genotyp GG eine Gendosis von 0, während eine Person mit Genotyp TT eine Dosis von 2 hätte (G als Referenz-Allel, T als Effekt-Allel).

Es ist zu beachten, dass das Genom eine Haploblock-Struktur aufweist. Ein Haploblock ist definiert als Kombination von SNPs, die in paarweise hohem Kopplungsungleichgewicht (Linkage-disequilibrium, LD) auf einem Chromosom liegen und typischerweise gemeinsam vererbt werden [16] (s. Abb. 5). Das heißt, SNPs sind nicht per se stochastisch unabhängig voneinander. Benachbarte SNPs werden häufiger zusammen vererbt und sind daher mitunter stark korreliert. Ein wichtiges Maß des LDs ist r^2 . Es hängt von den Allelfrequenzen ab und ist definiert als das Quadrat des Pearson-Korrelationskoeffizienten r . Eine Konsequenz daraus ist, dass man in einer GWAS nicht zwingend eine kausale Variante detektiert, sondern einen SNP in hohem LD mit der kausalen Variante (indirekte Assoziation).

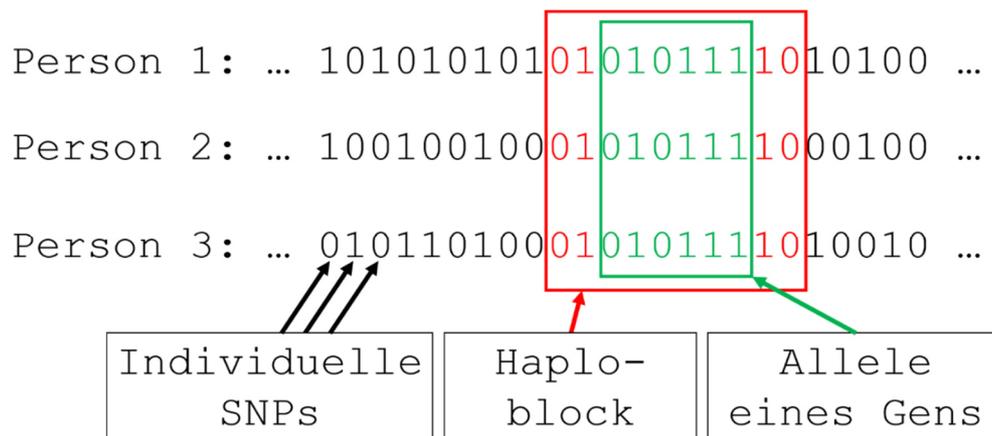


Abbildung 5: Ein Beispiel für die Haploblock-Struktur des Genoms. Von drei Personen wurde der Haplotyp eines Allels bestimmt. Der Haplotyp im Bereich eines Gens für alle drei Personen ist in diesem Beispiel gleich (roter Kasten). Ursache hierfür ist die hohe Konservierung von funktionellen Bereichen wie die eines wichtigen Gens, d.h. diese Bereiche werden en bloc vererbt. Innerhalb des Haploblocks ist das paarweise LD sehr hoch; das LD mit SNPs außerhalb des Blocks ist tendenziell niedriger.

1.2 Fragestellung der Arbeit

Die koronare Herzkrankheit ist immer noch die häufigste Todesursache in den Industrienationen (WHO 2015, [17]). Subklinische Formen der Atherosklerose, wie Karotis-Plaques, gelten als Risikofaktoren für Schlaganfall und KHK. Eine genetische Prädisposition für atherosklerotische Erkrankungen stellt ein von anderen Faktoren unabhängiges Risiko dar. Während der genetische Einfluss auf KHK bereits weitgehend in Meta-Studien beschrieben wurde, ist der Einfluss der Genetik auf die Atherosklerose der Karotis noch nicht abschließend geklärt. Das Ziel dieser Arbeit ist daher die Klärung der Frage, in wie weit Genetik einen Einfluss auf die Anfälligkeit für Karotis-Plaques hat. Um dieses Ziel zu erreichen, werden die folgenden drei Fragestellungen bearbeitet.

Zuerst muss die Frage beantwortet werden, ob und welche genetische Varianten mit Karotis-Plaques assoziiert sind. Hier ist es von Interesse, nicht nur die reine Prävalenz zu testen, sondern auch die Anzahl an Plaques zu betrachten, mit der Fragestellung, ob es genetische Unterschiede bei Patienten mit nur einem Karotis-Plaque gegenüber denen mit mehreren Plaques gibt. Um diese Fragen ausführlich zu beantworten, wurden genomweite Assoziationsstudien zu Karotis-Plaques sowohl mit als auch ohne Kontrollen durchgeführt. Die Analyse ohne Kontrollen soll prüfen, ob der Schweregrad zusätzliche Varianten aufweist, die bei der Analyse der Prävalenz nicht detektiert werden können.

Karotis-Plaques sind nur eine Ausprägung der Atherosklerose. Daher ist es naheliegend, einen gemeinsamen Pathomechanismus für Karotis-Plaques und anderen Formen der Atherosklerose, beispielsweise KHK, anzunehmen. Somit geht eine zweite Analyse der Frage nach, wie sich bekannte KHK-Loci zu Karotis-Plaques verhalten. Diese Fragestellung umfasst nicht nur eine Prüfung der signifikanten Loci und einer damit verbundenen Enrichment-Analyse, sondern auch die Erstellung eines genetischen Risiko-Scores aus KHK-Loci und anschließende Auswertung des Scores bezüglich Karotis-Plaques.

Ein interessanter KHK-Locus ist das Gen *PCSK9*, dessen Genprodukt einen Einfluss auf den Lipidmetabolismus hat. Eine Lipidstörung ist ebenfalls ein Risikofaktor für Atherosklerose in der Karotis. Es ist daher plausibel anzunehmen, dass dieser Locus auch mit Karotis-Veränderungen assoziiert ist, jedoch die Power der Analysen bislang nicht ausreichten, um diesen Effekt zu detektieren. Anhand dieses Locus soll exemplarisch die Frage geklärt werden, wie die genetischen Risiko-Loci auf Karotis-Plaques wirken und ob es kausale Effekte von *PCSK9* auf subklinische Formen der Atherosklerose gibt. Um diese Frage zu beantworten, wird zunächst eine GWAS zu *PCSK9*-Plasmaspiegeln durchgeführt. Die dort detektierten Varianten werden anschließend für kausale Analysen mittels Mendelscher Randomisierung genutzt.

1.3 Material und Methoden

Im Folgenden werden kurz die beiden LIFE Studien vorgestellt, die die Datengrundlage für die weiteren Analysen bilden, und die betrachteten Phänotypen der Atherosklerose definiert. Anschließend werden die wichtigsten Methoden umrissen, die in dieser Arbeit verwendet wurden. Das umfasst sowohl die Grundlagen (Genotyp-Calling, Imputation und GWAS-Pipeline) als auch speziellere Methoden wie die Enrichment Analyse und die Mendelsche Randomisierung.

1.3.1 Beschreibung der LIFE-Studien

Das Leipziger Forschungszentrum für Zivilisationserkrankungen (LIFE, Leipzig Interdisciplinary research cluster of genetic Factors, clinical phenotypes and Environment) umfasst unter anderem zwei große Erwachsenenkohorten: LIFE-Adult und LIFE-Heart.

LIFE-Adult ist eine Querschnittsstudie der Bevölkerung Leipzigs [18]. Etwa 10,000 Teilnehmer wurden zwischen 2011 und 2014 eingeladen und umfangreich untersucht. Dabei wurden viele verschiedene Parameter erhoben, die Rückschlüsse auf die allgemeine Gesundheit, sowie auf spezifische Zivilisationskrankheiten wie etwa Bluthochdruck, Übergewicht, Diabetes und Herz- und Gefäßkrankheiten zulassen. Ein Fokus lag hierbei auf der subklinischen Atherosklerose, daher wurde die Karotis aller Teilnehmer mittels Ultraschall untersucht, die cIMT vermessen und Plaques in den verschiedenen Segmenten der Karotis notiert (s. auch Kap. 1.3.2). Zusätzlich wurde der Knöchel-Arm-Index (ABI) mittels Vicorder bestimmt, um den Gefäßstatus in den peripheren Gefäßen zu erfassen. Bei etwa der Hälfte der Probanden (n=4985) liegen genetische Daten vor, die mittels genomweiter SNP-Arrays bestimmt wurden. Ein Follow-Up der Studie ist in Planung.

Im Rahmen der LIFE-Leipzig Heart Studie (LIFE-Heart) wurden etwa 7000 Patienten des Leipziger Herzzentrums im Zeitraum von 2006 bis 2014 rekrutiert [19]. Einschlusskriterien waren Verdacht auf koronare Herzerkrankung, manifeste Herzerkrankung oder Myokardinfarkt (MI). Hier erfolgte eine tiefgehende Untersuchung zum Gefäßstatus durch Koronarangiographie. Im Verdachtskollektiv (n=3500) wurde der Status der Karotis-Gefäße analog zu LIFE-Adult mittels Karotis-Ultraschall erfasst. Der ABI wurde mittels Doppler Ultraschall bestimmt. Auch hier wurden genomweite SNP-Arrays zur genetischen Charakterisierung verwendet (n=5700). Zusätzlich zu den Labormessungen von Lipidparametern wurden die PCSK9-Plasmaspiegel für Patienten mit Verdacht auf KHK (n=3500) bestimmt.

Beide Studien entsprechen den ethischen Standards der Deklaration von Helsinki und wurden von der Ethikkommission der Medizinischen Fakultät der Universität Leipzig genehmigt (Registrierungsnummern LIFE-Adult: 263-2009-14122009; LIFE-Heart: 276-2005). Zusätzlich ist LIFE-Heart registriert bei ClinicalTrials.gov (Registrierungsnummer: NCT00497887). Alle Teilnehmer haben ihr schriftliches Einverständnis zu genetischen Analysen gegeben.

1.3.2 Betrachtete Phänotypen der Atherosklerose

In dieser Arbeit wird die ASVD in drei verschiedenen Lokalisationen betrachtet: in den Herzgefäßen, in der Karotis und in den Beingefäßen. Da der Schwerpunkt dieser Arbeit auf den Karotis-Arterien liegt, wird auf die Definition der Karotis-Phänotypen vertieft eingegangen. Die Phänotyp-Definitionen der Herz- und Beingefäße werden nur kurz dargestellt.

Atherosklerose in der Karotis

Der Atherosklerose-Status der Karotisgefäße wurde in beiden Studien mittels Ultraschall gemäß einer SOP (standard operating procedure) erfasst [20]. Hierbei wurden folgende Arterien analysiert: *Arteria carotis communis* (CCA), *Sinus caroticus* (Bulbus), *A. carotis interna* (ICA) und *A. carotis externa* (ECA), jeweils rechts und links (siehe Abb. 6A). Ein atherosklerotischer Plaque wurde definiert als Intima-Media-Dicke (IMT) > 15 mm oder fokale Intima-Media-Verdickung auf $+0.5$ mm oder $>50\%$ der benachbarten IMT, wobei die Messung für CCA 3 cm distal und für ICA und ECA 1.5 cm proximal durchgeführt wurde [20, 21] (siehe Abb. 6B).

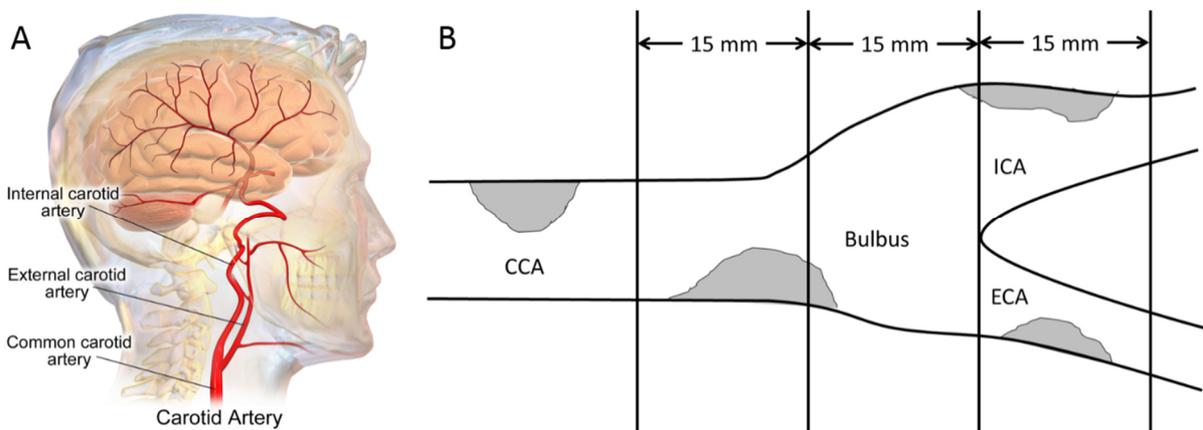


Abbildung 6: A) Schematischer Verlauf der Karotis-Gefäße (Carotid Artery) am Hals entlang (aus [22]). Eingezeichnet sind sowohl die Hauptschlagader (Common carotid artery, CCA), als auch die externe und interne Karotis (external carotid artery, ECA; bzw. internal carotid artery, ICA). Die Aufspaltung der CCA zur ECA und ICA stellt den Bulbus dar. B) Schematische Abbildung des Messbereichs für die Plaque-Detektion. Ein Plaque wurde definiert als Intima-Media-Dicke (IMT) > 15 mm oder fokale Intima-Media-Verdickung auf $+0.5$ mm oder $>50\%$ der benachbarten IMT. Die Messung für CCA wurde 3 cm distal und für ICA und ECA 1.5 cm proximal der Bifurkation durchgeführt.

Insgesamt wurde damit die Plaque-Präsenz (ja/nein = 1/0) in der Karotis an acht verschiedenen Stellen detektiert (abgekürzt als x_{ij} , wobei $i \in \{CCA, \text{Bulbus}, ECA, ICA\}$ und $j \in \{\text{rechts}, \text{links}\}$). Eine fehlende oder nicht beurteilbare Messung wurde als NA (not available, nicht verfügbar) notiert. Daraus lassen sich drei Phänotypen ableiten:

- Plaque-Prävalenz PP:

$$PP = \begin{cases} 1 & \text{falls mindestens ein } x_{ij} = 1 \\ 0 & \text{falls alle } x_{ij} = 0 \text{ und höchstens 2 } x_{ij} \text{ NA} \\ \text{NA} & \text{falls alle } x_{ij} = 0 \text{ und mehr als 2 } x_{ij} \text{ NA} \end{cases}$$

- Plaque-Score PS in CCA und Bulbus:

$$PS = \sum x_{ij}, \text{ wobei } i \in \{\text{CCA, Bulbus}\} \text{ und NA, falls mehr als ein } x_{ij} \text{ NA}$$

- Plaque-Score PS8 über alle Lokalisationen:

$$PS8 = \sum x_{ij}, \text{ wobei } i \in \{\text{CCA, Bulbus, ICA, ECA}\} \text{ und NA, falls mehr als zwei } x_{ij} \text{ NA}$$

Dies bedeutet für die Phänotypdefinition, dass die Plaque-Präsenz 1 oder 2 Fehlwerte aufweisen darf, bevor der Phänotyp PS oder PS8 selbst auf „fehlend“ gesetzt wird. Neben der Plaque-Belastung wurde die Intima-Media-Dicke (IMT) der ACC rechts und links vermessen (EchoPAC Dimension 06, GE Medical Systems, Munich, Germany) [4], wobei hier der Mittelwert der Mittelwerte beider Seiten als Phänotyp (cIMT) verwendet wurde.

Atherosklerose in den Koronararterien

Der Gefäßstatus der Koronararterien wurde durch eine Koronarangiographie bestimmt. Dies ist ein invasiver Eingriff, der nur bei Patienten der LIFE-Heart Studie im Herzzentrum Leipzig indiziert war.

In der Herzkatheter-Untersuchung wurden drei große Arterien berücksichtigt (*R. circumflexus*, *R. interventricularis anterior* und *A. coronaria dextra*, s. Abb. 7). Um den Koronarstatus zu beschreiben, wurden zwei Phänotypen definiert. Der erste Phänotyp ist der CAD-Status (coronary artery disease $\hat{=}$ KHK), wobei CAD definiert ist als eine Stenose $\geq 50\%$ in mindestens einem der drei Gefäße. Eingeschränkt auf das Verdachtskollektiv von LIFE-Heart ($n=3500$) wurde bei 42% der Patienten eine solche Stenose festgestellt.

Zusätzlich zum CAD-Status wurde als zweiter Phänotyp der Schweregrad der KHK im NVD50 Score erfasst (number of vessel diseased with stenosis of $\geq 50\%$ luminal reduction). NVD50 ist definiert als 0, falls lediglich Sklerosen mit weniger als 25% luminaler Reduktion detektiert wurden, 0.5, falls eine Sklerose zwischen 25% und 50% detektiert wurde und 1, 2 bzw. 3, für ein, zwei bzw. drei betroffene Gefäße mit $\geq 50\%$ Volumenreduktion.

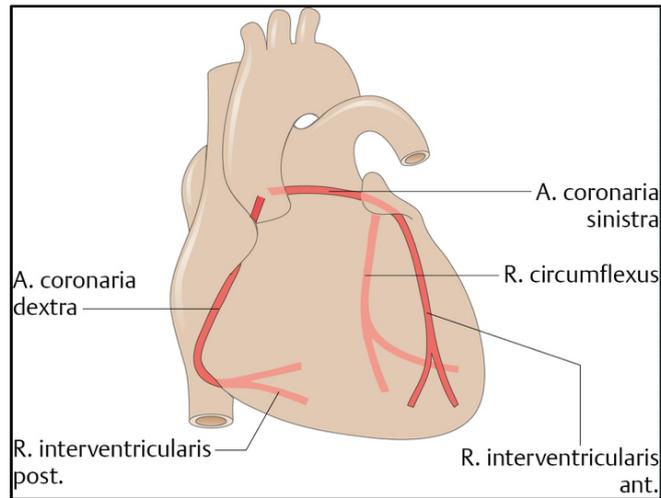


Abbildung 7: Schematischer Verlauf der Koronar-Gefäße (aus [23]). Von den gezeigten Gefäßen wurde der Stenosegrad in R. circumflexus, R. interventricularis anterior und A. coronaria dextra untersucht. Eine KHK wurde diagnostiziert bei einer Stenose von $\geq 50\%$ Reduktion des Gefäßlumens.

Atherosklerose in der Peripherie

Zur Erfassung des Gefäßstatus in den peripheren Gefäßen wurde der Knöchel-Arm-Index ermittelt (ABI). Dies wurde mit unterschiedlichen Methoden in LIFE-Heart und LIFE-Adult bestimmt.

In LIFE-Adult wurde der ABI mittels Photoplethysmographie (PPG) mit dem Gerät Vicorder (SMT Medical, Deutschland/Skidmore Medical, UK) gemessen. Pro Arm bzw. Knöchel wurden je drei PPG basierte Verschlussdruckmessungen durchgeführt. Der ABI wurde anschließend als Minimum der Quotienten der Verschlussdrücke von Knöchel zu Arm von rechts und links berechnet:

$$ABI = \min\left(\frac{VD_R}{VD_A}, \frac{VD_L}{VD_A}\right),$$

wobei VD_R bzw. VD_L der mittlere Verschlussdruck am rechten bzw. linken Knöchel ist und VD_A der mittlere Verschlussdruck beider Arme.

In LIFE-Heart wurde der Doppler Ultraschall angewandt (Huntleigh Mini-Dopplex, Kempfen, Germany). Beide Knöchel und der rechte Arm wurden je zweimal gemessen. Anschließend wurden die Werte gemittelt und die Quotienten wie oben gebildet. Auch hier wurde der minimale Wert des ABI von linkem und rechtem Knöchel als Phänotyp für die Analysen verwendet.

Aus dem kontinuierlichen Phänotyp ABI wurde die periphere arterielle Verschlusskrankheit (PAD, peripheral artery disease) als $ABI < 0.9$ festgelegt. Zusätzlich wurden alle Fälle, bei denen eine PAD aus der Anamnese bekannt ist, auf $PAD = 1$ gesetzt.

1.3.3 Grundlegende Methoden der Genotypisierung

Genotyp-Calling

Für beide LIFE Studien wurde ein Affymetrix SNP-Array verwendet, wobei für ein Teil der LIFE-Heart Studie ein modifizierter Array vorlag, der zusätzliche Kandidaten-SNPs der KHK beinhaltet. Das Calling wurde mit der jeweiligen aktuellen Version der „Affymetrix Power Tools“ gemäß der Affymetrix Guideline [24] durchgeführt. Dabei wurden Samples und SNPs gefiltert. Insgesamt gab es zum Zeitpunkt der Auswertung Genetikdaten von etwa 580,000 SNPs zu 4985 Samples in LIFE-Adult und 5700 in LIFE-Heart.

Sample-Qualitätskriterien	Schranke	Bedeutung
Dish QC	< 0.82	Kontraststärke des Arrays
Sample Call Rate	< 0.97	Qualität der Genotyp-Bestimmung pro Sample
XY-Intensität	Falsches Cluster, Ausreißer	Falsche Geschlechtszuordnung Gonosomale Chromosomenaberration
Verwandtschaftsanalyse	> 0.6	Klärung der Familienstruktur Kontrolle auf Duplikate
Hauptkomponentenanalyse (PCA)	> 6*SD	Detektion von ethnischen Ausreißern

Tabelle 1: Sample-Qualitätskriterien und ihre jeweiligen Schranken. Ausreißer im XY-Intensitäts-Plot waren Samples mit deutlich erhöhter Y-Intensität bei Männern bzw. deutlich verringerter oder erhöhter X-Intensität bei Frauen. Die Verwandtschaft wurde anhand der Methode von Wang [25] bestimmt. Für die Hauptkomponentenanalyse (PCA) wurden die ersten 10 Hauptkomponenten (PCs) bestimmt. Als Ausreißer wurden hier Samples definiert, die in einer der ersten 10 PCs mehr als 6 SD (Standardabweichung) vom jeweiligen Mittelwert abweichen.

Neben den von Affymetrix empfohlenen Sample-Filterungen bezüglich Signalkontrast und Sample-Call-Rate wurden Samples mit falscher Geschlechtszuordnung, hoher paarweiser Verwandtschaft und ethnische Ausreißer bezüglich der ersten 10 Eigenvektoren gefiltert. Für X-chromosomale Analysen wurden Ausreißer bezüglich der X- und Y-Intensitäten gefiltert (z.B. vermutliche X-Frauen, XXX-Frauen, XXY-Männer). Alle Sample-Kriterien sind in Tabelle 1 zusammengefasst.

Die SNPs wurden hinsichtlich Cluster-Kriterien von Affymetrix (FDL, HetSO, HomRO, nMA) gefiltert [24]. Zusätzlich wurde auf SNP-Call-Rate, Abweichung vom Hardy-Weinberg Gleichgewicht (HWE), Batcheffekte der Arrays und MAF kontrolliert. Für X-chromosomale SNPs wurden Kriterien verwendet, wie sie von König et al. [26] vorgeschlagen wurden. Dazu gehört, dass bei allen X-chromosomalen SNPs die heterozygoten Genotypen der Männer auf „fehlend“ gesetzt werden. Alle SNP-Kriterien sind in Tabelle 2 zusammengefasst.

Imputation

Ein wichtiges Maß für die Qualität eines SNP-Microarrays ist die Coverage. Diese beschreibt den Anteil an Referenz-SNPs, die in hinreichend hohem LD mit den Array-SNPs sind. Die Coverage ist abhängig von der verwendeten Referenz, der ethnischen Herkunft der Population, dem verwendeten LD-Niveau und dem Cut-Off für seltene Varianten.

SNP-Qualitätskriterien	Schranke	Bedeutung
Call Rate (CR)	< 97% (< 98%)	Qualität der Genotyp-Bestimmung pro SNP (Chr. X: getrennt für Männer und Frauen)
Hardy-Weinberg Gleichgewicht (HWE)	< 10^{-6} (< 10^{-4})	p-Wert von Fisher's Exakter Test (Chr. X: eingeschränkt auf Frauen)
Platten-Assoziation (PA)	< 10^{-7}	p-Wert von Chi-Quadrat Test
Minor Allel Frequenz (MAF)	= 0 (< 1%)	Filterung seltener/monomorpher Varianten (Chr. X: getrennt für Männer und Frauen)
Anzahl Minor Allels (nMA)	< 2	Filterung seltener/monomorpher SNPs
Fishers Linear Discriminant (FLD)	< 3.6	Minimaler Abstand zwischen den Clustern (bzgl. X-Achse)
Heterozygous Cluster Strength Offset (HetSO)	< -0.1	Abstand des AB-Clusters zu AA bzw. BB (bzgl. Y-Achse)
Homozygote Ratio Offset (HomRO)	< -0.9 (3 Cluster) < 0.3 (2 Cluster) < 0.6 (1 Cluster)	Verteilung der Cluster (bzgl. der X-Achse)

Tabelle 2: SNP-Qualitätskriterien und deren Schranken. Die Cluster-Kriterien nMA, FLD, HetSO und HomRO wurden direkt durch die Powertools berechnet und gemäß der Affymetrix-Empfehlung angewendet [24]. Die Call Rate, HWE-Abweichung, Plattenassoziation und MAF wurden mittels der Statistik-Software R berechnet und gemäß eigener Schranken verwendet. Für X-chromosomale SNPs wurden CR, HWE und MAF getrennt nach Männer und Frauen (CR, MAF) bzw. nur für Frauen (HWE) berechnet und Schranken gemäß der Empfehlung von König et al. [26] verwendet.

Die verwendeten Arrays haben eine hohe Coverage von etwa 90% (Referenz 1000 Genomes Project [27], CEU Population, $r^2 > 0.8$, $MAF \geq 1\%$). Man kann mittels der Überschneidung an SNPs in Referenz und eigenen Daten sowie unter Nutzung von LD die fehlenden Genotypen abschätzen. Dieses Verfahren nennt man Genotyp-Imputation. Es gliedert sich grob in drei Schritte: Zuerst müssen die SNP-Daten an die Referenz angepasst werden. Dafür sollen die Angaben zu Chromosom, RS-ID, Position auf dem Chromosom und die beiden Allele übereinstimmen. Als Referenz wurde zunächst 1000 Genomes Phase 1, Version 3 (2012) verwendet (Kapitel 2.1), später wurde der Datensatz auf die Referenz 1000 Genomes Phase 3, Version 5 (2015) aktualisiert (Kapitel 2.2 und 2.3).

Anschließend führt man eine sogenannte Phasierung durch, bei der die Haplotypen der SNP-Daten geschätzt werden. Mittels eines Hidden-Markov-Modells und der Referenz-Haplotypen wird für die gegebenen SNPs der wahrscheinlichste Haplotyp als Kombination von Referenzhaplotypen modelliert (Software SHAPEIT Version v2.r837 [28]).

Im letzten Schritt wird dann aus den modellierten Haplotypen der wahrscheinlichste Genotyp der benachbarten SNPs geschätzt (Software IMPUTE2 Version 2.3.2 [29]). Die Güte der Schätzung wird im info-Score gespeichert. Insgesamt wird durch diesen Vorgang die SNP-Zahl von ~580,000 auf etwa 38 Mio. SNPs (Phase 1) bzw. 85 Mio. SNPs (Phase 3) erhöht, wobei etwa 20 Mio. SNPs (Phase 1) bzw. 38 Mio. SNPs (Phase 3) mit einem info-Score > 0.3 eine ausreichende Imputationsqualität hatten.

1.3.4 Grundlagen der genomweiten Assoziationsanalyse

Eine genomweite Assoziationsanalyse (GWAS) besteht aus vielen einzelnen Analyse- und Auswerteschritten, die in einer GWAS-Pipeline der AG Genetische Statistik verbunden wurden. Eine Übersicht ist in Abbildung 8 angegeben.

Als Input werden die imputierten Genotypen und die zugehörigen Phänotypen benötigt. Die Pipeline berechnet die Assoziation jedes SNPs mit dem Phänotyp mittels PLINK oder SNPTEST [30] (Version 2.5.2), wobei man verschiedene genetische Modelle wählen kann (Standard: additiver Alleleffekt). Ein übliches Modell ist das lineare Regressionsmodell mit der Genosis G_{ij} als Einflussgröße und der Phänotyp X_i als kontinuierliche Zielgröße:

$$X_i \sim \beta_0 + \beta_j * G_{ij} + \beta_i * U_i + \varepsilon_i,$$

$$i = 1, \dots, n \text{ Patienten}, j = 1, \dots, m \text{ SNPs},$$

wobei β_j den Regressionsparameter des SNPs, U_i die Kovariablen (z.B. Alter und Geschlecht), β_i die Regressionsparameter der Kovariablen und ε_i den Fehlerterm darstellen. Am Ende der Assoziationsberechnung hat man zu jedem der SNPs den Effektschätzer $\hat{\beta}_j$, den zugehörigen Standardfehler $se(\hat{\beta}_j)$ und den p -Wert.

Um möglichst valide Hits zu erhalten, wird die SNP-Liste auf eine Top-Liste reduziert. Hierbei sind die wichtigsten Filterkriterien die MAF, der info-Score und der p -Wert. Die Grenze für den p -Wert kann für die Top-Liste auf $p < 1 \times 10^{-6}$ gesetzt werden. Dadurch erhält man neben den genomweit signifikanten Varianten ($p < 5 \times 10^{-8}$, Bonferroni-Korrektur) auch suggestive Hits, die eine genomweit signifikante Variante unterstützen oder neue Loci aufzeigen, für die eventuell die Power der Analyse nicht hoch genug war.

Zusätzlich wird das LD der Varianten analysiert. Hier wird ein *priority pruning* angewendet, d.h. dass von zwei SNPs in hohem LD die Variante mit kleinerem p -Wert in der Top-Liste verbleibt.

Um die Funktionalität der SNPs zu analysieren, folgt eine Annotation der Top-Liste, das heißt ein direkter Vergleich der Ergebnisse mit öffentlichen Ressourcen:

- Annotation der umliegenden Gene (± 250 kb) und ggf. eine Beschreibung der funktionellen Relevanz (z.B. Lokalisation in einem Exon) mittels der Ensembl Datenbank [31].
- Listung aller bereits als assoziiert berichteten Phänotypen des Tophits bzw. des Locus. Dazu wird die Top-Liste mit dem GWAS Katalog [32] verglichen und Phänotypassoziationen mit SNPs in LD $r^2 > 0.3$ notiert.
- Um zu klären, ob SNPs der Top-Liste einen Einfluss auf Expressionsniveaus von Genen haben, werden die Top-SNPs mit eQTL-Datenbanken (z.B. GTEx [33]) abgeglichen. Hier wird ebenfalls ein LD-Cutoff von 0.3 verwendet und sowohl cis- als auch trans-regulierte Gene notiert.

- Die bei der Genannotation gefundenen Gene und die eQTL Gene werden mit Pathway-Datenbanken wie DOSE [34] verglichen, um zu prüfen, ob es innerhalb biologischer Pathways Häufungen der gefundenen Gene gibt.

Das Ergebnis der Pipeline ist daher nicht nur eine Liste der SNP-Assoziationen, sondern auch eine Bewertung der Neuheit einer Assoziation, deren Einfluss auf die Genexpression und welche biologischen Pathways verstärkt betroffen sind. Zusätzlich werden für die Interpretation relevante Plots erzeugt (z.B. QQ-Plots, Manhattan-Plot und Regional Assoziation Plot).

Um die Power zur Bestimmung genetischer Assoziationen zu erhöhen, kann man mehrere GWAS zum gleichen Phänotyp im Rahmen einer Meta-Analyse zusammenfassen. In dieser Arbeit wurden LIFE-Heart und LIFE-Adult beispielsweise für den Endpunkt PS meta-analysiert, da in beiden Studien die Karotissonographie gleich durchgeführt wurde und der Endpunkt PS nach gleicher Definition erstellt wurde.

Es gibt mehrere Methoden, um eine Meta-GWAS durchzuführen. Eine der bekanntesten ist die *fixed-effect* Methode [35]: hier wird angenommen, dass der genetische Effekt aller Studien der gleiche ist. Meistens wird eine Gewichtung durch die inverse Varianz vorgenommen:

$$\hat{\beta}_{Fj} = \frac{\sum_k \hat{\beta}_{jk} * w_{jk}}{\sum_k w_{jk}}, \text{ wobei } w_{jk} = se(\hat{\beta}_{jk})^{-2} \text{ und } var(\hat{\beta}_{Fj}) = \frac{1}{\sum_k w_{jk}},$$

bei $k=1, \dots, K$ Studien und $j=1, \dots, J$ SNPs. Anschließend kann die entstandene SNP-Liste analog zur GWAS-Pipeline annotiert werden.

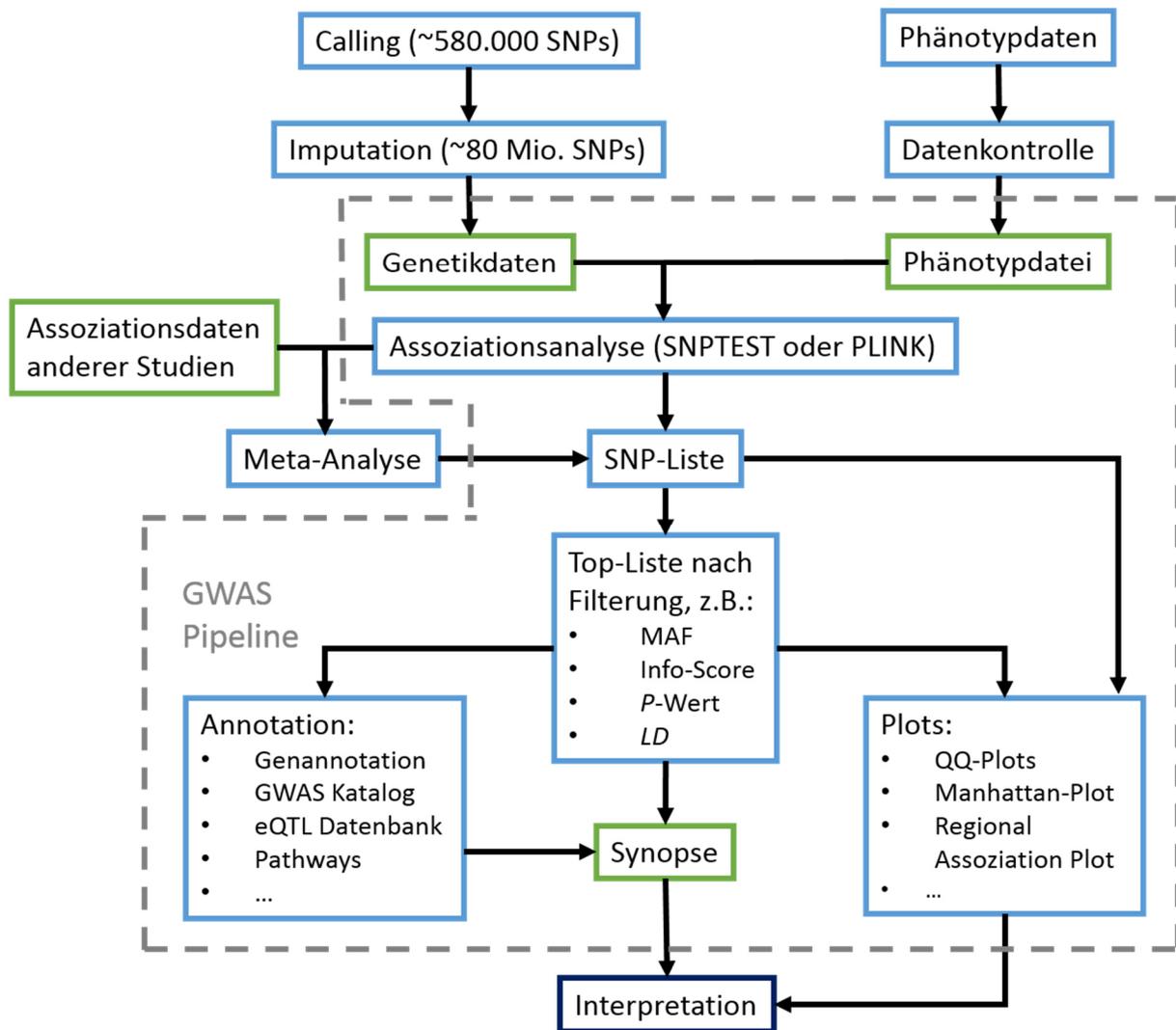


Abbildung 8: Schematische und vereinfachte Darstellung des Ablaufs einer GWAS bzw. Meta-GWAS nach einer Pipeline der AG Genetische Statistik des IMISE. Für eine GWAS benötigt man neben den Genetikdaten auch die geprüften und richtig zugeordneten Phänotypdaten und Kovariablen. Zum Start der Pipeline werden neben den SNP-Filter-Kriterien auch das genetische Modell (z.B. additiv) und das Regressionsmodell angegeben. Für die Meta-Analyse werden die ungefilterten Assoziationsergebnisse extrahiert und zusammen mit den Ergebnissen der anderen Studie auf die gleiche SNP-Menge eingeschränkt. Nach der Meta-Analyse kann die SNP-Liste wieder in die GWAS-Pipeline zurückgespielt und annotiert werden.

1.3.5 Enrichment Analyse

Plaques in der Karotis sind mit dem Vorliegen koronarer Stenosen assoziiert [36]. Beide werden durch ähnliche Risikofaktoren beeinflusst, was einen gemeinsamen Pathomechanismus nahelegt. Daher war es bei dieser Arbeit auch von Interesse zu testen, ob und wie viele genetische Risikoloci der KHK auch mit Karotis-Plaques assoziiert sind. Um zu prüfen, ob die beobachtete Überschneidung mehr als zufällig ist, wurden Enrichment-Analysen durchgeführt. Diese werden hier kurz beschrieben.

Es werden m Risiko-SNPs des Phänotyps Y auf Assoziation mit Phänotyp X getestet. Man betrachtet nun zwei Aspekte: die Signifikanz und die Konkordanz. Es seien n SNPs signifikant,

d.h. der Effekt des SNPs auf X unterscheidet sich signifikant von null, und k SNPs konkordant, d.h. die Effektrichtung des SNPs auf X ist die gleiche wie auf Y . Daraus lässt sich eine Vierfelder-Tafel erstellen, mit der man eine Odds-Ratio (OR) bestimmen kann (schematisch in Tab. 3 aufgeführt).

	Nicht konkordant	Konkordant	
Nicht signifikant	a	b	$m-n$
Signifikant	c	d	n
	$m-k$	k	m

Tabelle 3: Beispielhafte Vierfeldertafel in der Enrichment-Analyse. Eine Liste von m SNPs wird auf Signifikanz und Konkordanz geprüft (gepaartes Setting). Hier bedeutet Signifikanz „von Null verschiedener Effekt“ zu einem Signifikanzniveau von beispielsweise 5% und Konkordanz, dass das Vorzeichen des Effekts, bezogen auf das gleiche Effekt-Allel, mit dem in der Literatur übereinstimmt.

Um eine signifikante Anreicherung auf einfache Art zu testen, nutzt man nur die d SNPs, die sowohl signifikant als auch konkordant sind. Diese können mit einem einfachen Binomialtest zum Signifikanz-Niveau 2.5% getestet werden (d Treffer bei m Versuchen). Die Nullhypothese lautet „Die (zufällige) Treffer-Wahrscheinlichkeit liegt bei 2.5%“ mit der Überlegung, dass zufällig die Assoziation von 5% der SNPs signifikant wird und davon zufällig 50% konkordant sind. Daher ist es die Frage, ob die Beobachtung von d signifikanten und konkordanten SNPs signifikant größer ist als die $2.5\% \cdot m$ zufällig erwarteten SNPs.

Wie in Kapitel 1.1.3 beschrieben, werden in einer GWAS nicht zwingend die kausalen Varianten detektiert. Daher ist die Analyse von SNPs in LD mit den Literatur-SNPs sinnvoll ($r^2 > 0.5$). Man testet, ob die mit dem Literatur-SNP korrelierten Varianten mit dem Phänotyp X assoziiert sind. Der am besten assoziierte SNP wird anschließend als Proxy für den Literatur-SNP verwendet. Dies kann die Zahl von n assoziierte Literatur-SNPs auf n^* assoziierte Proxy-SNPs erhöhen. Hier ist allerdings keine Aussage über Konkordanz mehr möglich, da für den Proxy-SNP typischerweise keine Effekt-Richtung für den Phänotyp Y gegeben ist. Um den Anteil assoziierter Loci zu evaluieren, wird ein Permutationstest verwendet. Dabei kann unabhängig von der Verteilung getestet werden, ob ein beobachteter Zustand ein Extremfall ist oder nicht. Das Vorgehen beruht auf Resampling: Der Phänotyp X wird j -mal permutiert und erneut auf Assoziation mit den m Proxy-SNPs getestet. Pro Resampling wird die Anzahl der signifikanten Proxy-SNPs notiert und als zufällige Verteilung betrachtet. Wenn nun die beobachtete Anzahl von n^* signifikanten Proxy-SNPs größer ist als das 95%-Quantil der zufälligen Verteilung, dann kann man die Nullhypothese von zufällig assoziierten Loci zum Signifikanzniveau 5% ablehnen. Wenn die Verteilung bekannt ist, kann man direkt einen p -Wert bestimmen.

Um auf signifikante Anreicherung zu testen, kann man alternativ einen genetischen Risiko-Score (GRS) verwenden. Die Idee dahinter ist, die m viele Risikoloci von Y in einem Parameter zusammenzufassen und diesen dann auf Assoziation mit Phänotyp X zu testen. Dazu werden die Produkte der Effekte mit der jeweiligen Gendosis über alle m SNPs aufsummiert:

$$GRS = \sum_j \beta_j * G_j,$$

wobei β_j den Effektschätzer des j -ten Risiko-SNPs auf den Phänotyp Y und G_j die Gendosis des j -ten SNPs darstellt ($j=1, \dots, m$). Alle β_j werden bezüglich des Risikoallels bestimmt, d.h. $\beta_j > 0$. Je höher der GRS, desto höher das Risiko beziehungsweise die Werte von Phänotyp Y . Anschließend kann getestet werden, ob der GRS ebenfalls mit X assoziiert ist.

1.3.6 Mendelsche Randomisierung (MR)

Ein fundamentales Problem der Epidemiologie ist die Unterscheidung von Korrelation und Kausalität [37]. Um herauszufinden, durch welchen Risikofaktor X eine bestimmte Krankheit Y ausgelöst wird oder welchen Effekt ein bestimmtes Medikament verursacht, wird meistens eine prospektive randomisierte Studie durchgeführt. Im einfachsten Fall werden die Studienteilnehmer in zwei Gruppen aufgeteilt: eine Fallgruppe, die eine bestimmte Behandlung bezüglich des Risikofaktors X erhält, und eine Kontrollgruppe, die keine Intervention erfährt. Der betrachtete Risikofaktor ist in diesem Fall der auf Kausalität zu prüfende Faktor. Wenn die Gruppen balanciert sind, das heißt, wenn sie bezüglich aller anderen Störfaktoren U gleichverteilt sind, dann ist jede Änderung im Krankheitsbild alleine auf die Änderung des Risikofaktors zurückzuführen (s. Abb. 9). Dies wird typischerweise durch die Randomisierung des Studienkollektives erreicht. Allerdings sind solche Studien mit hohen Kosten und häufig langer Studiendauer verbunden. Unter Umständen sind sie sogar ethisch nicht durchführbar.

Um diese Probleme zu umgehen, wurde die „Mendelsche Randomisierung“ (MR) entwickelt, bei der die Randomisierung mittels Genetik erfolgt, also eine „instrumentelle Variablen [IV] Analyse mit genetischen Instrumenten“ [38]. Der Name leitet sich von zwei Mendelschen Regeln der Vererbung ab: der Segregationsregel und der Unabhängigkeitsregel [39]. Die erste Regel besagt, dass pro Gamet nur eine Kopie des Allels vorliegt, wobei die Aufteilung zufällig stattfindet. Die zweite Regel bedeutet, dass SNPs, die in sehr niedrigem LD sind, unabhängig von einander vererbt werden.

Drei grundlegende Bedingungen müssen erfüllt sein, um die MR-Methode anwenden zu können:

- (1) Die genetische Variante G ist mit dem Risikofaktor X assoziiert.
- (2) Die genetische Variante G ist unabhängig von allen Störfaktoren U , die den Zusammenhang von Risikofaktor X und Krankheit Y beeinflussen könnten.
- (3) Es gibt keine Beziehung zwischen der genetischen Variante G und Krankheit Y , außer über den Risikofaktor X [40].

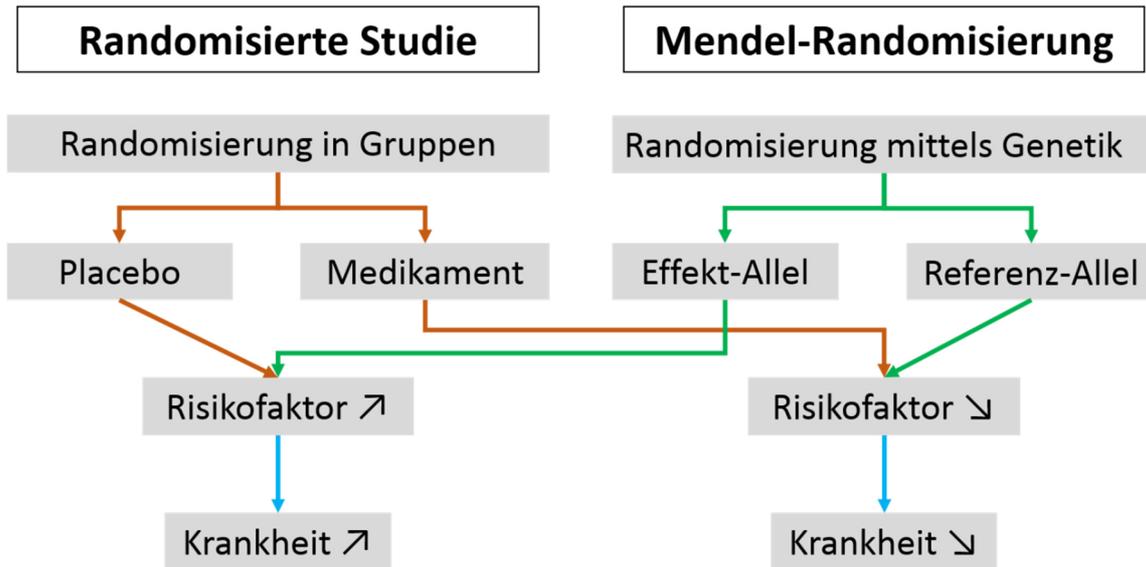


Abbildung 9: Vergleich zwischen Randomisierter Studie und Mendelscher Randomisierung. Das Effekt-Allel übernimmt in der Mendelschen Randomisierung die Rolle des Placebos, da beide den Risikofaktor erhöhen. Das Referenz-Allel bzw. das Medikament bewirken beide eine Senkung des Risikofaktors.

Die drei Bedingungen sind in einem azyklischen gerichteten Graphen zusammengefasst (s. Abb. 10). Die erste Bedingung kann mittels GWAS geprüft werden. Je stärker die Assoziation von G auf X ist, desto stärker ist G als Instrument. Die anderen beiden Bedingungen sind schwer nachweisbar und können meist nur plausibilisiert werden. Dazu können die Bradford Hill Kriterien verwendet werden [41].

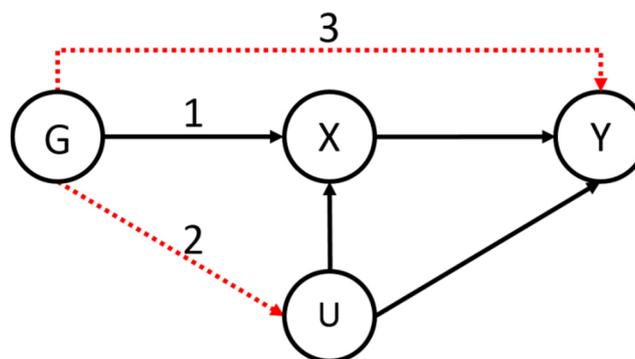


Abbildung 10: Gerichteter azyklischer Graph der Mendelschen Randomisierung. Die schwarzen Pfeile sind gültige Beziehungen, die rot-gestrichelten Pfeile dürfen nicht vorkommen. G steht für die genetische Variante, X für den Risikofaktor, der auf eine kausale Beziehung mit der Krankheit Y getestet wird. U bezeichnet alle bekannten Störgrößen von X und Y . Die Nummern der Graphen bezeichnen die jeweilige Bedingung, die erfüllt sein muss.

In genomweiten Assoziationsstudien wurde eine Vielzahl an genetischen Assoziationen mit klinischen Phänotypen gezeigt. Diese genetischen Risikoloci können für kausale Analysen genutzt werden: die Allele der Risikoloci werden gemäß der Unabhängigkeitsregel zufällig weitergegeben und sind in der Regel unabhängig voneinander (niedriges LD, da teilweise auf anderen Chromosomen). Dadurch ist eine Gruppen-Aufteilung durch die Genetik möglich. Je nach Genotyp wird der auf Kausalität zu prüfende Risikofaktor verstärkt oder geschwächt.

Zusätzlich kann plausibilisiert werden, dass der Genotyp nicht von anderen Faktoren U beeinflusst wird.

Für die Berechnung des kausalen Schätzers gibt es mehrere Methoden, sowohl für einzelne IVs als auch für mehrere. Die einfachste ist die Quotienten-Methode: Der kausale Effekt des Risikofaktors X auf den kontinuierlichen Phänotyp Y kann geschätzt werden, indem man die genetische Assoziation mit dem Phänotyp, $\hat{\beta}_Y$, durch genetische Assoziation mit dem Risikofaktor, $\hat{\beta}_X$, teilt [37]:

$$\beta_{IV} = \hat{\beta}_Y / \hat{\beta}_X \text{ und } se_{IV} = se(\hat{\beta}_Y) / \hat{\beta}_X$$

Bei mehreren unkorrelierten Instrumenten gibt es verschiedene Ansätze zur Schätzung von kausalen Effekten. Ein Ansatz ist die Verwendung eines GRS, in dem man alle Risiko-Loki in einem Instrument zusammenfasst. Alternativ kann man pro Instrument den kausalen Effekt einzeln schätzen und dann eine *fixed-effect* Meta-Analyse durchführen (Inverse-variance weighted, IVW) [42]:

$$\beta_{IV}(IVW) = \frac{\sum_j (\hat{\beta}_{Yj} / \hat{\beta}_{Xj}) * \hat{\beta}_{Xj}^2 se(\hat{\beta}_{Yj})^{-2}}{\sum_j \hat{\beta}_{Xj}^2 se(\hat{\beta}_{Yj})^{-2}} = \frac{\sum_j \hat{\beta}_{Yj} \hat{\beta}_{Xj} se(\hat{\beta}_{Yj})^{-2}}{\sum_j \hat{\beta}_{Xj}^2 se(\hat{\beta}_{Yj})^{-2}}, \text{ für } j \text{ IVs,}$$

$$\text{und } se_{IV}(IVW) = \sqrt{\frac{1}{\sum_j \hat{\beta}_{Xj}^2 se(\hat{\beta}_{Yj})^{-2}}} = \sqrt{\frac{1}{\sum_j w_j}}$$

Falls die Instrumente korreliert sind, kann man diese Methode modifizieren und auf die Korrelation adjustieren [42]. Die reine Gewichtung durch die inverse Varianz wird um die paarweise Korrelation erweitert:

$$w_{j_1, j_2} = \sum_{j_1} \sum_{j_2} \hat{\beta}_{X, j_1}^2 \hat{\beta}_{X, j_2}^2 se(\hat{\beta}_{Y, j_1})^{-2} se(\hat{\beta}_{Y, j_2})^{-2} r_{j_1, j_2}^2$$

2 Ergebnisse

In diesem Teil der Dissertation werden die Ergebnisse vorgestellt. Diese umfassen zwei Publikationen, die hier kurz beschrieben werden und anschließend vollständig aufgeführt sind. (Die Anhänge der Publikationen sind nicht aufgeführt. Sie sind online auf den Seiten der Journale verfügbar.) Zusätzlich werden weitere, bisher unveröffentlichte Ergebnisse kleiner Teilprojekte beschrieben, die keinen Eingang in die Publikationen erlangt haben. Dies umfasst einen Vergleich des genetischen Risikos der verwendeten Studien, eine genetische Analyse der Karotis-Plaques ohne Kontrollen und eine Haplotypenanalyse des *PCSK9* Locus.

2.1 „Genome-wide analysis identifies novel loci of plaque burden in carotid artery“

Pott et al. [43], DOI: <http://dx.doi.org/10.1016/j.atherosclerosis.2017.02.018>

Eingereicht: 23. September 2016; Akzeptiert: 22. Februar 2017

Plaques in der Karotis sind bekannte Marker der subklinischen Atherosklerose, und man nimmt an, dass es einen gemeinsamen Pathomechanismus der KHK und Karotis-Plaques gibt. Doch während es bereits viele GWAS und Meta-GWAS zur KHK gibt, fehlen solche für die Karotis-Plaques weitgehend. Das Ziel dieser Publikation war daher, genetische Varianten zu identifizieren, die einen signifikanten Einfluss auf Karotis-Plaques haben, und diese mit den bekannten KHK-Loki zu vergleichen.

Um dieses Ziel zu erreichen, wurde eine Meta-GWAS mit LIFE-Heart (n=3152) und LIFE-Adult (n=4037) bezüglich dem in der Einleitung beschriebenen Plaque-Score PS durchgeführt. Anschließend wurde mit einer Enrichment-Analyse von KHK-Loki (58 SNPs) und Schlaganfall-Loki (17 SNPs) auf Anreicherung getestet.

Zwei genomweit signifikante SNPs wurden in der Meta-GWAS detektiert: ein Hit im 9p21 Locus, der schon für KHK beschrieben wurde. Der zweite Hit wurde auf Chromosom 10 in einem Intron von *SFXN2* identifiziert. Die biologische Bedeutung des Locus ist unklar, allerdings gibt es SNPs im schwachen LD mit diesem Hit, die als Risiko-Loki für BMI, Bluthochdruck und KHK beschrieben wurden.

Von den 58 getesteten KHK-SNPs waren acht mit PS nominal assoziiert ($p < 0.05$), wobei bei sechs der SNPs die Effektrichtung übereinstimmte ($OR = 4.6$, $p = 3.2 \times 10^{-3}$). Mittels *priority-pruning* wurden SNPs in hohem LD ($r^2 > 0.5$) mit den KHK-SNPs detektiert, die stärker mit PS assoziiert sind. Unter Verwendung dieser SNP-Liste waren 17 KHK-Loki mit PS assoziiert ($p < 0.05$), was ebenfalls eine deutliche Anreicherung darstellt (empirischer p-Wert aus dem Permutationstest: $p = 1.03 \times 10^{-8}$). Auch bei Nutzung eines GRS für KHK konnte eine starke Assoziation mit PS detektiert werden ($\beta_{CAD} = 0.264$, $p = 1.99 \times 10^{-6}$). Zusätzlich konnte eine Korrelation der Effektstärken auf PS und KHK gezeigt werden ($r = 0.4$, $p = 0.002$), wobei die Effekte auf PS schwächer waren als auf KHK ($p = 1.5 \times 10^{-7}$).

Es wurde ebenfalls eine Anreicherung der beschriebenen Schlaganfall-Loki beobachtet: hier waren sechs der 17 Loki mit PS assoziiert (empirischer p-Wert: $p=2.44 \times 10^{-4}$). Eingeschränkt auf die elf Risikoloki eines Schlaganfalls in den großen Gefäßen (*large artery strokes*, LAS) wurde die Assoziation des entsprechenden GRS auf PS signifikant ($\beta_{LAS} = 0.114, p = 0.039$). Dies sind starke Hinweise auf eine über-zufällige Überlappung von Karotis-Plaques-Loki mit KHK- bzw. Schlaganfall-Loki.

Die Enrichment-Analyse wurde mit PS8 als Phänotyp wiederholt, der neben CCA und Bulbus auch die ICA und ECA abbildet. Interessanterweise verstärkte dies die Assoziation des GRS der Schlaganfall-Loki ($\beta_{Stroke} = 0.225, p = 6.8 \times 10^{-3}, \beta_{LAS} = 0.296, p = 9.4 \times 10^{-4}$). Der stärkere Überlapp in genetischen Risikofaktoren zwischen PS8 und Schlaganfall kann dadurch erklärt werden, dass das Risiko eines Schlaganfalls stärker von Plaques in ICA beeinflusst wird.



Genome-wide meta-analysis identifies novel loci of plaque burden in carotid artery



Janne Pott^{a,b}, Ralph Burkhardt^{b,c}, Frank Beutner^{b,d}, Katrin Horn^a, Andrej Teren^{b,d}, Holger Kirsten^{a,b}, Lesca M. Holdt^{b,e}, Gerhard Schuler^d, Daniel Teupser^{b,e}, Markus Loeffler^{a,b}, Joachim Thiery^{b,c}, Markus Scholz^{a,b,*}

^a Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany

^b LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany

^c Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital, Leipzig, Germany

^d Heart Center Leipzig, Leipzig, Germany

^e Institute for Laboratory Medicine, Ludwig-Maximilians University, Munich, Germany

ARTICLE INFO

Article history:

Received 23 September 2016

Received in revised form

13 February 2017

Accepted 22 February 2017

Available online 24 February 2017

Keywords:

Atherosclerosis

Carotid artery plaque

Genome-wide association analysis

ABSTRACT

Background and aims: Carotid artery plaque is an established marker of subclinical atherosclerosis and common patho-mechanisms with coronary artery disease (CAD) are hypothesized. We aimed to identify genetic variants associated with carotid plaque and to examine the potential shared genetic basis with CAD.

Methods: After investigating the reliability of plaque detection, we performed a genome-wide meta-association study in two independent cohorts (LIFE-Adult, $n = 4037$ and LIFE-Heart, $n = 3152$) for carotid plaque score (PS), defined as the sum of the plaque load of common carotid artery and carotid bulb. Further, we analyzed whether previously reported CAD and stroke loci were also associated with PS.

Results: We identified two loci with genome-wide significance for PS. One locus is the known CAD-locus at chromosome 9p21 (lead SNP rs9644862, $p = 8.73 \times 10^{-12}$). We also describe a novel locus on chromosome 10q24 within the *SFXN2* gene as the most probable candidate (lead SNP rs2902548, $p = 1.97 \times 10^{-8}$). In addition, 17 out of 58 known CAD loci and six of 17 known stroke loci were associated with PS at a nominal level of significance.

Conclusions: We showed that PS is a reliable trait to analyze genetics of atherosclerosis. Two new loci of genome-wide significant association with PS were found. The observed non-random overlap of CAD and PS associations strengthens the hypothesis of a shared genetic basis for these atherosclerotic manifestations.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Coronary artery disease (CAD) is a complex disease determined by numerous environmental and genetic factors [1]. Over the last several years, genome-wide association analyses have led to the identification of several genetic loci associated with CAD and myocardial infarction (MI). The most recent meta-analysis of Nikpay et al. [2] included about 185,000 cases and controls adding further evidence and resulting in a total of 58 loci considered as

well-established. However, these studies typically comprise a highly heterogeneous mixture of assessments and information to determine disease status, e.g. anamnestic data, clinical records, coronary angiography or acute myocardial infarction. Therefore, to gain a better understanding of the underlying genetic patho-mechanisms, it appears worthwhile to study whether these loci are also related to other manifestations of atherosclerosis.

Carotid intima-media thickness (cIMT) and carotid artery plaque are promising assessment for this purpose. Intima-media thickness was reported to be predictive for cardiovascular events [3,4]. However, the importance of cIMT in comparison to established risk scores such as the Framingham Risk Score was challenged by others [5]. Studies have also shown an association of carotid parameters with prevalent CAD with the predictive value of carotid plaque

* Corresponding author. Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany.

E-mail address: markus.scholz@imise.uni-leipzig.de (M. Scholz).

outperforming that of intima-media thickness [6,7]. In a large cohort of patients with suspected CAD receiving coronary angiography (LIFE-Heart), we have recently also found a strong correlation of plaque status in the carotid artery and lesions in coronary vessels with an odds ratio (OR) > 3.7 [8]. Again, carotid plaque showed a considerably stronger association with CAD than intima-media thickness proposing carotid plaque as a non-invasively assessable marker of coronary lesions. Based on these findings, we hypothesized that there may be a shared genetic basis for carotid artery plaques and CAD.

Here, we performed a large genome-wide association analysis of carotid plaques in two independent cohorts collected in the LIFE Research Center of Civilization diseases. While previous studies focused on plaque prevalence [9,10] or plaque size [11], we used the number of carotid plaques as primary endpoint.

In a secondary analysis, we investigated whether there is an enrichment of plaque associations for CAD risk loci to detect possible genetic similarities and differences of these atherosclerotic phenotypes.

In analogy, we analyzed loci associated with ischemic stroke (IS) as found by Dichgans et al. [12] and Pulit et al. [13]. Both authors defined subtypes of stroke using the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification system [14]. One subtype is large artery stroke (LAS), which requires a stenosis of greater than 50% of the extracranial *internal carotid artery* (ICA). We analyzed this sub-phenotype separately since plaque assessment was available in proximal ICA in our cohorts.

2. Materials and methods

2.1. Cohort description

2.1.1. LIFE-Adult

LIFE-Adult is a population-based cohort of 10,000 adult inhabitants of the city of Leipzig, Germany. Participants are well characterized regarding life-style and environmental risk factors and clinical and subclinical signs of diseases such as cardiovascular diseases, type 2 diabetes or cognitive impairment. Detailed description of the cohort can be found elsewhere [15]. CAD status in LIFE-Adult was determined by the following anamnestic criteria: MI, stent implantation during a coronary angiography or diagnosed CAD.

2.1.2. LIFE-Heart

LIFE-Heart is a cohort of patients with suspected and confirmed stable coronary artery disease or myocardial infarction as first manifestation of CAD. All patients underwent coronary angiography and vascular phenotyping at the Heart Center of the University of Leipzig. Details of the study can be found elsewhere [16]. A total of about 7000 patients were recruited so far. CAD status is defined by the presence of at least one coronary lesion with more than 50% luminal reduction.

Both LIFE-Adult and LIFE-Heart meet the ethical standards of the Declaration of Helsinki. They have been approved by the Ethics Committee of the Medical Faculty of the University Leipzig, Germany (LIFE-Adult: Reg. No 263-2009-14122009, LIFE-Heart: Reg. No 276-2005). LIFE-Heart is registered at ClinicalTrials.gov (NCT00497887). Written informed consent including agreement with genetic analyses was obtained from all participants.

A description of basic parameters of LIFE-Adult in comparison with those of LIFE-Heart is shown in [Table 1](#).

2.2. Carotid ultra-sound and plaque assessment

For both cohorts, carotid ultrasound was performed using the same standard operating procedures. Subjects with cervical spine disorder, wounds at the scanning area and patients with acute myocardial infarction in LIFE-Heart were excluded from carotid ultrasound.

Eligible patients were scanned at both sides of four anatomical regions: *common carotid artery* (CCA), *carotid bulb* (Bulb), proximal parts of *internal carotid artery* (ICA) and *external carotid artery* (ECA) respectively. High-resolution B-mode ultrasound images of carotid vessels were acquired using the GE Vivid ultrasound platform with a 12.0-MHz linear-array transducer (GE-Healthcare). For the assessments, subjects were in supine position.

Carotid artery plaque was defined according to the American Society of Echocardiography Intima-Media Thickness Task Force [17]. In detail, a lesion was counted as plaque if echogenic thickening of intimal reflection that extends into the arterial lumen at least 0.5 mm or 50% of the surrounding intima-media thickness or thickness of intima and media >1.5 mm. Plaque presence was documented as 'present' or 'absent' or 'missing' if the quality of the image was insufficient. Details can be found elsewhere [8].

Prior to LIFE recruitment, we performed a feasibility study of N = 47 volunteers receiving repeated plaque readings by six different investigators. We observed a high intra- ($\kappa = 0.93$ for the overall agreement) and inter-rater reliability ($\kappa = 0.88$ for the overall agreement with the consented plaque assessment, data not shown). Based on the single plaque assessments, we defined the plaque score (PS) as follows. PS is the sum of the plaque assessments at CCA and Bulb on both sides. Thus, PS takes values from 0 to 4. Single missing values are counted as zeros for this purpose. If there is more than one missing value, the score is set to missing.

A possible alternative to this phenotype definition is a more comprehensive evaluation of all four scanned carotid areas on both side (PS8, with values in between 0 and 8). Plaque scores PS and PS8 were highly correlated ($r = 0.91$ for both LIFE-Heart and LIFE-Adult). However, as ICA and ECA were more difficult to scan, there is a substantially higher percentage of missingness for PS8 compared to PS. Therefore, we decided to use PS as our primary endpoint. However, for consistency we verified our main results using this alternative plaque phenotype. For the purpose of comparisons with established stroke loci, we analyzed the four anatomical regions separately considering scores of corresponding plaque burden, which take values 0, 1 and 2.

Carotid ultra-sound was available for 9858 participants of LIFE-Adult and 3501 patients of LIFE-Heart. Distributions of PS in both cohorts are presented in [Table 1](#).

2.3. Genotyping

2.3.1. LIFE-Adult

A total of 5101 randomly selected individuals were genotyped using the genome-wide SNP array *Affymetrix Axiom CEU1*. Genotype calling was performed following the best practice steps recommended by Affymetrix [18]. The software "*Affymetrix Power Tools*" (APT, version 1.17.0) was used with the latest library "*Axiom GenomeWide CEU 1 Array Plate, Analysis Files, release 6*".

We filtered 73 samples failing the dishQC criteria (dishQC ≥ 0.82) and sample call rate criteria (SCR $\geq 97\%$) in the initial calling round. Genotypes of 587,352 SNPs were determined in the final calling round. Sex-mismatches were filtered. Additionally, the intensities of gonosomal SNPs were plotted to check for further abnormalities as proposed by Laurie et al. [19]. Cryptic relatedness was assessed according to Wang [20]. Duplicates were removed keeping the sample with the higher quality. Principal

Table 1
Patient and participants' characteristics of LIFE-Heart and LIFE-Adult cohorts.

Parameter	LIFE-Heart	LIFE-Adult	p value
Women/Men (with Chr X)	1111/2041 (1110/2039)	2182/1855 (2179/1852)	<0.001
Age (years)	61.7 ± 10.7	62.1 ± 11.3	ns
BMI (kg/m ²)	29.7 ± 5.0	27.2 ± 4.3	<0.001
Diabetes: known, in therapy or acute	963	422	<0.001
Smokers: former/current	1330/659	1154/672	<0.001
SBP/DBP (mmHg)	139/84	131/76	<0.001
Hypertension: known, in therapy or acute	2839	2290	<0.001
Lipid therapy	1192	649	<0.001
Cholesterol (mmol/L)	5.44 ± 1.21	5.71 ± 1.07	<0.001
HDLC (mmol/L)	1.34 ± 0.41	1.66 ± 0.47	<0.001
LDLC (mmol/L)	3.33 ± 1.04	3.55 ± 0.95	<0.001
CAD	1259	271	
Plaque score (PS)			<0.001
	0	1820	
	1	956	
	2	902	
	3	245	
	4	114	

All parameters are restricted to patients/participants genotyped for the autosomes and with valid PS. For the continuous parameters, the unit is in parenthesis and the arithmetic mean and standard deviation are given. All categorical parameters were tested with a Chi-squared test, continuous parameters were tested with a Mann-Whitney U test. Plaque score was tested with proportional odds regression. All parameters except age show significant differences between the studies. CAD status was not compared due to different definitions (LIFE-Heart: coronary angiography, LIFE-Adult: self-reported).

component analysis (PCA) as suggested by Patterson et al. [21] was performed using PLINK [22,23] (version 1.9). Ethnical outliers (>6 SD of first ten PCs) were removed, which resulted in a final set of 4985 unrelated high quality samples for autosomal analyses and 4978 for gonosomal analyses.

Autosomal SNPs were filtered for call rate (CR < 97%), the cluster plot quality criteria as suggested by Affymetrix (i.e. Fisher's Linear Discriminant, heterozygous strength offset, homozygote ratio offset), p value of exact test for Hardy-Weinberg equilibrium (HWE < 10⁻⁶), monomorphic SNPs and p value of plate association (< 10⁻⁷, i.e. test for dependency of allele frequency to plate). A total of 532,875 SNPs fulfilled all quality criteria.

For X-chromosomal SNPs, filter criteria proposed by König et al. [24] were applied (p value of exact test for HWE for women < 10⁻⁴, CR < 98%, and minor allele frequency (MAF) < 1% after setting all heterozygous calls in male samples on missing). This resulted in 13,554 X-chromosomal SNPs eligible for analysis.

Finally, all 4985 samples were imputed, using 1000 Genomes Phase 1, Version 3 [25] (2012) as reference, SHAPEIT [26] (version v2.r790) for phasing, and IMPUTE2 [27] (version 2.3.2) for imputation. Imputation for X-chromosomal SNPs was performed using the same reference panel and software, but with the subset samples eligible for X-chromosomal analysis as mentioned above.

2.3.2. LIFE-Heart

LIFE-Heart samples were genotyped with either *Affymetrix Axiom CEU1* or *Affymetrix Axiom CADLIFE*. The latter is an array containing Axiom CEU as genome-wide backbone and an additional custom content of about 62,500 SNPs from CAD loci. Genotype calling relied on *Affymetrix Power Tools* (APT, version 1.17.0 for Axiom-CADLIFE, version 1.16.1 for Axiom-CEU) with their latest libraries (Axiom CADLIFE1, release 3 respectively Axiom Genome-Wide CEU 1 Array Plate, Analysis Files, release 6). Genotype calling and SNP filtering was performed separately for the two array products. For sample filtering, high quality SNPs in the intersection of both arrays were used. The same steps of sample and SNP filtering were performed as in LIFE-Adult. In summary, 5700 samples and 504,593 SNPs fulfilled all quality criteria for autosomal analyses. For X-chromosomal analyses, additional 12 samples were removed and a total of 12,715 SNPs fulfilled the specified quality

criteria.

Imputation was performed using the same reference, software and software settings as in LIFE-Adult. The intersection of SNPs of *Affymetrix Axiom CEU1* and *Affymetrix Axiom CADLIFE* was used for this purpose.

2.4. Statistical analysis

PS was available for 4037 successfully genotyped samples of LIFE-Adult and 3152 samples of LIFE-Heart sub cohort of patients initially admitted for suspected CAD. Separate genome-wide analysis was performed for each cohort first. The analysis was executed with PLINK [22,23] (version 1.9) using the additive frequentist model and expected genotype counts, adjusting for sex and age (LIFE-Heart: beta = 0.54 for males and beta = 0.03/year for age; LIFE-Adult: beta = 0.44 for males and beta = 0.03/year for age; all betas significant with p < 0.001). X-chromosomal SNPs were analyzed assuming total X inactivation. PS was treated as continuous since the difference of test statistics compared to those of proportional odds logistic regression appeared to be small (correlation of test statistics r > 0.99 for both cohorts).

SNPs not in the intersection of the studies, with MAF < 1% or info-score < 0.3 in one of the studies were initially filtered, leaving 9,846,747 million SNPs for further analysis. A fixed effect model (FEM) meta-analysis was then performed for all remaining SNPs using the package *meta* of R. A genome-wide significance threshold of 5.0 × 10⁻⁸ was applied. The study has a power of 80% respectively 90% to detect variants explaining 0.5% respectively 0.6% of variance.

To determine independent variants, we applied a priority pruning using linkage disequilibrium data from 1000 Genomes Phase 1, Version 3 (2012), restricted to individuals with European ancestry. Variants which are in linkage disequilibrium (LD) with an association of higher significance were removed if r² ≥ 0.5.

We performed a comprehensive annotation of our top-SNPs using a number of bioinformatic resources: We reported known genome-wide association study (GWAS) hits, using the *GWAS catalogue* [28], if in linkage disequilibrium (r² > 0.2) with one of our top-hits. Analogously, we add expression quantitative trait loci (eQTL) data as explained in Kirsten et al. [29]. Pathway enrichment

was performed for each SNP by including physically nearby genes, all within 50 kb distance and up to four within 250 kb distance, and eQTLs in LD ($r^2 > 0.2$) with the variant. Pathways were retrieved from KEGG, GO, DOSE [30], and reactome. We also calculated deleteriousness scores according to Kircher et al. [31].

In addition, we tried to replicate five SNPs which were reported to be associated with carotid plaque prevalence (Bis et al. [9]) with our score. To analyze a possible shared genetic background of carotid plaques and CAD, we analyzed 58 loci for which associations with CAD were described in the literature [2]. Of the 58 top-SNPs of these loci, ten were directly genotyped in both of our cohorts. All other SNPs were successfully imputed. To analyze enrichment of significant associations with concordant direction of effects, we compared this percentage with 2.5% expected by chance.

We also analyzed SNPs in LD ($r^2 > 0.5$) with the reported CAD SNPs and determined the SNP with strongest plaque association per locus. This results in a higher number of CAD loci showing nominal significance for plaque. To compare this number with the one expected by chance, we performed permutation analysis. 10,000 permutations of phenotypes (PS, sex, age) and genotypes were analyzed. An empirical p value was calculated assuming a Poisson distribution of the number of significant loci. Additionally, we calculated a genetic risk score (GRS) of CAD using the beta estimates of Nikpay et al. [2] in order to analyze its association with our trait.

Similarly, we analyzed loci associated with cIMT [9,32], ischemic stroke (IS) [12,13] and large artery stroke (LAS) [12].

3. Results

3.1. Genome-wide significant hits for plaque score

Meta-analysis of 9,576,485 SNPs in the LIFE-Adult and LIFE-Heart studies showed no signs of inflation ($\lambda = 1.004$). QQ-Plot and Manhattan plot can be found in Fig. 1. We identified two loci at chromosome 9p21 and 10q24 with genome-wide significant hits ($p < 5 \times 10^{-8}$) in our study (Table 2, further results are summarized in Supplementary Table 1).

The strongest association with PS was observed for SNP rs9644862, located at 9p21 in the sequence of *CDKN2B-AS1* (also known as *ANRIL*), i.e., the best established common locus of CAD

association. This SNP showed robust associations with similar effect sizes in both cohorts (LIFE-Heart: $\beta = 0.174$, $p = 3.36 \times 10^{-7}$, LIFE-Adult: $\beta = 0.111$, $p = 1.87 \times 10^{-6}$, FEM: $\beta = 0.131$, $p = 8.73 \times 10^{-12}$). No interactions with sex were observed (LIFE-Heart: $p = 0.53$, LIFE-Adult: $p = 0.47$). Our lead SNP rs9644862 is in linkage disequilibrium ($r^2 = 0.75$) with rs2891168, which was reported as top-SNP in the most recent GWAS of CAD [2] (see also Supplementary Table 2). For our top-SNP no direct eQTLs were observed. But it is in strong LD ($r^2 = 0.75$) with rs4977574 which is an expression quantitative trait nucleotide (eQTN) of *CDKN2B* [33]. Other, partly stronger eQTLs but in weaker LD with the top-hit are presented in Supplementary Table 3.

After pruning with $r^2 = 0.5$, another SNP at the 9p21 locus (rs7853090) also showed genome-wide significance. This SNP is in weak LD with the top-SNP ($r^2 = 0.28$). A regional association plot of the locus is shown in Fig. 2A.

Besides the 9p21 locus, we identified a second locus at 10q24 reaching genome-wide significance (lead SNP rs2902548; FEM: $\beta = -0.141$, $p = 1.97 \times 10^{-8}$, Table 2). The lead SNP is located in the intron of *SFXN2*. Other nearby genes are *ARL3* (13 kb), *WPL1P* (16 kb) and *TRIM8* (69 kb, Fig. 2B). SNP rs2902548 is an eQTN for several *cis*-regulated genes including those mentioned above (Supplementary Table 3). The strongest eQTL ($p = 8.6 \times 10^{-26}$, The GTEx Consortium [34]) was observed for *SFXN2* in esophagus tissue, and a weaker one in left ventricular heart tissue ($p = 6.1 \times 10^{-10}$, The GTEx Consortium [34]). There is also an eQTL-SNP in LD with rs2902548 at *ARL3* in tissue of the artery aorta ($p = 1.5 \times 10^{-5}$, $r^2 = 0.56$, The GTEx Consortium [34]). Restricted to blood tissue, the strongest eQTL was found for *SFXN2*, *C10orf32*, and *ARL3* ($p = 3.6 \times 10^{-15}$, $p = 1.8 \times 10^{-11}$, and $p = 4.2 \times 10^{-10}$, respectively, Westra et al. [35]). SNP rs2902548 is also in weak LD ($r^2 < 0.3$) with several SNPs previously reported for cardiovascular risk factors such as BMI and hypertension (Supplementary Table 2). Further, it is in weak LD ($r^2 = 0.20$) with SNP rs12413409 which has been associated with CAD [12,33,36]. But conversely, the latter was only weakly associated with PS ($p = 0.01$).

3.2. Suggestive plaque score loci identified in the meta-analysis

A total of five SNPs showed suggestive association ($p < 1 \times 10^{-6}$) with PS but failed genome-wide significance. While two of them

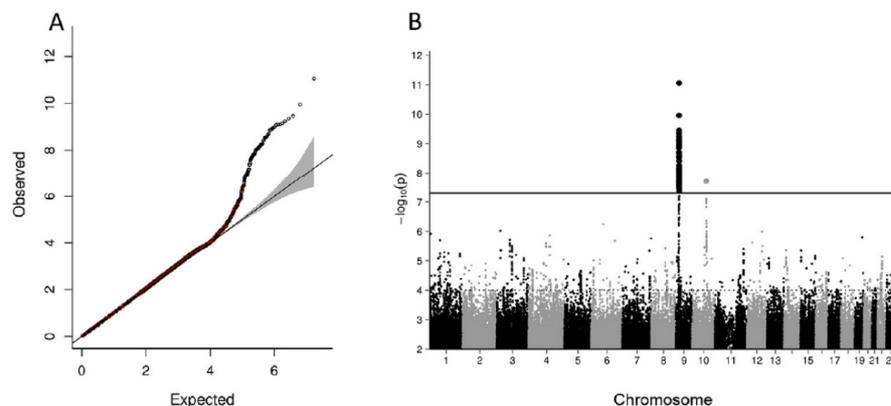


Fig. 1. QQ-Plot and Manhattan-Plot.

(A) QQ-Plot of p values in our meta-analysis. SNPs with info score less than 0.8 are plotted as triangles, and those with MAF less than 0.05 are shown in red. There is no strong inflation ($\lambda = 1.0041$). (B) Distribution of log transformed p values in our meta-analysis with trait PS. The bold line marks genome-wide significance ($p = 5 \times 10^{-8}$). Two loci reach genome-wide significance, one at chromosome 9 (*CDKN2B-AS1*), and the other one at chromosome 10 (*SFXN2-WBP1L*). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.1 „Genome-wide analysis identifies novel loci of plaque burden in carotid artery“

36

J. Pott et al. / *Atherosclerosis* 259 (2017) 32–40

Table 2
Genome-wide significant and suggestive SNPs.

SNP Infos					Combined – fixed effect model				LIFE-Heart			LIFE-Adult		
Locus	Lead SNP	Nearby gene (distance)	Other SNPs $r^2 < 0.5$ and $p < 1 \times 10^{-6}$	Effect/other allele	β	SE	p value	Q	EAF	p value	β	EAF	p value	β
9p21	rs9644862	<i>CDKN2B-AS1</i> (0 kb)	rs7853090, rs1333050, rs10811650	G/T	0.131	0.019	8.73×10^{-12}	2.341	0.453	3.36×10^{-7}	0.174	0.425	1.87×10^{-6}	0.111
10q24	rs2902548	<i>SFXN2</i> (0 kb)/ <i>WBP1L</i> (16 kb)	rs9663711	T/C	-0.141	0.025	1.97×10^{-8}	1.681	0.184	3.94×10^{-2}	-0.093	0.183	7.80×10^{-8}	-0.163
6q13	rs141249844	<i>KCNQ5</i> (0 kb)		T/C	0.475	0.095	5.79×10^{-7}	0.431	0.022	3.21×10^{-4}	0.554	0.020	4.25×10^{-4}	0.426
3p24	rs1349287	<i>KCNH8</i> (130 kb)		A/T	0.208	0.042	9.54×10^{-7}	0.176	0.060	1.67×10^{-3}	0.234	0.058	1.59×10^{-4}	0.196

SNPs associated with carotid plaque at genome-wide significant levels and at suggestive levels ($p < 1 \times 10^{-6}$) are shown. Beta estimates and p values of LIFE-Heart and LIFE-Adult were obtained by linear regression analysis of plaque score adjusting for age and sex.

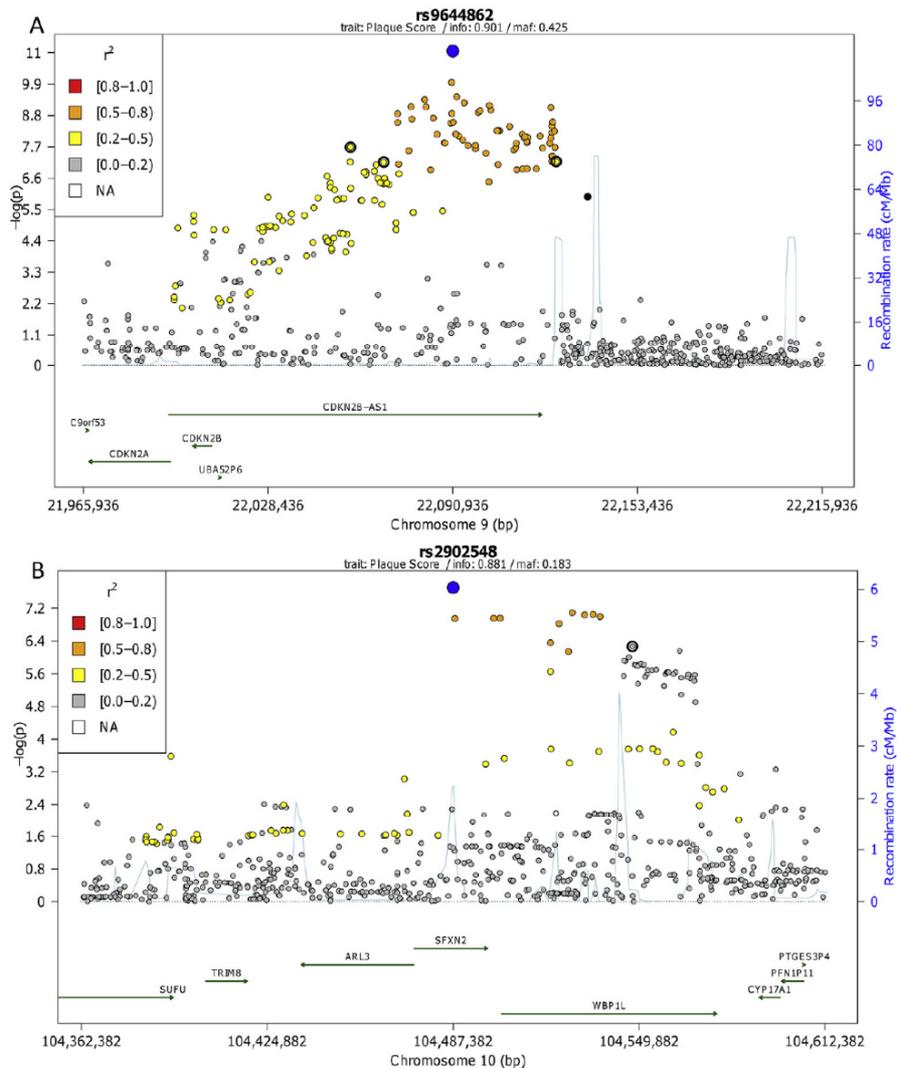


Fig. 2. Regional association plot.

Regional association plots of the two genome-wide significant loci at chromosome (A) 9 and (B) 10. The lead SNP of each locus is shown in blue (rs9644862 and rs2902548 in A and B respectively). All other SNPs are colored according to their linkage disequilibrium to the lead SNP. Additional suggestive SNPs ($p < 1 \times 10^{-6}$ and pairwise $r^2 < 0.5$) are encircled in black.

correspond to the 9p21 locus and one to 10q24, two additional new loci were found: At chromosome 6q13, the lead SNP rs141249844 is located in the intron of *KCNQ5* and at chromosome 3p24 the lead SNP rs1349287 is located in close proximity to *KCNH8*. Statistics of all loci are summarized in Table 2. Interestingly, both genes belong to the calcium channel family.

3.3. Verification of top and suggestive SNPs with PS8

We verified our results with an alternative plaque endpoint, namely PS8. The 9p21 locus remained genome-wide significant for PS8 (beta = 0.193, $p = 7.34 \times 10^{-10}$). The second hit on 10q24 reached suggestive evidence for PS8 (beta = -0.211, $p = 2.63 \times 10^{-7}$). For the suggestive loci at 6q13 and 3p24 no suggestive significance was found for the phenotypic alternatives ($p = 5.35 \times 10^{-5}$ and $p = 3.08 \times 10^{-4}$, respectively, see also Supplementary Table 4). Interestingly, higher absolute betas were observed for PS8 which however did not result in better p values due to the smaller sample size available for PS8.

3.4. Replication of SNPs associated with plaque prevalence

A recent meta-GWAS for carotid plaque prevalence (Bis et al. [9]) reported two SNPs with genome-wide significance and three SNPs with suggestive p values. We tested, whether these 5 SNPs were also associated with carotid plaque score in our study.

We found evidence for replication in the LIFE studies (nominal significance and same direction of effect) for four of the five SNPs, including the two SNPs with genome-wide significance in Bis et al. [9] (Table 3 and Supplementary Tables 4 and 5). Only SNP rs17045031/*LRIG1* at chromosome 3p14 failed replication. The same results were observed if analyzing plaque prevalence instead of PS (not shown).

3.5. Analysis of known CAD loci for association with plaque score

To investigate if the development of atherosclerotic plaques in the carotid and coronary arteries may share a common genetic background, we analyzed whether the 58 lead SNPs at known CAD-loci were also associated with PS (see Table 3 and Supplementary Tables 4 and 5). CAD association statistics were retrieved from Nikpay et al. [2], the most recent meta-GWAS on CAD.

The CAD lead SNP at the chromosome 9p21 locus, rs2891168, also reached genome-wide significance with PS ($p = 2.31 \times 10^{-8}$). This SNP is in high LD ($r^2 = 0.75$) with the PS lead SNP rs9644862 of our study. In addition, seven other CAD lead SNPs reached nominal significance for PS and of those, five SNPs showed concordant direction of effects (Table 3). Thus, concordant SNPs with nominal significance are enriched in this analysis (OR = 4.6, $p = 3.2 \times 10^{-3}$).

Additionally, the literature OR of the CAD SNPs are correlated with the beta estimates ($r = 0.4$, $p = 0.0017$, after transforming the OR as suggested by Chinn [37], see also Supplementary Fig. 1). However, the beta estimates for PS were on average smaller than those for CAD (t -test of difference, after recoding for effect allele, $p = 1.5 \times 10^{-7}$). Restricted to the eight nominal significant SNPs, correlation increased ($r = 0.7$, $p = 0.049$), and the difference became insignificant ($p = 0.68$). Using PS8 instead did not improve the results.

We next included sets of proxy SNPs that are in high LD ($r^2 > 0.5$) with the lead SNP of a CAD locus in our analysis. Of note, this included only 57 SNPs, as two CAD SNPs were tagged by the same proxy SNP. The number of CAD loci that were also significantly associated with PS ($p < 0.05$) increased to 16, suggesting that these CAD loci might also be associated with carotid plaque. To verify this result, we performed a permutation test with 10,000 randomly

combined sets of genotypes and phenotypes. Assuming the data to be Poisson distributed, the empirical p value is 1.03×10^{-8} , which indicates a strong enrichment. We also considered the combined effect of all CAD SNPs by calculating a GRS. The GRS showed a strong association with PS (beta = 0.264, $p = 1.99 \times 10^{-6}$).

3.6. Analysis of known cIMT, IS, and LAS loci for association with plaque score

For four SNPs, association with cIMT [9] was reported. One of them showed nominal significance with PS, and the direction of effects are the same. The locus of this SNP is *APOC1*, which is also reported to be a CAD and plaque prevalence locus (with suggestive evidence) [2,9].

We found no significant association with the eight SNPs of the most recently published stroke GWAS [13]. Dichgans et al. [12] reported 17 SNPs in association with IS or LAS. Three of them showed nominal significance with our PS trait, namely rs1333047 at 9p21 (reported for LAS), rs7937106 at 11q22 (reported for LAS), and rs4792143 at 17p12 (reported for IS). For both LAS SNPs the effect with PS had the same direction as reported. However, the SNP reported for IS has opposite direction of effect in our data. Using best proxy SNPs in high LD ($r^2 > 0.5$) there were six loci significantly associated with PS ($p < 0.05$). A permutation test was performed and resulted in an empirical p -value of 2.44×10^{-4} , indicating a non-random overlap of stroke loci and loci associated with carotid plaque score.

In addition, we calculated the genetic risk score for stroke using the published ORs of Dichgans et al. [12] and tested the association between this GRS and PS, PS8 and plaque burden at each of the four anatomical regions investigated (see methods and Supplementary Table 6). There was no significant association between PS and GRS using all 17 stroke SNPs (beta = 0.0658, $p = 0.20$). Restricting the GRS on the eleven LAS SNPs, the association becomes significant (beta = 0.114, $p = 0.039$). We repeated the analysis with PS8, which resulted in significant association for all 17 SNPs (beta = 0.225, $p = 6.8 \times 10^{-3}$) as well as LAS SNPs only (beta = 0.296, $p = 9.4 \times 10^{-4}$). Interestingly, this association is mainly triggered by plaque in ICA.

3.7. Comparison of LIFE-Heart and LIFE-Adult

Prevalence of carotid plaque is higher in LIFE-Heart. Therefore, we compared the risk allele frequencies (RAF) of our genome-wide significant and suggestive PS hits between LIFE-Heart and LIFE-Adult. The same was done for the RAF of the 58 CAD SNPs and the corresponding GRS. Additionally, we compared CAD cases and controls (noCAD) of LIFE-Heart with respect to these frequencies. Frequencies are shown in Supplementary Table 7.

For the PS hits, three of them had higher RAF in LIFE-Heart compared to LIFE-Adult (see also Table 2). This also applies for the comparison of CAD and noCAD.

For the 58 CAD SNPs we found no enrichment of risk variants in LIFE-Heart, as only 35 SNPs had higher RAF (binomial test, $p = 0.15$). In contrast, comparing CAD and noCAD in LIFE-Heart we observed 39 SNPs with higher RAF in cases (binomial test, $p = 0.012$). Results are displayed in Supplementary Fig. 2.

Similarly, we compared the GRS between our cohorts and between cases and controls of LIFE-Heart. As expected, subjects of LIFE-Adult have smaller risk scores than those of LIFE-Heart (t -test, $p = 0.005$). Controls of LIFE-Heart have smaller GRS than cases (t -test, $p = 0.001$). Of note, scores in LIFE-Adult and LIFE-Heart are significantly associated with PS (LIFE-Adult: beta = 0.22, $p = 8.7 \times 10^{-4}$, LIFE-Heart: beta = 0.36, $p = 3.7 \times 10^{-4}$).

Table 3
Results of previously reported SNPs in our study.

	SNP Infos					Literature $\beta \pm SE/OR$ (95% CI)	Association of lead SNP with PS		Best associated tag SNP			
	Locus	Lead SNP	Nearby gene	Effect/ other	EAF Heart Adult		p value	β	tag SNP	p value	β	
Bis et al. [9]	3p14	rs17045031	<i>LRIG1</i>	A/G	0.032 0.036	-0.297 ± 0.059	4.00×10^{-7}	9.95×10^{-2}	-0.082	rs76405716	5.63×10^{-2}	-0.093
	4q31	rs1878406	<i>EDNRA</i>	T/C	0.141 0.141	0.199 ± 0.029	6.90×10^{-12}	1.86×10^{-2}	0.061	rs6820938	6.37×10^{-4}	0.081
	7q22	rs17398575	<i>PIK3CG</i>	A/G	0.250 0.243	0.162 ± 0.023	2.30×10^{-12}	4.40×10^{-2}	0.043	rs12705390	3.63×10^{-3}	0.065
	19p13	rs6511720	<i>LDLR</i>	T/G	0.086 0.099	-0.178 ± 0.033	1.00×10^{-7}	4.66×10^{-2}	-0.063	rs17242381	8.51×10^{-3}	-0.082
	19q13	rs445925	<i>APOC1</i>	A/G	0.100 0.113	0.82 (0.76, 0.89)	4.00×10^{-6}	7.34×10^{-3}	-0.098	rs7412	1.41×10^{-3}	-0.124
Nikpay et al. [2]	1p32	rs9970807 ^a	<i>PPAP2B</i>	T/C	0.080 0.080	0.88 (0.85, 0.91)	5.00×10^{-14}	2.47×10^{-1}	-0.039	rs72664358	1.49×10^{-2}	-0.060
	1q41	rs67180937	<i>MIA3</i>	G/T	0.743 0.747	1.08 (1.06, 1.11)	1.01×10^{-12}	5.58×10^{-2}	0.040	rs28787398	1.23×10^{-3}	0.065
	2p21	chr2:44074126:D ^a	<i>ABCG5- ABCG8</i>	C/CGT	0.246 0.251	0.94 (0.92, 0.96)	2.60×10^{-8}	2.05×10^{-1}	-0.027	rs7598542	4.31×10^{-2}	-0.045
	2p11	rs7568458	<i>VAMP5-VAMP8- GDCX</i>	A/T	0.445 0.460	1.06 (1.04, 1.08)	3.62×10^{-10}	3.45×10^{-1}	0.017	rs6747828	1.29×10^{-2}	0.050
	4q32	rs72689147 ^a	<i>GUCY1A3</i>	T/G	0.190 0.201	0.93 (0.91, 0.95)	6.07×10^{-9}	4.57×10^{-2}	-0.046	rs11731886	3.60×10^{-2}	-0.045
	6q25	rs55730499 ^{a,b}	<i>SLC22A3-LPAL2- LPA</i>	T/C	0.059 0.058	0.93 (0.92, 0.96)	1.85×10^{-8}	8.88×10^{-2}	0.067	rs10455872	4.66×10^{-2}	0.078
	6q26	rs4252185 ^a	<i>PLG</i>	C/T	0.073 0.070	1.00 (0.98, 1.02)	9.60×10^{-1}	2.23×10^{-1}	0.049	rs10455872	4.66×10^{-2}	0.078
	8q24	rs2954029 ^a	<i>TRIB1</i>	T/A	0.487 0.492	0.96 (0.94, 0.97)	2.61×10^{-6}	1.37×10^{-3}	-0.059	rs2980868	7.18×10^{-4}	-0.062
	9p21	rs2891168	<i>CDKN2B</i>	G/A	0.480 0.468	1.21 (1.19, 1.24)	2.29×10^{-98}	2.31×10^{-8}	0.101	rs9644862	8.73×10^{-12}	0.131
	10p11	rs2487928	<i>KIAA1462</i>	A/G	0.473 0.488	1.06 (1.04, 1.08)	4.41×10^{-11}	4.98×10^{-2}	0.036	rs2505084	5.63×10^{-3}	0.050
	10q24	rs11191416 ^a	<i>CYP17A1-CNNM2- NT5C2</i>	G/T	0.111 0.109	0.93 (0.90, 0.95)	4.65×10^{-9}	5.17×10^{-3}	-0.081	rs284841	1.71×10^{-4}	0.105
	12q21	rs2681472 ^{a,b}	<i>ATP2B1</i>	G/A	0.145 0.149	0.93 (0.92, 0.96)	1.03×10^{-9}	2.20×10^{-2}	0.059	rs4842666	4.57×10^{-3}	0.076
	13q34	rs11838776 ^b	<i>COL4A1/A2</i>	A/G	0.296 0.299	1.04 (1.02, 1.06)	7.13×10^{-5}	1.43×10^{-1}	-0.030	rs9515199	1.75×10^{-2}	0.045
	15q25	rs4468572 ^{a,b}	<i>ADAMTS7</i>	C/T	0.590 0.596	0.93 (0.91, 0.95)	4.52×10^{-9}	2.84×10^{-1}	0.021	rs12899452	4.85×10^{-2}	0.039
	17p11	rs12936587 ^b	<i>RAI1-PEMT- RASD1</i>	A/G	0.434 0.433	1.04 (1.02, 1.06)	1.84×10^{-5}	4.21×10^{-2}	-0.038	rs58251514	1.72×10^{-2}	-0.046
19p13	rs56289821 ^a	<i>LDLR</i>	A/G	0.087 0.100	0.88 (0.85, 0.90)	4.44×10^{-15}	6.91×10^{-2}	-0.058	rs17242381	8.51×10^{-3}	-0.082	
19q13	rs4420638	<i>APOE- APOC1</i>	G/A	0.147 0.170	1.10 (1.07, 1.13)	7.07×10^{-11}	9.27×10^{-3}	0.069	rs429358	7.63×10^{-4}	0.104	

SNPs of Bis et al. [9] (trait plaque prevalence) and Nikpay et al. [2] (trait CAD) reaching at least nominal significance in our study (trait PS). Tag-SNP denotes the best SNP (minimal p value) in LD with the reported top-SNP ($r^2 > 0.5$). Some SNPs were reversely coded in the reference (marked with ^a). For better comparison, we transformed their OR into its inverse. After transformation, six CAD-SNPs showed a different direction of effect (marked with ^b) when compared to our PS trait. For further information, see Supplementary Tables 4 and 5. Significant direct associations of literature SNPs with PS are shown in bold.

^a SNPs were reversely coded in the reference.

^b SNPs with a different direction of effect.

4. Discussion

Carotid artery plaque was proposed as a marker to support prediction of prevalent and incident coronary events [6–8]. However, assessing plaque status in sufficiently large cohorts is demanding and only one meta-GWAS was published for this trait so far [9]. Moreover, there is no generally accepted gold standard of carotid plaque assessment, resulting in large heterogeneity of the trait between studies. Here, we performed a genetic meta-analysis of two large epidemiologic studies using the same operating procedures of plaque assessment, thereby guaranteeing a high degree of comparability and high intra- and inter-rater reliability as shown by our feasibility study. We also propose to consider the number of affected segments as a more refined trait. We have previously shown [8] that a plaque score considering the number of affected

segments is a better predictor for obstructive CAD than plaque prevalence alone [9]. For our analyses, we relied on a plaque score of the segments CCA and bulb. A score also including ICA and ECA was considered. Although effect sizes appeared to be stronger compared to PS, the higher percentage of missingness resulted in inferior power. Nevertheless, our study is the first GWAS exploiting plaque score as primary phenotype and resulted in the identification of two loci with genome-wide significance not yet identified with plaque prevalence in genome-wide studies.

The strongest association was observed for SNPs at the 9p21 locus, the most widely replicated locus of CAD. Our lead SNP is in high LD with the lead SNP of this locus. The underlying pathomechanism of this locus is subject of intense research and evidence was collected that the non-coding RNAs of *CDKN2B-AS1* expressed at this locus play a key role [38–40]. The locus is known

to be associated with several other vascular phenotypes and cardiac events [41,42].

The situation for the second strongest locus at chromosome 10 is less obvious. Our top-hit is in an intron of *SFXN2* and the expression of this gene is strongly controlled by the SNP in several tissues. Thus, *SFXN2* is the most plausible candidate here. However, our lead SNP is in low LD with the CAD lead SNP (rs12413409) [33]. While the latter shows only marginal association with PS ($p = 0.01$), proxy SNPs considered by Nikpay et al. [2] (rs11191416) and us (rs284841) showed some association ($p = 5.17 \times 10^{-3}$ and $p = 1.71 \times 10^{-4}$, for rs11191416 and rs284841, respectively). The SNP rs284841 is an intron modifier in *WBP1L*. The same applies for rs9663711 showing suggestive significance in our data. Therefore, *WBP1L* is another plausible candidate gene for this locus.

Both top-SNPs of our genome-wide significant loci are not described in the study of Bis et al. [9] Conversely, we aimed at replicating their loci in our data. It turned out that both genome-wide significant SNPs reported in Bis et al. showed at least nominal significance with the same direction of effect in our data. We further found supportive evidence, i.e. same direction of effect, for the other three SNPs reported as suggestive. Two of them showed nominal significance. This analysis shows that further studies and meta-analyses are required to underpin associations with carotid plaque phenotypes.

Another interesting question is to which extent there is a common genetic origin of different atherosclerotic manifestations. Although the correlation of carotid plaque and cardiovascular lesion is well established [6–8], it is not clear how this translates into common genetic patho-mechanisms. Therefore, we present association statistics of reported top-SNPs of CAD loci with our phenotype, which takes the number of plaques into account. Although only one of the SNPs showed genome-wide significance in our data, we observed an enrichment of concordant effects with at least nominal significance. Effect sizes of CAD and PS are positively correlated. This strongly supports our hypotheses of common genetic mechanisms underlying the development of atherosclerosis in carotid and coronary vessels. In line with this, the strongest CAD locus 9p21 also showed strongest association with our plaque score. The GRS of CAD loci was also associated with PS.

In analogy to the analysis of CAD loci, we investigated the association of cIMT and stroke loci with our phenotype. Regarding cIMT, one of the four reported loci was also associated with PS (*APOC1*). Regarding stroke, six of the 17 loci were associated which was significantly more than expected by random. Interestingly, the genetic risk score of stroke variants showed stronger association with PS8 than PS, which was mostly triggered by a strong association of LAS loci with ICA. This is in well agreement of the pathophysiology of LAS, i.e. plaques in ICA are required for the definition of LAS [14].

In conclusion, we propose carotid plaque score as a valuable trait of genetic association analysis. By performing the first GWAS of this phenotype, we identified two loci of genetic association. While the 9p21 locus is robustly associated with carotid artery plaque score, our hit at chromosome 10 requires further replication in independent cohorts as well as identification of the causal gene. Using our plaque score, we could replicate four hits associated with carotid plaque prevalence. Comparisons of plaque score and CAD associations strongly suggest that there is a larger overlap of genetic causes of the two atherosclerotic manifestations. Likewise, we observed an overlap with genetic loci associated with stroke. This overlap becomes stronger for LAS loci analyzed for genetic associations with plaque scores comprising ICA. Larger studies and meta-analyses are required to confine this overlap and possible differences in genetic patho-mechanisms.

Conflict of interest

The authors declared they do not have anything to disclose regarding conflict of interest with respect to this manuscript.

Financial support

LIFE-Heart and LIFE-Adult are funded by the Leipzig Research Center for Civilization Diseases (LIFE). LIFE is an organizational unit affiliated to the Medical Faculty of the University of Leipzig. LIFE is funded by means of the European Union, by the European Regional Development Fund (ERDF) and by funds of the Free State of Saxony within the framework of the excellence initiative.

Author contributions

JP: data analysis, manuscript writing, RB: genotyping, contribution to discussion, FB, AT: carotid plaque phenotyping, KH: data analysis, HK: data analysis, GS: study design of LIFE-Heart, LMH: genotyping, DT: study design of LIFE-Heart, genotyping, ML: design and conduction of LIFE-Adult, JT: design and conduction of LIFE-Heart, MS: study design, manuscript writing. All authors read and approved the final manuscript.

Acknowledgments

We thank the participants of LIFE-Adult and the patients of LIFE-Heart very much for their time and blood samples. We thank Kerstin Wirkner very much for running the LIFE study center. We thank Sylvia Henger very much for data quality control. We thank Kay Olscher and Annegret Unger very much for technical assistance.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.atherosclerosis.2017.02.018>

References

- [1] T. Kessler, J. Erdmann, H. Schunkert, Genetics of coronary artery disease and myocardial infarction—2013, *Curr. Cardiol. Rep.* 15 (2013) 368, <http://dx.doi.org/10.1007/s11886-013-0368-0>.
- [2] M. Nikpay, A. Goel, H.-H. Won, L.M. Hall, et al., A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease, *Nat. Genet.* 47 (2015) 1121–1130, <http://dx.doi.org/10.1038/ng.3396>.
- [3] M.W. Lorenz, H.S. Markus, M.L. Bots, M. Rosvall, et al., Prediction of clinical cardiovascular events with carotid intima-media thickness: a systematic review and meta-analysis, *Circulation* 115 (2007) 459–467, <http://dx.doi.org/10.1161/CIRCULATIONAHA.106.628875>.
- [4] M.L. Bots, K. Sutton-Tyrrell, Lessons from the past and promises for the future for carotid intima-media thickness, *J. Am. Coll. Cardiol.* 60 (2012) 1599–1604, <http://dx.doi.org/10.1016/j.jacc.2011.12.061>.
- [5] H.M. den Ruijter, S.A.E. Peters, T.J. Anderson, A.R. Britton, et al., Common carotid intima-media thickness measurements in cardiovascular risk prediction: a meta-analysis, *JAMA* 308 (2012) 796–803, <http://dx.doi.org/10.1001/jama.2012.9630>.
- [6] H. Sillesen, P. Muntendam, A. Adourian, R. Entrekina, et al., Carotid plaque burden as a measure of subclinical atherosclerosis: comparison with other tests for subclinical arterial disease in the High Risk Plaque Biomage study, *JACC. Cardiovasc. imaging* 5 (2012) 681–689, <http://dx.doi.org/10.1016/j.jcmg.2012.03.013>.
- [7] Y. Inaba, J.A. Chen, S.R. Bergmann, Carotid plaque, compared with carotid intima-media thickness, more accurately predicts coronary artery disease events: a meta-analysis, *Atherosclerosis* 220 (2012) 128–133, <http://dx.doi.org/10.1016/j.atherosclerosis.2011.06.044>.
- [8] A. Weissgerber, M. Scholz, A. Teren, M. Sandri, et al., The value of noncoronary atherosclerosis for identifying coronary artery disease: results of the Leipzig LIFE heart study, *Clin. Res. Cardiol. official J. Ger. Cardiac Soc.* 105 (2016) 172–181, <http://dx.doi.org/10.1007/s00392-015-0900-x>.
- [9] J.C. Bis, M. Kavousi, N. Franceschini, A. Isaacs, et al., Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque, *Nat. Genet.* 43 (2011) 940–947, <http://dx.doi.org/10.1038/ng.920>.

- [10] M. den Hoed, R.J. Strawbridge, P. Almgren, S. Gustafsson, et al., GWAS-identified loci for coronary heart disease are associated with intima-media thickness and plaque presence at the carotid artery bulb, *Atherosclerosis* 239 (2015) 304–310, <http://dx.doi.org/10.1016/j.atherosclerosis.2015.01.032>.
- [11] D. Della-Morte, L. Wang, A. Beecham, S.H. Blanton, et al., Novel genetic variants modify the effect of smoking on carotid plaque burden in Hispanics, *J. neurological Sci.* 344 (2014) 27–31, <http://dx.doi.org/10.1016/j.jns.2014.06.006>.
- [12] M. Dichgans, R. Malik, I.R. König, J. Rosand, et al., Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants, *Stroke; a J. Cereb. circulation* 45 (2014) 24–36, <http://dx.doi.org/10.1161/STROKEAHA.113.002707>.
- [13] NINDS Stroke Genetics Network (SIGN), International Stroke Genetics Consortium (ISGC), Loci associated with ischaemic stroke and its subtypes (SIGN): a genome-wide association study, *Lancet Neurology* 15 (2016) 174–184, [http://dx.doi.org/10.1016/S1474-4422\(15\)00338-5](http://dx.doi.org/10.1016/S1474-4422(15)00338-5).
- [14] H.P.J.R. Adams, B.H. Bendixen, L.J. Kappelle, J. Biller, et al., Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in acute stroke treatment, *Stroke* 24 (1993) 35–41.
- [15] M. Loeffler, C. Engel, P. Ahnert, D. Alfermann, et al., The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany, *BMC public health* 15 (2015) 691, <http://dx.doi.org/10.1186/s12889-015-1983-z>.
- [16] F. Beutner, D. Teupser, S. Gielen, L.M. Holdt, et al., Rationale and design of the Leipzig (LIFE) Heart Study: phenotyping and cardiovascular characteristics of patients with coronary artery disease, *PLoS one* 6 (2011) e29070, <http://dx.doi.org/10.1371/journal.pone.0029070>.
- [17] J.H. Stein, C.E. Korcarz, R.T. Hurst, E. Lonn, et al., Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American society of echocardiography carotid intima-media thickness task force. Endorsed by the Society for Vascular Medicine, *J. Am. Soc. Echocardiogr. official Publ. Am. Soc. Echocardiogr.* 21 (2008) 93–111, <http://dx.doi.org/10.1016/j.jecho.2007.11.011> quiz 189–90.
- [18] Affymetrix, I., *Axiom Genotyping Solution Data Analysis Guide*.
- [19] C.C. Laurie, K.F. Doheny, D.B. Mirel, E.W. Pugh, et al., Quality control and quality assurance in genotypic data for genome-wide association studies, *Genet. Epidemiol.* 34 (2010) 591–602, <http://dx.doi.org/10.1002/gepi.20516>.
- [20] J. Wang, An estimator for pairwise relatedness using molecular markers, *Genetics* 160 (2002) 1203–1215.
- [21] N. Patterson, A.L. Price, D. Reich, Population structure and eigenanalysis, *PLoS Genet.* 2 (2006) e190, <http://dx.doi.org/10.1371/journal.pgen.0020190>.
- [22] C.C. Chang, C.C. Chow, L.C. Tellier, S. Vattikuti, et al., Second-generation PLINK: rising to the challenge of larger and richer datasets, *GigaScience* 4 (2015) 7, <http://dx.doi.org/10.1186/s13742-015-0047-8>.
- [23] Shaun Purcell, C. C., PLINK.
- [24] I.R. König, C. Loley, J. Erdmann, A. Ziegler, How to include chromosome X in your genome-wide association study, *Genet. Epidemiol.* 38 (2014) 97–103, <http://dx.doi.org/10.1002/gepi.21782>.
- [25] A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, et al., A global reference for human genetic variation, *Nature* 526 (2015) 68–74, <http://dx.doi.org/10.1038/nature15393>.
- [26] O. Delaneau, B. Howie, A.J. Cox, J.-F. Zagury, et al., Haplotype estimation using sequencing reads, *Am. J. Hum. Genet.* 93 (2013) 687–696, <http://dx.doi.org/10.1016/j.ajhg.2013.09.002>.
- [27] B.N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies, *PLoS Genet.* 5 (2009) e1000529, <http://dx.doi.org/10.1371/journal.pgen.1000529>.
- [28] D. Welter, J. MacArthur, J. Morales, T. Burdett, et al., The NHGRI GWAS Catalog, a curated resource of SNP-trait associations, *Nucleic acids Res.* 42 (2014) D1001–D1006, <http://dx.doi.org/10.1093/nar/gkt1229>.
- [29] H. Kirsten, H. Al-Hasani, L. Holdt, A. Gross, et al., Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding lincRNAs, *Hum. Mol. Genet.* 24 (2015) 4746–4763, <http://dx.doi.org/10.1093/hmg/ddv194>.
- [30] G. Yu, L.-G. Wang, G.-R. Yan, Q.-Y. He, DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis, *Bioinforma. Oxf. Engl.* 31 (2015) 608–609, <http://dx.doi.org/10.1093/bioinformatics/btu684>.
- [31] M. Kircher, D.M. Witten, P. Jain, B.J. O’Roak, et al., A general framework for estimating the relative pathogenicity of human genetic variants, *Nat. Genet.* 46 (2014) 310–315, <http://dx.doi.org/10.1038/ng.2892>.
- [32] M.H. Geisel, S. Coassin, N. Hessler, M. Bauer, et al., Update of the effect estimates for common variants associated with carotid intima media thickness within four independent samples: the Bonn IMT Family Study, the Heinz Nixdorf Recall Study, the SAPHIR Study and the Bruneck Study, *Atherosclerosis* 249 (2016) 83–87, <http://dx.doi.org/10.1016/j.atherosclerosis.2016.03.042>.
- [33] H. Schunkert, I.R. König, S. Kathiresan, M.P. Reilly, et al., Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease, *Nat. Genet.* 43 (2011) 333–338, <http://dx.doi.org/10.1038/ng.784>.
- [34] Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans, *Sci. (New York, N.Y.)* 348 (2015) 648–660, <http://dx.doi.org/10.1126/science.1262110>.
- [35] H.-J. Westra, M.J. Peters, T. Esko, H. Yaghootkar, et al., Systematic identification of trans eQTLs as putative drivers of known disease associations, *Nat. Genet.* 45 (2013) 1238–1243, <http://dx.doi.org/10.1038/ng.2756>.
- [36] A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease, *Nat. Genet.* 43 (2011) 339–344, <http://dx.doi.org/10.1038/ng.782>.
- [37] S. Chinn, A simple method for converting an odds ratio to effect size for use in meta-analysis, *Statistics Med.* 19 (2000) 3127–3131.
- [38] L.M. Holdt, S. Hoffmann, K. Sass, D. Langenberger, et al., Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks, *PLoS Genet.* 9 (2013) e1003588, <http://dx.doi.org/10.1371/journal.pgen.1003588>.
- [39] L.M. Holdt, K. Sass, G. Gabel, H. Bergert, et al., Expression of Chr9p21 genes CDKN2B (p15^{INK4b}), CDKN2A (p16^{INK4a}), p14^(ARF) and MTAP in human atherosclerotic plaque, *Atherosclerosis* 214 (2011) 264–270, <http://dx.doi.org/10.1016/j.atherosclerosis.2010.06.029>.
- [40] L.M. Holdt, A. Stahringer, K. Sass, G. Pichler, et al., Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans, *Nat. Commun.* 7 (2016) 12429, <http://dx.doi.org/10.1038/ncomms12429>.
- [41] L.M. Holdt, F. Beutner, M. Scholz, S. Gielen, et al., ANRIL expression is associated with atherosclerosis risk at chromosome 9p21, *Arteriosclerosis, Thrombosis, Vasc. Biol.* 30 (2010) 620–627, <http://dx.doi.org/10.1161/ATVBAHA.109.196832>.
- [42] L.M. Holdt, D. Teupser, Recent studies of the human chromosome 9p21 locus, which is associated with atherosclerosis in human populations, *Arteriosclerosis, Thrombosis, Vasc. Biol.* 32 (2012) 196–206, <http://dx.doi.org/10.1161/ATVBAHA.111.232678>.

2.2 „Genetic regulation of PCSK9 plasma levels and its impact on atherosclerotic vascular disease phenotypes“

Pott et al. [44], DOI: <https://doi.org/10.1161/CIRCGEN.117.001992>

Eingereicht: 20. Oktober 2017; Akzeptiert: 5. März 2018

Die Proproteinkonvertase Subtilisin/Kexin Typ 9 (PCSK9) ist ein Schlüsselfaktor im Lipidmetabolismus. Die Inhibition von PCSK9 stellt eine neue Therapiemöglichkeit für Patienten mit Hypercholesterinämie dar. Obwohl bereits eine Reduktion von kardiovaskulären Ereignissen durch PCSK9-Hemmer beschrieben wurde, ist die Rolle von Plasma-PCSK9 als Biomarker noch nicht eindeutig geklärt. Das Ziel dieser Publikation war es, genetische Varianten zu identifizieren, die einen signifikanten Einfluss auf PCSK9-Plasmaspiegel haben, und hinsichtlich kausaler Effekte von PCSK9 auf verschiedene Formen der Atherosklerose zu testen.

Dazu wurde eine GWAS im Verdachtskollektiv der LIFE-Heart Studie durchgeführt (n=3290). Da eine Lipidtherapie mit Statinen einen signifikanten Einfluss auf das PCSK9-Spiegel hat, wurde zur Überprüfung der Ergebnisse eine GWAS in der statin-freien Subgruppe der Patienten ausgeführt (n=2022).

In der GWAS konnten Assoziationen von vier unabhängigen Varianten (paarweises LD $r^2 < 0.1$) im *PCSK9* Gen-Lokus mit Plasma-PCSK9 detektiert werden. In der statin-freien Subgruppe waren diese Varianten ebenfalls assoziiert, jedoch erreichten hier nur drei der SNPs genomweite Signifikanz aufgrund der Fallzahlreduktion. Der Top-Hit war die bekannte Missense-Mutation R46L (rs11591147, $p = 1.94 \times 10^{-17}$), die eine Reduktion der Phosphorylierung an Ser47 bewirkt. Da die Phosphorylierung vor Abbau schützen würde, wird die Variante mit Lys46 verstärkt degradiert, was in einem niedrigeren PCSK9-Plasmaspiegel resultiert [45, 46].

Neben dem starken Hit im PCSK9 Lokus konnten ein weiterer SNP mit genomweiter Signifikanz und sieben mit suggestiver Signifikanz detektiert werden. Die Validierung dieser Loci und Analyse möglicher funktioneller Gene steht noch aus.

Anschließend wurden die vier detektierten Varianten auf Assoziation mit Formen der Atherosklerose in der Karotis getestet. Keiner der vier SNPs war mit PS assoziiert, und nur einer nominal mit cIMT (rs2479409, $p = 0.028$).

Abschließend wurde eine Mendelsche Randomisierung mit den vier Varianten als IVs durchgeführt. Hier wurden kausale Effekte von PCSK9 auf Atherosklerose in koronaren Gefäßen (CAD, NVD50), Karotisgefäßen (PS, cIMT) und in der Peripherie (ABI, PAD) analysiert. Kausale Effekte auf die Koronar- und Karotis-Gefäße konnten gezeigt werden, wohingegen kein solcher Effekt auf periphere Gefäße nachweisbar war.

Circulation: Genomic and Precision Medicine

ORIGINAL ARTICLE

Genetic Regulation of PCSK9 (Proprotein Convertase Subtilisin/Kexin Type 9) Plasma Levels and Its Impact on Atherosclerotic Vascular Disease Phenotypes

See Editorial by Paquette and Baass

BACKGROUND: Inhibition of PCSK9 (proprotein convertase subtilisin/kexin type 9) is a novel strategy to treat hypercholesterolemia and reduce cardiovascular events. However, the potential role of circulating plasma PCSK9 concentrations as a diagnostic and predictive biomarker remains uncertain as of now. Here, we aimed to identify genetic variants associated with plasma PCSK9 and investigate possible causal effects on atherosclerotic vascular disease phenotypes.

METHODS: We performed the first genome-wide association study of plasma PCSK9 levels in a cohort of suspected and confirmed coronary artery disease (LIFE-Heart; n=3290).

RESULTS: Several independent variants at the *PCSK9* gene locus were associated with circulating PCSK9 levels at genome-wide significance (lead SNP rs11591147, PCSK9-R46L; $P=1.94 \times 10^{-17}$). We discovered 4 independent *PCSK9* SNPs explaining 4.4% of the variance of plasma PCSK9. In addition, we identified a genome-wide significant locus at chromosome 7p22.1 (rs6957201; $P=7.01 \times 10^{-9}$) and 7 suggestive hits ($P < 1 \times 10^{-6}$). Using MR (Mendelian Randomization), we detected significant causal effects of circulating PCSK9 on coronary artery disease status and severity, carotid plaques, and intima-media thickness.

CONCLUSIONS: Variants at the *PCSK9* gene locus seem to be the major genetic determinants of plasma PCSK9 levels with 4 independent variants at the *PCSK9* gene locus expressing allelic heterogeneity. The detected MR estimates support the hypothesis of a causal effect of PCSK9 on coronary artery disease and other vascular phenotypes. Other observed genetic associations for PCSK9 require validation in independent cohorts.

CLINICAL TRIAL REGISTRATION: URL: <http://www.clinicaltrials.gov>. Unique Identifier: NCT00497887.

Janne Pott, MS
Valentin Schlegel
Andrej Teren, MD
Katrin Horn, MS
Holger Kirsten, PhD
Christina Bluecher, MS
Juergen Kratzsch, PhD
Markus Loeffler, MD, PhD
Joachim Thiery, MD
Ralph Burkhardt, MD*
Markus Scholz, PhD*

*Drs Burkhardt and Scholz contributed equally to this work.

Key Words: atherosclerosis ■ genome-wide association study

© 2018 American Heart Association, Inc.

<http://circgenetics.ahajournals.org>

CLINICAL PERSPECTIVE

PCSK9 (proprotein convertase subtilisin/kexin type 9) plays an essential role in regulating lipid metabolism. Functional inhibition of PCSK9 is a novel treatment strategy for hyperlipidemia. In our genome-wide association study, we identified 4 independent genetic variants at the *PCSK9* locus at genome-wide significance, explaining 4.4% of the total variance of PCSK9. In an MR study (Mendelian randomization), we detected causal effects of PCSK9 plasma levels on coronary artery disease, number of coronary vessels with stenosis, and carotid artery plaques but not peripheral artery disease. This increases confidence that PCSK9 is a valid target not only to treat hyperlipidemia but also to reduce risk or severity of atherosclerotic vascular disease in carotid and coronary arteries.

Elevated plasma levels of LDL-C (low-density lipoprotein cholesterol) are a major modifiable risk factor for atherosclerotic cardiovascular disease. Therapeutic interventions facilitating the hepatic uptake of LDL via the LDLR (LDL receptor) pathway are proven to reduce cardiovascular events.¹ PCSK9 (proprotein convertase subtilisin/kexin type 9) plays an essential role in lipid metabolism as a key regulator of plasma LDL-C.² The protein is mainly produced and secreted by the liver. It binds LDLR on the cell surface of hepatocytes and targets them for lysosomal degradation, resulting in increased LDL-C. Rare gain-of-function mutations in *PCSK9* were identified to cause autosomal dominant hypercholesterolemia, whereas loss-of-function mutations in *PCSK9* were found to be associated with low LDL-C and reduced risk of coronary artery disease (CAD).^{3,4} Moreover, genome-wide association analysis showed that common variants in the *PCSK9* gene locus are associated with LDL-C and the risk for CAD.^{5,6} Therefore, inhibiting PCSK9 function has emerged as a promising strategy to treat hypercholesterolemia. Indeed, 2 monoclonal antibodies against PCSK9 (Alirocumab and Evolocumab) have been approved^{7,8} and shown to markedly reduce LDL-C in a wide range of patients. The use of evolocumab was also associated with a significant reduction of future cardiovascular events in a large-scale outcome trial.⁸

Recent studies indicate that the functional effects of PCSK9 may not be limited to the LDLR pathway. PCSK9 was implicated to affect several other atherogenic risk factors, such as the metabolism of triglyceride-rich lipoproteins, degradation of the very-low-density-lipoprotein receptor, inflammatory response, and glucose metabolism.² In this context, circulating PCSK9 concentrations have gained interest as a potential biomarker

for risk stratification. Current findings on the relation between PCSK9 plasma levels and cardiovascular disease are equivocal, but a meta-analysis of 7 studies reported that subjects with high PCSK9 levels have a 23% higher risk for total cardiovascular events when compared with patients with low PCSK9 levels.⁹

Although considerable variation in circulating PCSK9 levels has been noted in different cohort studies,⁹⁻¹¹ information on the genetic factors that contribute to this variability is limited. A recent GWAS (Genome-Wide Association Study) of PCSK9 plasma levels failed to detect genome-wide significant variants (genome-wide threshold $P < 5 \times 10^{-8}$) in 2 cohorts of healthy, middle-aged Swedes, which might be because of limited statistical power given the moderate sample size ($N=1215$).¹⁰ In addition, causal links between PCSK9 plasma levels and atherosclerotic vascular disease (ASVD) phenotypes have not been studied using genetic instruments to date.

In the present study, we performed genome-wide association analysis for circulating PCSK9 levels in a large cohort of patients receiving coronary angiography for suspected CAD ($N=3290$; LIFE-Heart study). Further, we analyzed allelic heterogeneity of the *PCSK9* gene locus and applied MR (Mendelian Randomization) to identify causal effects of circulating PCSK9 levels on various atherosclerotic disease phenotypes including additional data of the large population-based LIFE-Adult study.

METHODS

The data that support the findings of this study are available from the LIFE Research Center for Civilization diseases on qualified request. Requests for access to more detailed summary statistics, replication results, and analytic methods will be considered by the authors.

Cohort Descriptions

LIFE-Heart is an observational study of patients collected at the Heart Center of Leipzig, Germany. A total of ≈ 7000 patients were recruited with either suspected stable CAD or myocardial infarction. Study design and a detailed description of patients can be found elsewhere.¹² For the present analysis, we only included patients with suspected stable CAD subjected to coronary angiography. PCSK9 plasma levels and genetic data were available for $N=3358$ of these patients.

LIFE-Adult is a population-based cohort study of adult residents of the city of Leipzig, Germany. A total of ≈ 10000 participants have been recruited and characterized including subclinical atherosclerosis phenotypes.¹³ For a total of 4985 LIFE-Adult participants, genetic data were available. These samples were used to improve power of the MR analysis.

Both LIFE-Heart and LIFE-Adult meet the ethical standards of the Declaration of Helsinki and were approved by the Ethics Committee of the Medical Faculty of the University Leipzig, Germany (LIFE-Heart: Reg. No. 276-2005; LIFE-Adult: Reg. No. 263-2009-14122009). Written informed consents including agreement with genetic analyses were obtained from all patients of LIFE-Heart and participants of LIFE-Adult.

Assessment of Atherosclerotic Phenotypes

We analyzed 6 ASVD phenotypes at 3 locations: coronary arteries: CAD status and number of coronary vessels with stenosis (NVD50); carotid arteries: intima-media thickness (cIMT), and plaque score (PS; with values between 0 and 4 counting the number of vessels with plaque); and peripheral arteries: ankle-brachial index (ABI) and peripheral artery disease (PAD) status (defined as $ABI < 0.9$). Coronary phenotypes are only available for LIFE-Heart. Details of the measurements are provided in the [Data Supplement](#).

Blood Lipid and PCSK9 Measurement

Venous blood samples were taken before coronary angiography in patients of the LIFE-Heart study. Laboratory measurements were performed on the same day using an automated Roche Cobas 8000 Clinical Chemistry analyzer (Roche Diagnostics, Mannheim, Germany). Total cholesterol, LDL-C, and HDL-C (high-density lipoprotein cholesterol) were determined by homogeneous enzymatic colorimetric assays. ApoA1 and ApoB (apolipoproteins) were determined with Roche immunoturbidimetric assays. Total PCSK9 levels were analyzed in serum samples (previously stored at -80°C) using a commercial assay (Quantikine Human PCSK9 immunoassay; R&D Systems). Further details of PCSK9 measurement are given in the [Data Supplement](#).

SNP Genotyping and Imputation

LIFE-Heart samples were genotyped using Affymetrix Axiom CEU1 or Affymetrix Axiom CADLIFE genome-wide SNP arrays. The latter essentially contains Axiom CEU as genome-wide backbone and an additional custom content of $\approx 62\,500$ SNPs from known CAD loci. Genotype calling was performed using Affymetrix Power Tools (v1.17.0 for Axiom CADLIFE and v1.16.1 for Axiom CEU) with their latest libraries (Axiom CADLIFE1, release 3 and Axiom Genome-Wide CEU 1 Array Plate, Analysis Files, release 6, respectively). LIFE-Adult samples were genotyped with Affymetrix Axiom CEU1 SNP array, and calling relied on Affymetrix Power Tools (version 1.17.0) with the same library as LIFE-Heart.

An in-house pipeline performed SNP and sample quality control. Details can be found in the [Data Supplement](#). For LIFE-Heart, 5700 (5688) samples and 504 593 (12 715) SNPs fulfilled all quality criteria for autosomal (X-chromosomal, respectively) analyses. For LIFE-Adult, 4985 (4978) samples and 532 875 (13 554) SNPs fulfilled all quality criteria.

Imputation was performed using 1000 Genomes Phase 3, Version 5¹⁴ (2015) as reference, SHAPEIT¹⁵ (version v2.r837) for phasing, and IMPUTE2¹⁶ (version 2.3.2) for genotype estimation. For chromosome X, the same reference and software was used.

SNPs with info score < 0.5 or minor allele frequency $< 1\%$ were filtered. A total of 9 882 017 SNPs were used for genome-wide analyses, 9 579 329 for autosomal, and 302 688 for X-chromosomal analysis.

Statistical Analysis

Genome-Wide Association Analysis

PCSK9 and BMI were log-transformed for all analyses.

Before any GWAS, we controlled several parameters for association with plasma PCSK9. Only sex, age, smoking, and

statin therapy were significantly associated in a multivariate model. In addition, we found a significant sex–age interaction ($P=0.0041$), but decided against an inclusion of this covariate for GWA because it was expected that this interaction can hardly influence the SNP effect. Nevertheless, for the purpose of validation, we checked all genome-wide and suggestive GWA hits with this model and a model containing no covariates (univariate analysis). No relevant differences were found.

We performed 2 genome-wide association analyses for PCSK9 levels in LIFE-Heart: the first GWAS included all patients ($N=3290$ patients with complete covariates), and the second was restricted to those without statin therapy ($N=2022$ patients with complete covariates, denoted as statin-free subset). Both analyses were executed with SNPTTEST¹⁷ (version 2.5.2) using the additive frequentist model and expected genotype counts. We adjusted for sex, age, and current smoking in both analyses, and additionally, for lipid therapy in the first analysis. The threshold for genome-wide significance was set to $P < 5 \times 10^{-8}$. Associations with $P < 1 \times 10^{-6}$ were considered as suggestive and presented as list of top SNPs.

X-chromosomal SNPs were analyzed assuming total X inactivation, that is, the gene doses of females are halved according to König et al.¹⁸ To consider possible alternatives, we tested X-chromosomal top hits with a model without X inactivation and analyzed sex–SNP interactions. No such interactions were found at genome-wide significance.

Linkage disequilibrium between markers was calculated using data from 1000 Genomes Phase 3, Version 5 (2015)¹⁴ for European samples. Priority pruning of the top list was performed as follows: first, variants of the list of top SNPs which are in linkage disequilibrium with an association of higher significance were considered as tagged by the other variant if $r^2 \geq 0.5$. Then, to analyze the PCSK9 locus in detail, we used a stricter threshold of $r^2 \geq 0.1$.

A comprehensive annotation was applied to all SNPs of our top list using the following bioinformatics resources: Ensemble,¹⁹ GWAS catalogue,²⁰ expression quantitative trait loci data,^{21,22} and pathways from KEGG, GO, DOSE,²³ and Reactome.²⁴ Deleteriousness scores for nonsynonymous coding SNPs were calculated according to Kircher et al.²⁵ and Bendl et al.²⁶ Further details are summarized in the [Data Supplement](#).

Relation of PCSK9 and ASVD

Vascular phenotypes (CAD, NVD50, PS, cIMT, PAD, and ABI) were tested for association with PCSK9 levels and SNPs discovered in our GWAS. CAD and PAD status were analyzed by logistic regression, the ordered categories of NVD50 and PS by proportional odds logistic regression, and cIMT and ABI by linear regression.

MR Analysis

We performed an MR study using 4 independent PCSK9 variants with pairwise $r^2 < 0.1$ as instrumental variables (IVs), plasma PCSK9 levels as exposure, and the vascular phenotypes as outcomes. Variants were assumed to satisfy MR assumptions for IVs as explained in the [Data Supplement](#) (the relevant directed acyclic graph is given in Figure I in the [Data Supplement](#)).

LIFE-Heart was used to calculate the regression coefficients of IV on plasma PCSK9 and of IV on CAD and NVD50. LIFE-Heart and LIFE-Adult were used to calculate regression

coefficients of IV on PS, cIMT mean, PAD, and ABI. We performed a fixed-effect model meta-analysis to generate combined estimates of IV on outcome. We then calculated the combined inverse-variance weighted causal estimate $\hat{\beta}_{IV}$ for each outcome.²⁷ To test whether our positive causal results are confounded by pleiotropy, we applied the modified Q test.²⁸ Test was negative for all outcomes.

If not stated otherwise, statistical analyses were performed with the software R.²⁹ For all analyses other than genome-wide SNP association, we applied a (nonadjusted) significance threshold of 5%, that is, our results are exploratory rather than confirmative.

RESULTS

Relationships of Cardiovascular Risk Factors and Plasma Lipids to Circulating PCSK9 Levels

The characteristics of the LIFE-Heart study population are summarized in Table 1. Plasma PCSK9 levels were associated with sex, age, BMI, diabetes mellitus, smoking status, hypertension, and statins in univariate analysis (Table I A in the [Data Supplement](#)). In a multivariate analysis, only sex, age, smoking status, and statins remained significant. Similar effect sizes of sex, age, and smoking were observed in the statin-free subset (Table I B in the [Data Supplement](#)). Male sex and age were associated with lower PCSK9 levels, that is, these factors have an opposite effect on PCSK9 than on CAD risk, whereas smoking and statins increased PCSK9. We observed a significant and qualitative interaction between sex and age (Figure II in the [Data Supplement](#); Table IC and ID in the [Data Supplement](#)).

PCSK9 levels were positively correlated with total cholesterol (Spearman $\rho=0.16$; $P=2.06\times 10^{-19}$), LDL-C ($\rho=0.10$; $P=8.66\times 10^{-9}$), ApoB ($\rho=0.14$; $P=1.19\times 10^{-16}$), and ApoA1 ($\rho=0.11$; $P=3.65\times 10^{-10}$) but not with HDL-C after controlling for sex, age, smoking, and statins (Table II in the [Data Supplement](#)). These correlations were stronger in the statin-free subset (Figure III in the [Data Supplement](#)).

Genome-Wide Association Analysis of PCSK9 Plasma Levels

To identify genetic variants associated with plasma PCSK9 levels, we performed GWA of all 3290 LIFE-Heart samples and then evaluated the consistency of the results in the statin-free subset ($n=2022$). In the primary GWAS, we identified a total of 31 variants (5 variants after pruning) with genome-wide significance. No inflation of test statistics was observed ($\lambda=1.006$). Manhattan plots and QQ plot are displayed in Figure 1 (statin-free subset analysis is given in Figure IV in the [Data Supplement](#)). Summary statistics and SNP annotations are displayed in Table III A in the [Data Supplement](#)

(statin-free subset is given in Table III B in the [Data Supplement](#)).

Genetic Variants at the PCSK9 Locus

The strongest associations with PCSK9 plasma level were observed on chromosome 1p32.3 at the *PCSK9* locus. Here, 4 independent SNPs ($r^2<0.1$) located within the gene or a distance of <60 kb were associated at genome-wide significance (Table 2; Tables III A and IV in the [Data Supplement](#)). A regional association plot of the locus is shown in Figure V A in the [Data Supplement](#). The lead SNP rs11591147 ($\beta=-0.315$; $P=1.94\times 10^{-17}$) is coding for the *PCSK9-R46L* missense mutation and was previously reported to associate with LDL-C^{5,30} and response to statin therapy.³¹ Further, the *PCSK9-R46L* allele was associated with lower plasma PCSK9 levels and CAD risk in candidate gene studies.¹¹

Associations with LDL-C and coronary heart disease were also reported for other identified SNPs at the *PCSK9* locus.^{6,32,33} These variants are in strong linkage disequilibrium ($r^2>0.5$) with several cis-eQTL-SNPs for *PCSK9* ($P=3.27\times 10^{-27}$;²² Tables IV and V in the [Data Supplement](#)), suggesting a potential functional effect via regulation of *PCSK9* transcription.

Association results were robust when restricting analysis to the statin-free subset. However, because of power loss and different priority pruning, only 3 SNPs reached genome-wide and 2 suggestive significance (regional association plot is given in Figure VI A in the [Data Supplement](#)). Of note, absolute values of β estimates increased for the majority of variants (Table 2; Table III B in the [Data Supplement](#); Figure VII in the [Data Supplement](#)). Association results were also robust against additional adjustments about the observed age–sex interaction and testing the variants in a univariate model (Table VI in the [Data Supplement](#)).

Because multiple SNPs in low linkage disequilibrium at the *PCSK9* locus were associated with plasma PCSK9 levels (Figure 2), we analyzed this allelic heterogeneity in more detail. Of the 4 independent SNPs initially identified at this locus, 3 SNPs (rs11591147, rs45448095, and 1:55520994) remained significant when tested in a multivariate model. Together they explain 4.4% variance of plasma PCSK9. In a combined analysis of these SNPs and all covariates, the SNPs remained significantly associated, and the explained variance increased to 19.4%.

FBXL18 Locus

Besides the *PCSK9* locus, we identified a second locus with genome-wide significance on chromosome 7 (Table 2; Figures V B and VI C in the [Data Supplement](#)). The lead SNP is rs6957201 ($\beta=-0.163$; $P=7.01\times 10^{-9}$). However, the SNP has low minor allele frequency (3%) and inferior imputation quality (info score of 0.549) reducing its credibility. Association results were also consistent in the statin-free subset ($\beta=-0.167$;

Table 1. Characteristics of the LIFE-Heart Study Population

Parameter	All	Men vs Women	Lipid vs No Therapy
Men/women	2167/1191	...	0.448
Age, y	61.7±10.8	<0.001*	<0.001†
Current smoker	699	<0.001*	0.701
Statin therapy	1270	0.448	...
BMI, kg/m ²	29.7±5.0	0.083	<0.001†
Type 2 diabetes mellitus‡	1038	0.125	<0.001†
Hypertension‡	3024	0.721	<0.001†
Systolic BP, mm Hg	138.5±18.6	0.172	0.074
Diastolic BP, mm Hg	83.5±10.9	0.477	<0.001
PCSK9, ng/mL	230.9 (188.9–282.6)	<0.001	<0.001†
Cholesterol, mmol/L	5.43±1.21	<0.001	<0.001
HDL-C, mmol/L	1.34±0.41	<0.001	<0.001
LDL-C, mmol/L	3.32±1.04	0.056	<0.001
ApoA, g/L	1.51±0.29	<0.001	0.070
ApoB, g/L	1.00±0.28	0.223	<0.001
NVD50		<0.001*	<0.001†
0	1230		
0.5	632		
1	503		
2	405		
3	449		
Established CAD	1357	<0.001*	<0.001†
cIMT, mm	0.79±0.15	<0.001*	<0.001†
PS		<0.001*	<0.001†
0	1154		
1	549		
2	756		
3	371		
4	320		
PAD	355	0.127	<0.001†
ABI	1.08±0.19	<0.001*	<0.001

All parameters are restricted to genotyped patients with PCSK9 measurement. For the continuous parameters, the arithmetic mean and SD is given. We tested men against women and statin treatment against no treatment. All binary parameters were tested with a χ^2 test; continuous parameters were tested with a Mann–Whitney *U* test. NVD50 and PS were tested with proportional odds regression. ABI indicates ankle-brachial index; Apo, apolipoprotein; CAD, coronary artery disease; cIMT, carotid intima-media thickness; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; PAD, peripheral artery disease; and PS, plaque score.

*More cases or higher values in men for binary or continuous parameters, respectively.

†More cases or higher values in the statin group for the binary parameters or continuous parameters, respectively.

‡Known, in therapy or acute.

$P=1.70\times 10^{-6}$) and robust against additional adjustment for the observed age/sex interaction (Table VI in the Data Supplement).

Suggestive Loci

Applying a suggestive significance threshold of $P<1\times 10^{-6}$, we identified 3 loci when analyzing all patients (Table VI in the Data Supplement; regional association plots are given in Figure V in the Data Supplement). Those loci were at Xq27.3 (nearest gene *SPANXN4*), 22q12 (*TOM1*), and 13q13 (*RPS12P24*). Restricting GWA analysis to the statin-free subset revealed 4 suggestive loci (6q13 within *KCNQ5*, 18q12 near *SYT4*, 1p31 near *IFI44*, and 4p15; Table VI in the Data Supplement; Figure VI in the Data Supplement). Comprehensive annotation for all suggestive loci is provided in the Data Supplement.

Replication of Previously Reported SNPs

Chernogubova et al¹⁰ reported 6 SNPs at the *PCSK9* locus with nominal significance. Of those 6, 4 were significant in our data, 2 even on genome-wide level. Outside the *PCSK9* locus, we could not replicate the 19 reported hits (Table VII in the Data Supplement).

Theusch et al³⁴ reported 7 SNPs associated with the difference in PCSK9 levels before and after statin treatment. None of these SNPs showed significant association with PCSK9 plasma levels in our data (Table VII in the Data Supplement).

In addition, we looked up known nonsynonymous coding mutations in the *PCSK9* gene and tested them for association with circulating PCSK9 levels. Out of 50 reported mutations,^{35,36} only 5 were in our genotype data. All but one were significantly associated with PCSK9 plasma levels (Table VIII in the Data Supplement).

Further details on replication analyses are documented in the Data Supplement.

Associations With ASVD and MR

We next analyzed the relationship between circulating PCSK9 and the 4 genome-wide significant *PCSK9* SNPs with CAD and other atherosclerotic phenotypes available in LIFE-Heart (Table 3; Tables IX through XII in the Data Supplement).

Association of PCSK9 Variants With Vascular Phenotypes

We identified one *PCSK9* SNPs (1:55520994) that was associated with the number of affected coronary arteries (NVD50) at nominal significance ($P<0.05$). It was also significantly associated with the presence of CAD in the statin-free subset. One of the other *PCSK9* SNPs was associated with cIMT (rs2479409 for all patients). It was also associated with PAD in the statin-free subset (Table IX in the Data Supplement).

We also analyzed the SNP associations with plasma lipid traits. The SNP 1:55520994 was associated with total cholesterol, LDL-C, and ApoB in both, all subjects, and the statin-free subset. Including patient with statins

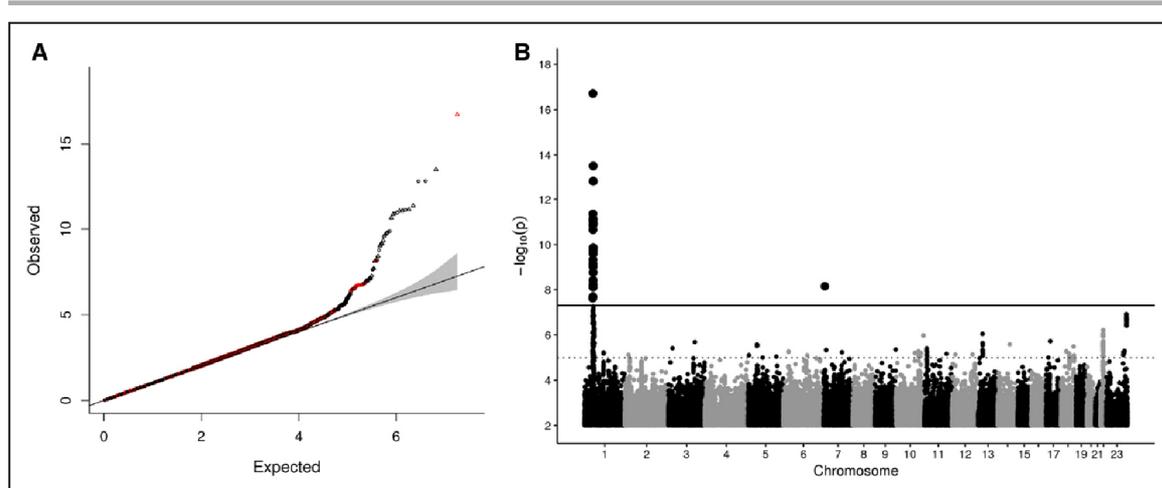


Figure 1. QQ plot and Manhattan plot.

A, QQ plot of P values of the GWAS of all study subjects. SNPs with info score <0.8 are plotted as triangles, and those with minor allele frequency <0.05 are colored red. There is no inflation in our data ($\lambda=1.006$). **B**, Distribution of log-transformed P values of the GWAS of all study subjects. The bold line marks genome-wide significance ($P=5\times 10^{-8}$). Two loci reached genome-wide significance, one at chromosome 1 (*PCSK9*) and the other one at chromosome 7 (*FBXL18*).

attenuated the associations with plasma lipids (Table IX in the [Data Supplement](#)).

Correlation of PCSK9 Plasma Levels With ASVD

PCSK9 plasma levels were significantly associated with CAD status and severity ($P=0.034$ and $P=0.031$, for CAD and NVD50, respectively) only in the statin-free subset (adjusted for all common risk factors: sex, age, smoking, hypertension, BMI, and type 2 diabetes mellitus). Including patients with statins reduced both the association and β estimators (Table X in the [Data Supplement](#)).

In addition, circulating PCSK9 levels were significantly associated with all analyzed subclinical atherosclerosis parameters (cIMT, $P=0.030$; PS, $P=0.004$; PAD, $P=0.032$; ABI, $P=6.36\times 10^{-6}$) in the univariate analysis of all subjects.

MR With PCSK9 Variants as IV and ASVD Phenotypes as Outcomes

Using the 4 independent ($r^2<0.1$), genome-wide significant *PCSK9* SNPs as IVs, we performed MR to analyze whether circulating PCSK9 plasma levels causally affect vascular phenotypes (Table 3).

We found significant causal effects for both CAD status and severity ($\hat{\beta}_{MR}$ (CAD)=0.960, $P=0.048$; $\hat{\beta}_{MR}$ (NVD50)=1.223, $P=0.005$) in the entire cohort. In the statin-free subset analysis, only $\hat{\beta}_{MR}$ (NVD50) reached significance. For the analysis of carotid and peripheral artery phenotypes, we used the combined information of the LIFE-Heart and LIFE-Adult studies (total $N=7647-7792$). We detected significant causal effects on PS ($\hat{\beta}_{MR}$ (PS)=0.605; $P=0.023$) and a trend for cIMT ($P=0.061$). Effects were robust in the statin-free subset.

We could not detect a significant causal effect for ABI or PAD in neither the full cohort nor the statin-free subset analysis. Scatterplots of SNP effects on PCSK9 levels and vascular phenotype are shown in Figures VIII and IX in the [Data Supplement](#).

DISCUSSION

PCSK9 plays an essential role in regulating lipid metabolism and functional inhibition of PCSK9 is a novel treatment strategy for hyperlipidemia. Here, we performed to date the largest genome-wide association study of circulating PCSK9 levels in 3290 patients of the LIFE-Heart study, a cohort focusing on CAD. We identified the *PCSK9* locus as a strong and genome-wide significant genetic determinant of PCSK9 plasma levels and detected considerable allelic heterogeneity. In addition, we discovered a genome-wide significant locus on chromosome 7 within *FBXL18* and several other suggestive loci that warrant independent validation. Availability of coronary angiography (in LIFE-Heart) and other atherosclerotic disease assessments (in LIFE-Heart and LIFE-Adult) allowed us to analyze the causal impact of circulating PCSK9 on these phenotypes for the first time. By MR, we detected causal relationships of PCSK9 for atherosclerosis at the coronary, and carotid arteries. These results provide further evidence that functional inhibition of PCSK9 may contribute to reduce ASVD risk.

Nongenetic Factors Influencing PCSK9

To analyze possible confounders of the observed associations, we first investigated factors influencing PCSK9

Table 2. Genome-Wide Significant SNPs of One of the GWAS

Flag	Locus	SNP Info			Effect/Other Allele	Study 1: All Patients (N=3290)				Study 2: Statin-Free Patients (N=2022)			
		Lead SNP	Nearby Genes (distance)	Info Score		β	SE	P Value	Exp Var	EAF	β	SE	P Value
1	1p32.3	rs11591147	PCSK9 (0 kb)	0.581	T/G	-0.315	0.037	1.94E-17*	0.018	0.012	5.91E-16*	0.030	0.012
1	1p32.3	1:55520994	PCSK9 (0 kb)	0.761	G/C/G	0.060	0.008	3.04E-14*	0.015	0.563	1.96E-12*	0.023	0.558
1	1p32.3	rs45448095†	PCSK9 (0 kb)	0.683	T/C	-0.085	0.012	4.40E-12*	0.012	0.133	3.19E-08	0.014	0.139
0	1p32.3	rs373507733†	PCSK9 (0 kb)	0.581	A/A/T	-0.071	0.011	2.24E-11	0.011	0.234	2.61E-08*	0.014	0.242
1	1p32.3	rs2479409	PCSK9 (0.57 kb)	0.646	A/G	-0.052	0.009	7.41E-09*	0.009	0.640	3.19E-07	0.012	0.640
0	7p22.1	rs6957201	FBXL18 (0 kb) TNRC18 (10 kb)	0.549	C/T	-0.163	0.028	7.01E-09*	0.009	0.971	1.70E-06	0.011	0.970

Model was adjusted for sex, age, current smoking, and statin treatment. EAF indicates effect allele frequency. ExpVar, explained variance; and Flag, SNPs used in Mendelian Randomization analysis.

* Genome-wide significant and independent hits.

† Pairwise linkage disequilibrium ($r^2 > 0.3$), tagged differently because of priority pruning.

plasma concentration in detail. Statins treatment significantly increases PCSK9 levels via the SREBP-2 (sterol regulatory element-binding protein-2) pathway.³⁷ Because 38% of the patients in LIFE-Heart received statins, we performed all analysis in the entire cohort (adjusted for statins) and in the statin-free subset. We found that the results of our GWAS were not confounded by statin treatment, suggesting that adjustment on statin use was sufficient for genetic association analyses. However, the partial correlations of PCSK9 to plasma lipid parameters were typically larger in the statin-free subset, indicating biased relations of PCSK9 to lipids under statin treatment. These results are in line with previous reports that statin use modifies the correlation between PCSK9 and plasma LDL-C.³⁸

Beyond statin treatment, age, sex, and smoking status were the strongest independent factors influencing PCSK9 plasma levels in our study population. These results are in accordance with observations of other cohort studies.^{10,11,39–42} Interestingly, we observed an interaction between sex and age, independent of the other factors. In detail, PCSK9 levels increase in women until an age of 51 to 60 years and then decrease with a similar slope but remaining on higher values than in men. This suggests a menopausal effect on PCSK9 levels and is in agreement with previous findings reporting a sex-specific regulation of PCSK9.^{11,43} However, limiting factors of this finding are the small sample size of young women in our cohort and missing information on menopausal status and hormone replacement therapy. The interaction between age and sex had no effect on the genetic associations reported in our study.

Genetic Factors Influencing Circulating PCSK9

Plasma levels of PCSK9 show substantial variation over a wide range of concentrations, but little is known on the underlying genetic determinants. Candidate gene approaches identified only 1 low-frequency variant (*PCSK9-R46L*) that was robustly associated with lower PCSK9 plasma levels across several studies,^{10,11} whereas a GWAS approach to discover common variants associated with plasma PCSK9 levels did not detect any genome-wide significant associations.¹⁰ Using a considerably larger discovery cohort, we now identified several variants at the *PCSK9* gene locus that were associated with circulating PCSK9 at genome-wide significance, both, in the entire cohort and in the statin-free subset. The strongest association was observed for SNP rs11591147, encoding the above-mentioned missense mutation *PCSK9-R46L*. This association is plausible as the mutation decreases the phosphorylation of Ser47, which is assumed to protect PCSK9 from proteolysis.^{35,44} Therefore, PCSK9 secretion into

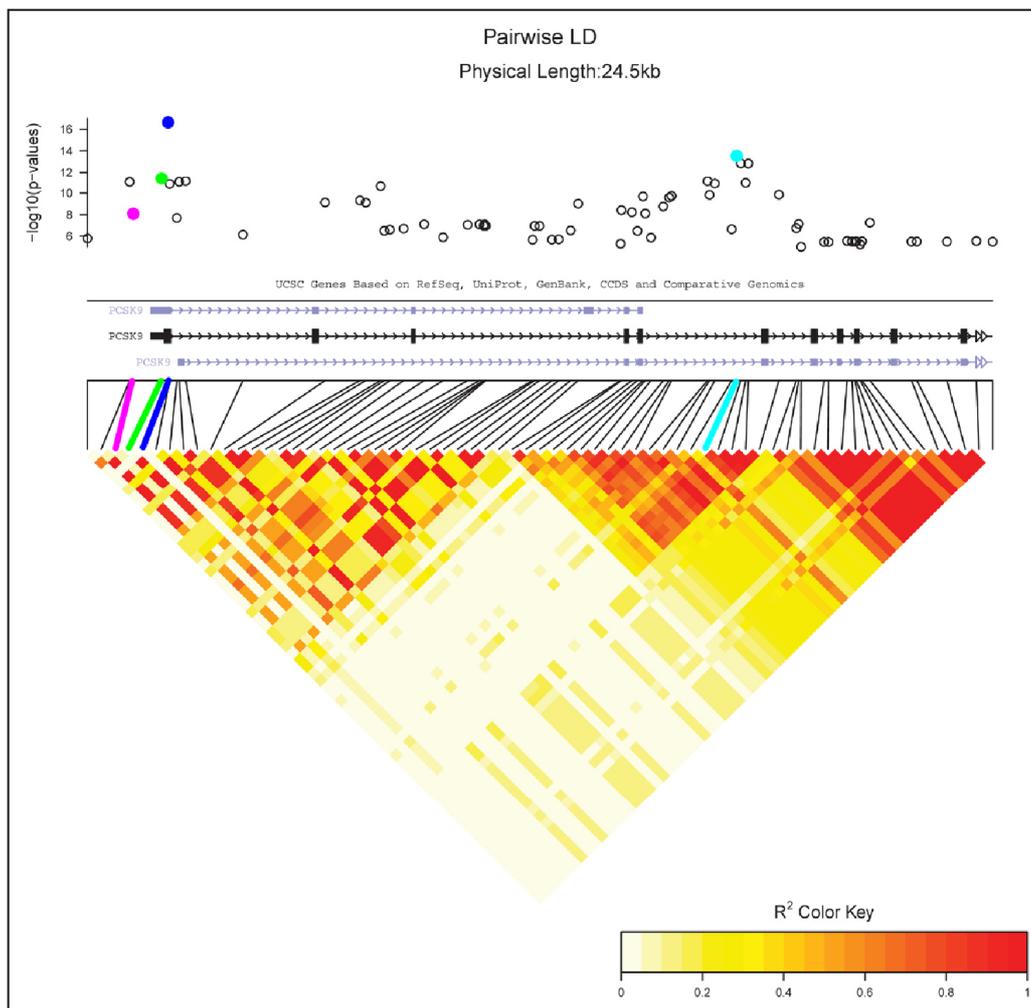


Figure 2. Heat map of pairwise linkage disequilibrium (r^2) for the SNPs at the *PCSK9* locus.

Gene range was added according to University of California, Santa Cruz Genes track (hg19). *P* values of the GWAS were shown according to the position in the gene, not in heat map. Colors mark the genome-wide significant SNPs with pairwise LD $r^2 < 0.1$ (blue: rs11591147, green: rs45448095, magenta: rs2479409, and cyan: rs1:55520994).

plasma decreases with rs11591147-T. We also detected 3 additional independent and multivariate significant SNPs, suggesting substantial allelic heterogeneity. Moreover, the *PCSK9* locus harbored several other suggestive SNPs including the rare nonsynonymous coding mutations A53V, V474I, and E670G that correlated with circulating PCSK9 levels. Together our findings suggest that variants at the *PCSK9* gene locus are the predominant genetic determinants for circulating PCSK9 levels in this population. Associations of the *PCSK9* variants with plasma lipids (CHOL, LDL-C, and ApoB) were much weaker and likely mediated by the effect of PCSK9 on lipid metabolism.

Besides the *PCSK9* locus, we identified a second genome-wide significant locus in a gene-rich area at chromosome 7p22. The lead SNP was located in

FBXL18, which encodes the F-box and leucine-rich repeat protein 18, whose function remains largely unknown. However, the SNP has inferior quality metrics. Follow-up studies especially replication in an independent population and functional assessments are necessary to validate this finding. Likewise, validation is also required for the suggestive hits ($P < 1 \times 10^{-6}$) identified in the entire cohort (Xq27.3 near *SPANXN4*, 22q12 near *TOM1*, and 13q13 near *RPS12P24*) or the statin-free subset (6q13 within *KCNQ5*, 18q12 near *SYT4*, 1p31 near *IFI44*, and 4p15). Interestingly, suggestive associations between *KCNQ5* and LDL-C were previously reported in the Framingham Heart Study Offspring Cohort,⁴⁵ and we have recently reported suggestive associations between *KCNQ5* and carotid atherosclerotic plaque burden.⁴⁶

Table 3. Summary of MR Studies Using Inverse-Variance Weighted Method

ASVD	N (All)		N (Subset)		All Patients				Statin-Free Patients						
	SE	P Value	β	SE	P Value	Correlation		Causality		SE	P Value	Correlation		Causality	
						β	SE	β	SE			β	SE	β	SE
CAD	0.117	0.101	0.192	0.486	0.048*	0.166	0.589	1.118	0.641	0.081	0.166	0.589	1.118	0.641	0.081
NVD50	0.104	0.075	0.185	0.434	0.005*	0.146	0.382	1.523	0.633	0.016*	0.146	0.382	1.523	0.633	0.016*
PS	0.104	0.004*	0.301	0.266	0.023*	0.144	0.440	0.769	0.290	0.008*	0.144	0.440	0.769	0.290	0.008*
cIMT	0.009	0.030*	-0.019	0.022	0.061	0.013	0.002*	0.045	0.023	0.054	0.013	0.002*	0.045	0.023	0.054
PAD	0.185	0.032*	0.398	0.708	0.412	0.282	0.646	-1.108	1.249	0.375	0.282	0.646	-1.108	1.249	0.375
ABI	0.011	6.36E-06*	-0.049	0.022	0.794	0.013	0.028*	0.023	0.021	0.287	0.013	0.028*	0.023	0.021	0.287

Univariate simple regression results of PCSK9 and the different ASVDs are provided in the Correlation columns. MR-based causal effects of PCSK9 on the ASVDs are given in the Causality columns. MR results for coronary phenotypes were calculated in LIFE-Heart only. ABI indicates ankle-brachial index; ASVD, atherosclerotic vascular disease; CAD, coronary artery disease; cIMT, carotid intima-media thickness; PAD, peripheral artery disease; and PS, plaque score.

*Significant P values (P<0.05).

PCSK9 Gene Variant and PCSK9 Plasma Levels in Relation to Vascular Phenotypes

Recent MR studies have used *PCSK9* SNPs as instruments to successfully show the causality of LDL-C on CAD or type 2 diabetes mellitus risk.^{47–50} However, variants at the *PCSK9* locus are much stronger associated with circulating PCSK9 protein levels. Hence, SNPs at the *PCSK9* gene locus are well suited to perform MR analysis of the causal relationship of PCSK9 protein levels and ASVD phenotypes. This is of particular clinical interest because clinical applications for functional inhibitors of PCSK9 are currently explored.

We confirmed the previously reported positive correlation of PCSK9 and PAD.⁵¹ In line with this, a negative correlation of PCSK9 and ABI was observed. However, in MR analysis, we found no causal effect.

Despite the lack of a raw correlation of PCSK9 with CAD and NVD50 for both the complete data set and the statin-free subjects, we detected causal effects of PCSK9 on coronary and carotid phenotypes. We suppose that the strong confounding factors age and male sex blur the raw correlations. On one hand, these factors increase the risk for CAD and NVD50; on the other hand, they reduce PCSK9.

Our study suggests that a genetic reduction of PCSK9 levels by 50% is associated with a reduction of CAD risk by 50%. Interestingly, the effect on carotid artery atherosclerosis was just half as strong as on CAD. On treatment strategies to inhibit PCSK9, the results of the MR study may not be overstated because the detected effect is a lifelong effect. Later onset of treatment with PCSK9 inhibitors might result in a more moderate effect.

Study Limitations

The present study has the following limitations. The GWAS was performed in a population at risk or with present CAD, that is, not within a population-based study. On the other hand, the availability of coronary angiography and vascular phenotyping allowed to analyze CAD severity in relation to PCSK9 and its genetic variants. The present study also included a high percentage of patients on statin treatment, which is known to increase PCSK9 levels. Therefore, all analyses were also performed in the statin-free subset and were robust.

We have to acknowledge that MR as a method to show causality relies on assumptions that inherently cannot be proven. In the present case, we demonstrated plausibility of the assumptions by our considerations shown in Mendelian Randomization Analysis in the [Data Supplement](#).

Last, the ELISA assay used to determine plasma PCSK9 levels does not differentiate between active and inactive PCSK9. It is known that PCSK9 circulates as

mature and furin-cleaved protein and that a fraction of PCSK9 is also lipoprotein bound.¹¹ Possible differences in associations of these subfractions with genetics and phenotypes could not be assessed.

Conclusions

In conclusion, by genome-wide association for circulating PCSK9 levels, we identified 4 independent genetic variants at the *PCSK9* locus at genome-wide significance, explaining 4.4% of the total variance of PCSK9. In addition, we describe 1 genome-wide significant locus on chromosome 7p22 and 7 suggestive loci associated with plasma PCSK9, which require independent replication and identification of the causal genes.

We performed the first MR study to address causal effects of PCSK9 on ASVD phenotypes. We detected significant causal effects for atherosclerosis at the coronary and carotid arteries. This increases confidence in PCSK9 as a valid target to reduce risk or severity of ASVD.

ARTICLE INFORMATION

Received October 20, 2017; accepted March 5, 2018.

The Data Supplement is available at <http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGEN.117.001992/-/DC1>.

Correspondence

Markus Scholz, PhD, Institute for Medical Informatics, Statistics and Epidemiology, Haertelstrasse 16-18, 04107 Leipzig, Germany, E-mail markus.scholz@imise.uni-leipzig.de or Ralph Burkhardt, MD, Institute of Laboratory Medicine, Liebigstraße 27, 04103 Leipzig, Germany, E-mail ralph.burkhardt@medizin.uni-leipzig.de

Affiliations

Institute for Medical Informatics, Statistics and Epidemiology, Leipzig, Germany (J.P., K.H., H.K., M.L., M.S.). LIFE Research Center for Civilization Diseases, University of Leipzig, Germany (J.P., V.S., A.T., H.K., C.B., M.L., J.T., R.B., M.S.). Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital, Leipzig, Germany (V.S., C.B., J.K., J.T., R.B.). Heart Center Leipzig, Germany (A.T.).

Acknowledgments

We thank the participants of the LIFE-Heart and the LIFE-Adult studies. We thank Sylvia Henger for data quality control, Kay Olscher and Annegret Unger for technical assistance, and Kerstin Wirkner for running the LIFE-Adult study center.

Sources of Funding

LIFE-Heart and LIFE-Adult are funded by the Leipzig Research Center for Civilization Diseases (LIFE). LIFE is an organizational unit affiliated to the Medical Faculty of the University of Leipzig. LIFE is funded by means of the European Union, by the European Regional Development Fund (ERDF) and by funds of the Free State of Saxony within the framework of the excellence initiative. Dr Burkhardt is supported by a grant from the German Research Foundation (DFG SFB1052-B7).

Disclosures

None.

REFERENCES

- Rader DJ. New therapeutic approaches to the treatment of dyslipidemia. *Cell Metab*. 2016;23:405–412. doi: 10.1016/j.cmet.2016.01.005.

- Norata GD, Tavori H, Pirillo A, Fazio S, Catapano AL. Biology of proprotein convertase subtilisin kexin 9: beyond low-density lipoprotein cholesterol lowering. *Cardiovasc Res*. 2016;112:429–442. doi: 10.1093/cvr/cvz194.
- Abifadel M, Varret M, Rabès JP, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet*. 2003;34:154–156. doi: 10.1038/ng1161.
- Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med*. 2006;354:1264–1272. doi: 10.1056/NEJMoa054013.
- Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet*. 2008;40:189–197. doi: 10.1038/ng.75.
- Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47:1121–1130. doi: 10.1038/ng.3396.
- Robinson JG, Farnier M, Krempf M, Bergeron J, Luc G, Averna M, et al; ODYSSEY LONG TERM Investigators. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. *N Engl J Med*. 2015;372:1489–1499. doi: 10.1056/NEJMoa1501031.
- Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al; FOURIER Steering Committee and Investigators. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med*. 2017;376:1713–1722. doi: 10.1056/NEJMoa1615664.
- Vlachopoulos C, Terentes-Printzios D, Georgiopoulou G, Skoumas I, Koutagiar I, Ioakeimidis N, et al. Prediction of cardiovascular events with levels of proprotein convertase subtilisin/kexin type 9: a systematic review and meta-analysis. *Atherosclerosis*. 2016;252:50–60. doi: 10.1016/j.atherosclerosis.2016.07.922.
- Chernogubova E, Strawbridge R, Mahdessian H, Mälärstig A, Krapivner S, Gigante B, et al. Common and low-frequency genetic variants in the PCSK9 locus influence circulating PCSK9 levels. *Arterioscler Thromb Vasc Biol*. 2012;32:1526–1534. doi: 10.1161/ATVBAHA.111.240549.
- Lakoski SG, Lagace TA, Cohen JC, Horton JD, Hobbs HH. Genetic and metabolic determinants of plasma PCSK9 levels. *J Clin Endocrinol Metab*. 2009;94:2537–2543. doi: 10.1210/jc.2009-0141.
- Beutner F, Teupser D, Gielen S, Holdt LM, Scholz M, Boudriot E, et al. Rationale and design of the Leipzig (LIFE) Heart Study: phenotyping and cardiovascular characteristics of patients with coronary artery disease. *PLoS One*. 2011;6:e29070. doi: 10.1371/journal.pone.0029070.
- Loeffler M, Engel C, Ahnert P, Alfermann D, Arelin K, Baber R, et al. The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health*. 2015;15:691. doi: 10.1186/s12889-015-1983-z.
- Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74. doi: 10.1038/nature15393.
- Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. Haplotype estimation using sequencing reads. *Am J Hum Genet*. 2013;93:687–696. doi: 10.1016/j.ajhg.2013.09.002.
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5:e1000529. doi: 10.1371/journal.pgen.1000529.
- Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39:906–913. doi: 10.1038/ng2088.
- König IR, Loley C, Erdmann J, Ziegler A. How to include chromosome X in your genome-wide association study. *Genet Epidemiol*. 2014;38:97–103. doi: 10.1002/gepi.21782.
- Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. *Database (Oxford)*. 2016;2016:baw093. doi: 10.1093/database/baw093.
- Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(Database issue):D1001–D1006. doi: 10.1093/nar/gkt1229.
- Kirsten H, Al-Hasani H, Holdt L, Gross A, Beutner F, Krohn K, et al. Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci. *Hum Mol Genet*. 2015;24:4746–4763. doi: 10.1093/hmg/ddv194.
- GTEx Consortium. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648–660. doi: 10.1126/science.1262110.

23. Yu G, Wang L-G, Yan G-R, He Q-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*. 2015;31:608–609. doi:10.1093/bioinformatics/btu684.
24. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst*. 2016;12:477–479. doi:10.1039/c5mb00663e.
25. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310–315. doi:10.1038/ng.2892.
26. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol*. 2014;10:e1003440. doi:10.1371/journal.pcbi.1003440.
27. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med*. 2016;35:1880–1906. doi:10.1002/sim.6835.
28. Bowden J, Del Greco M F, Minelli C, Lawlor D, Sheehan N, Thompson J, et al. Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption. 2017. <https://doi.org/10.1101/159442>. Accessed April 24, 2018.
29. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2016.
30. Surakka I, Horikoshi M, Mägi R, Sarin AP, Mahajan A, Lagou V, et al.; ENGAGE Consortium. The impact of low-frequency and rare variants on lipid levels. *Nat Genet*. 2015;47:589–597. doi:10.1038/ng.3300.
31. Chasman DI, Giulianini F, MacFadyen J, Barratt BJ, Nyberg F, Ridker PM. Genetic determinants of statin-induced low-density lipoprotein cholesterol reduction: the Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial. *Circ Cardiovasc Genet*. 2012;5:257–264. doi:10.1161/CIRCGENETICS.111.961144.
32. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet*. 2009;41:334–41. doi:10.1038/ng.327.
33. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al; Cardiogenics, CARDIoGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333–338. doi:10.1038/ng.784.
34. Theusch E, Medina MW, Rotter JI, Krauss RM. Ancestry and other genetic associations with plasma PCSK9 response to simvastatin. *Pharmacogenet Genomics*. 2014;24:492–500. doi:10.1097/FPC.0000000000000081.
35. Dron JS, Hegele RA. Complexity of mechanisms among human proprotein convertase subtilisin-kexin type 9 variants. *Curr Opin Lipidol*. 2017;28:161–169. doi:10.1097/MOL.0000000000000386.
36. Kotowski IK, Pertsemidis A, Luke A, Cooper RS, Vega GL, Cohen JC, et al. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet*. 2006;78:410–422. doi:10.1086/500615.
37. Sahebkar A, Simental-Mendia LE, Guerrero-Romero F, Golledge J, Watts GF. Effect of statin therapy on plasma proprotein convertase subtilisin kexin 9 (PCSK9) concentrations: a systematic review and meta-analysis of clinical trials. *Diabetes Obes Metab*. 2015;17:1042–1055. doi:10.1111/dom.12536.
38. Welder G, Zineh I, Pacanowski MA, Troutt JS, Cao G, Konrad RJ. High-dose atorvastatin causes a rapid sustained increase in human serum PCSK9 and disrupts its correlation with LDL cholesterol. *J Lipid Res*. 2010;51:2714–2721. doi:10.1194/jlr.M008144.
39. Baass A, Dubuc G, Tremblay M, Delvin EE, O’Loughlin J, Levy E, et al. Plasma PCSK9 is associated with age, sex, and multiple metabolic markers in a population-based sample of children and adolescents. *Clin Chem*. 2009;55:1637–1645. doi:10.1373/clinchem.2009.126987.
40. Leander K, Mälarstig A, Van’t Hooft FM, Hyde C, Hellénius ML, Troutt JS, et al. Circulating proprotein convertase subtilisin/kexin type 9 (PCSK9) predicts future risk of cardiovascular events independently of established risk factors. *Circulation*. 2016;133:1230–1239. doi:10.1161/CIRCULATIONAHA.115.018531.
41. Ridker PM, Rifai N, Bradwin G, Rose L. Plasma proprotein convertase subtilisin/kexin type 9 levels and the risk of first cardiovascular events. *Eur Heart J*. 2016;37:554–60. doi:10.1093/eurheartj/ehv568.
42. Zhu YM, Anderson TJ, Sikdar K, Fung M, McQueen MJ, Lonn EM, et al. Association of proprotein convertase subtilisin/kexin type 9 (PCSK9) with cardiovascular risk in primary prevention. *Arterioscler Thromb Vasc Biol*. 2015;35:2254–2259. doi:10.1161/ATVBAHA.115.306172.
43. Rusica M, Ferri N, Fogacci F, Rosticci M, Botta M, Marchiano S, et al. Circulating levels of proprotein convertase subtilisin/kexin type 9 and arterial stiffness in a large population sample: data from the Brisighella Heart Study. *J Am Heart Assoc* 2017. doi:10.1161/JAHA.117.005764.
44. Dewpura T, Raymond A, Hamelin J, Seidah NG, Mbikay M, Chrétien M, et al. PCSK9 is phosphorylated by a Golgi casein kinase-like kinase *ex vivo* and circulates as a phosphoprotein in humans. *FEBS J*. 2008;275:3480–3493. doi:10.1111/j.1742-4658.2008.06495.x.
45. Kathiresan S, Manning AK, Demissie S, D’Agostino RB, Surti A, Guiducci C, et al. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet*. 2007;8(suppl 1):S17. doi:10.1186/1471-2350-8-S1-S17.
46. Pott J, Burkhardt R, Beutner F, Horn K, Teren A, Kirsten H, et al. Genome-wide meta-analysis identifies novel loci of plaque burden in carotid artery. *Atherosclerosis*. 2017;259:32–40. doi:10.1016/j.atherosclerosis.2017.02.018.
47. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. *J Am Coll Cardiol*. 2012;60:2631–2639. doi:10.1016/j.jacc.2012.09.017.
48. Jansen H, Lieb W, Schunkert H. Mendelian randomization for the identification of causal pathways in atherosclerotic vascular disease. *Cardiovasc Drugs Ther*. 2016;30:41–49. doi:10.1007/s10557-016-6640-y.
49. Rosenson RS, Koenig W. Mendelian randomization analyses for selection of therapeutic targets for cardiovascular disease prevention: a note of circumspection. *Cardiovasc Drugs Ther*. 2016;30:65–74. doi:10.1007/s10557-016-6642-9.
50. Schmidt AF, Swerdlow DJ, Holmes MV, Patel RS, Fairhurst-Hunter Z, Lyall DM, et al; Lifelines Cohort study group; UCLEB consortium. PCSK9 genetic variants and risk of type 2 diabetes: a Mendelian Randomisation study. *Lancet Diabetes Endocrinol*. 2017;5:97–105. doi:10.1016/S2213-8587(16)30396-5.
51. Chao T-H, Chen I-C, Li Y-H, Lee P-T, Tseng S-Y. Plasma levels of proprotein convertase subtilisin/kexin type 9 are elevated in patients with peripheral artery disease and associated with metabolic disorders and dysfunction in circulating progenitor cells. *J Am Heart Assoc* 2016. doi:10.1161/JAHA.116.003497.

2.3 Zusätzliche Ergebnisse

In diesem Kapitel werden drei weitere Analysen beschrieben, die bislang nicht publiziert sind. Es handelt sich hierbei um folgende Aspekte bzw. Fragestellungen:

- Die Prävalenz von Karotis-Plaques unterscheidet sich stark zwischen den beiden LIFE-Studien. Es wird das genetische bzw. lebensstil-assoziierte Risiko für KHK (als Stellvertreter für Atherosklerose) zwischen LIFE-Adult und LIFE-Heart verglichen mit der Fragestellung, ob die unterschiedliche Prävalenz durch diese Risikoscores erklärt werden kann.
- Eine kürzlich publizierte genetische Analyse des KHK-Schweregrads konnte bis auf den bekannten 9p21 Locus keinen weiteren detektieren [47]. Es stellt sich daher die Frage, ob dies auch für die Schwere der Belastung mit Karotis-Plaques zutrifft. Dazu wird eine Meta-GWAS analog zur ersten Publikation (Kapitel 2.1) durchgeführt, allerdings eingeschränkt auf Patienten / Probanden mit Karotis-Plaques ($PS > 0$, Case-Only Studie).
- In der zweiten Publikation (Kapitel 2.2) wurden vier unabhängige SNPs im *PCSK9* Locus beschrieben, die allelische Heterogenität suggerieren. Deshalb wird eine Haplotypen-Analyse des *PCSK9* Locus durchgeführt mit dem Ziel, häufige Haplotypen zu detektieren und Gruppen mit ungefähr gleichen *PCSK9*-Spiegeln zu definieren. Abschließend soll die Frage geklärt werden, ob die Haplotypgruppen mehr Varianz erklären als die vier SNPs im multivariaten Modell.

2.3.1 Vergleich LIFE-Heart – LIFE-Adult

LIFE-Adult ist eine Querschnittsstudie, zu der zufällig 10,000 Personen aus Leipzig eingeladen wurden. LIFE-Heart ist zwar ebenfalls eine Querschnittsstudie, aber selektiert auf Patienten des Herzzentrums Leipzig, die mit Verdacht auf KHK, einer manifesten Herzerkrankung oder Myokardinfarkt rekrutiert wurden. Wie in der ersten Publikation [43] gezeigt werden konnte, gab es signifikante Unterschiede zwischen den Studien bezüglich fast aller Risikofaktoren der Atherosklerose (siehe Tabelle 1 der Publikation). Die einzige Ausnahme stellte das Alter der Teilnehmer dar, welches nur einen geringen Unterschied aufwies. Neben den Risikofaktoren unterscheidet sich auch die Verteilung des Plaquescores PS signifikant voneinander und die Plaqueprävalenz ist in LIFE-Heart signifikant höher (OR [95%-KI] = 2.26 [2.03; 2.52]). Adjustiert auf Alter und Geschlecht konnte ein signifikanter Studieneffekt auf PS detektiert werden ($\beta=0.359$, $p < 2 \times 10^{-16}$, s. Abb. 11).

Dies führt zur Frage, ob dieser Studieneffekt auf PS durch die Genetik erklärt werden kann und ob es zusätzlich ein unterschiedliches genetisches Risikoprofil für Atherosklerose gibt.

Methodik & Ergebnisse

Es wurde ein genetischer Risikoscore (GRS) anhand der 58 KHK-Loki von Nikpay et al. [3] bestimmt. Alle Allele wurden so kodiert, dass der Effekt auf KHK positiv ist, d.h. $\beta_j > 0$. Je höher der GRS, desto höher das Risiko für KHK.

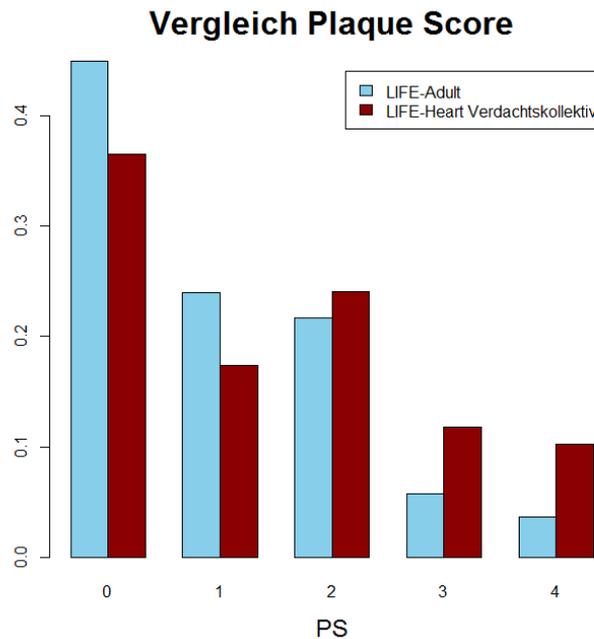


Abbildung 11: Barplots des Plaque-Scores PS. Die relativen Häufigkeiten der jeweiligen Plaque-Anzahl ist für LIFE-Adult (blau, n=4325) und das Verdachtskollektiv der LIFE-Heart (dunkelrot, n=3157) gezeigt. Man erkennt eine deutlich höhere Plaquebelastung in LIFE-Heart ($p < 2 \times 10^{-16}$).

Neben dem Vergleich des genetischen Risikos wurde auch das lebensstil-bedingte Risiko in zwei Scores erfasst, ähnlich zu Khera et al. [9]. Der erste Lebensstilscore (LS1) wurde als Summe folgender Parameter erstellt: Rauchen, Übergewicht und sportliche Aktivität. Es wird je einen Punkt vergeben, wenn der Proband/Patient aktueller Raucher ist bzw. einen BMI ≥ 30 kg/m² hat. Die sportliche Aktivität wurde in beiden Studien unterschiedlich erfasst. Die Teilnehmer von LIFE-Adult füllten im Selbstbericht den IPAQ (International Physical Activity Questionnaire) aus, dessen Auswertung drei Kategorien ergibt: hohe, mittlere und geringe Aktivität [48]. Hier wurde einen Punkt vergeben, wenn der Teilnehmer in der Kategorie mit geringer Aktivität ist. In LIFE-Heart wurde lediglich mittels Fragebogen abgefragt, ob man regelmäßig Sport treibt. Einen Punkt für fehlende sportliche Aktivität wurde hier vergeben, falls nicht regelmäßig mindestens einmal die Woche sportliche Betätigung von mehr als 15 Minuten angegeben wurde. Da sich die Assessments zur sportlichen Aktivität unterscheiden, ist die Aussagekraft dieses Scores eingeschränkt. Daher wurde ein zweiter Score (LS2) definiert, der nur das aktuelle Rauchverhalten und den BMI berücksichtigt. Eine Übersicht aller Parameter ist in Tabelle 4 gegeben.

Um das genetische Risiko der gesunden Probanden von LIFE-Adult (n=4325) zu vergleichen, wurden zwei Analysen durchgeführt: zum einen ein Vergleich mit Patienten des Verdachtskollektiv der LIFE-Heart (n=3157 mit Genetik), zum anderen einer mit den Patienten

mit manifester KHK- bzw. akuten Myokardinfarkten von LIFE-Heart (n=2543 mit Genetik). Die Verteilungen des GRS sind in Abbildung 12 gezeigt. Während sich im ersten Vergleich mit dem Verdachtskollektiv nur ein Trend zu höheren GRS-Werten zeigte (U-Test, $p=0.068$), konnte ein signifikanter Unterschied zwischen gesunden LIFE-Adult Probanden und KHK- bzw. MI-Patienten detektiert werden (U-Test, $p=0.002$).

	LIFE-Adult (n=4325)	LIFE-Heart (n=3157)	p-Wert
Männer (in %)	48.44	64.75	$< 2.2 \times 10^{-16}$
Alter (MW \pm SD)	62.14 \pm 11.15	61.72 \pm 10.71	0.104
BMI \geq 30 kg/m ² (in %)	25.71	42.07	$< 2.2 \times 10^{-16}$
Aktuelle Raucher (in %)	17.43	20.62	5.5×10^{-4}
Geringe sportliche Aktivität (in %)	8.95	62.97	$< 2.2 \times 10^{-16}$
LS1 (in %)	46.87; 38.80; 13.27; 1.06	12.26; 42.03; 39.69; 6.02	$< 2.2 \times 10^{-16}$
LS2 (in %)	60.30; 36.25; 3.45	44.25; 48.81; 6.94	$< 2.2 \times 10^{-16}$
GRS (MW \pm SD)	2.59 \pm 0.23	2.60 \pm 0.23	0.068
PS > 0 (in %)	55.05	63.45	4.0×10^{-13}

Tabelle 4: Studien-Charakteristika bezüglich der verwendeten Parameter in der Risikoscoreanalyse. Die p-Werte von Geschlecht, Rauchen, sportlicher Aktivität und PS wurden mittels χ^2 -Test bestimmt, die von Alter und GRS mittels U-Test und die der Lebensstil-Scores (LS1 und LS2) mittels Armitage Trend-Test. Die Daten wurden eingeschränkt auf vollständige Assessments.

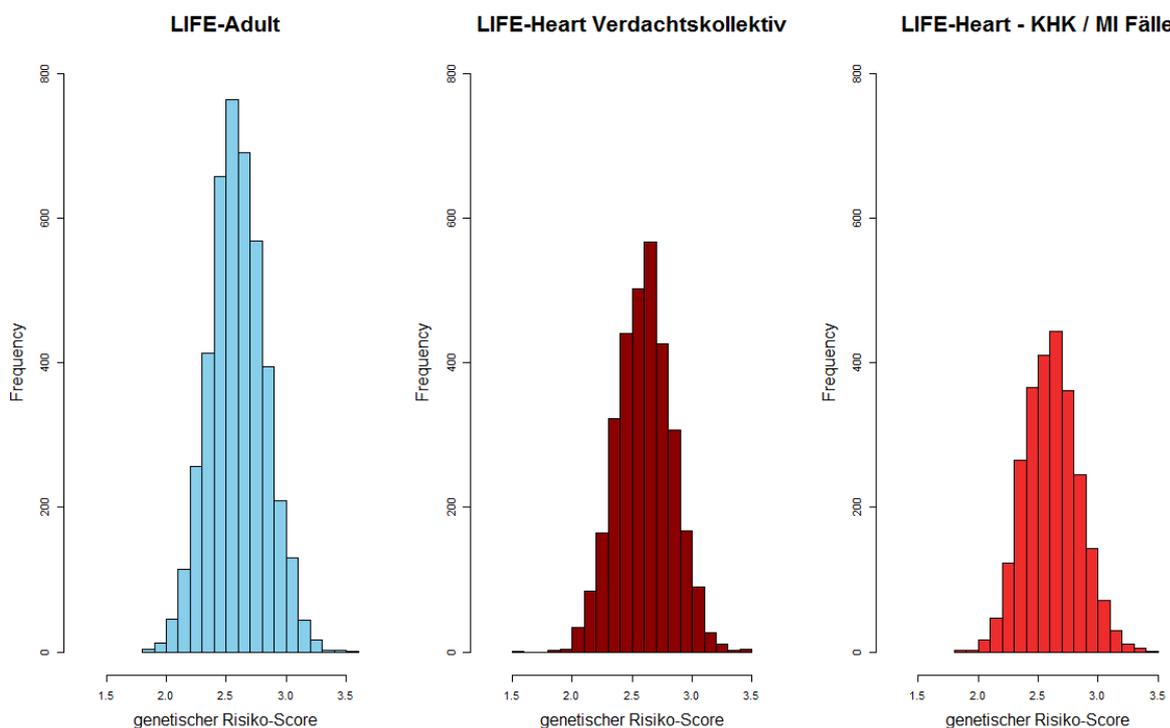


Abbildung 12: Histogramme des genetischen Risiko-Scores (GRS) in LIFE-Adult (links, n=4325), im Verdachtskollektiv der LIFE-Heart (mittig, n=3157), und in KHK bzw. MI-Patienten (rechts, n=2543), . Für den GRS wurden 58 KHK-SNPs mit ihren jeweiligen Effekten aus der Publikation von Nikpay et al. [3] verwendet, wobei alle SNPs auf einen positiven Effekt kodiert wurden. Je höher der GRS, desto höher ist das Risiko für KHK. Es gab einen signifikanten Unterschied zwischen LIFE-Adult und den KHK / MI-Patienten von LIFE-Heart ($p=0.002$) und einen Trend zwischen LIFE-Adult und dem Verdachtskollektiv ($p=0.068$). Es konnte kein signifikanter Unterschied zwischen dem Verdachtskollektiv und den KHK/MI-Patienten festgestellt werden.

Die Lebensstil-Scores wurden nur für das Verdachtskollektiv der LIFE-Heart erstellt. Hier ist für beide Scores ein deutlicher Trend zu erkennen: Patienten der LIFE-Heart Studie tendieren zu einem schlechteren Lebensstil (Armitage Trend-Test, $p < 2 \times 10^{-16}$, s. Abb. 13).

Um die Effekte der Genetik und des Lebensstils auf den Plaque-Score PS zu vergleichen, wurden zwei multivariate Modelle gerechnet, je eines für die beiden Lebensstil-Scores (zusätzlich adjustiert auf Alter und Geschlecht). Der Effekt des GRS ist über die Modelle und Studien hinweg sehr ähnlich (s. Tab. 5) und erklärt zwischen 0.3% und 0.4% der Varianz von PS. Der Effekt der Lebensstil-Scores ist stärker als der der Genetik, wobei der LS2 weniger Varianz erklärt als LS1. Dies war zu erwarten, da LS1 zusätzliche Information über die sportliche Aktivität enthält. Im Vergleich zwischen den Studien ist der Effekt des Lebensstils in LIFE-Heart größer. Unter Berücksichtigung der Genetik und Lebensstil wird der Studien-Effekt kleiner, bleibt aber signifikant ($\beta = 0.210$, $p = 8.8 \times 10^{-13}$).

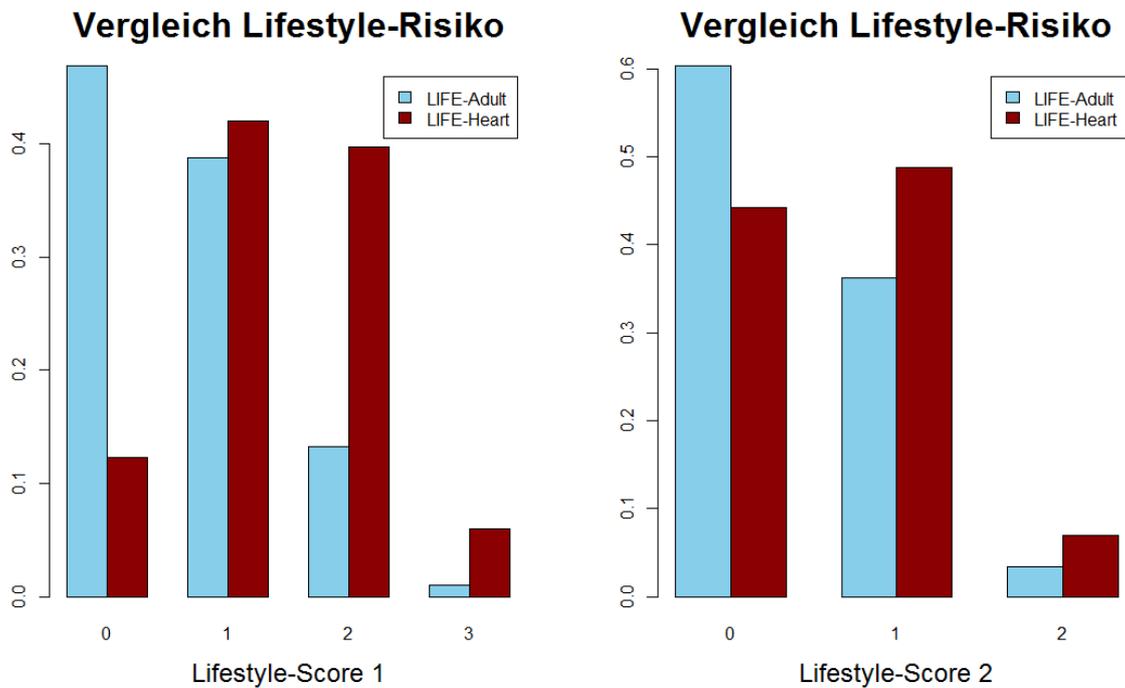


Abbildung 13: Barplots der beiden Lebensstil-Scores mit Berücksichtigung von sportlicher Aktivität (links) und ohne (rechts). Rauchen und Übergewicht wurde immer verwendet. Je höher der Score, desto höher das Risiko für Atherosklerose. In beiden Fällen sind die Lebensstil-scores in LIFE-Heart deutlich hin zu ungünstigeren Werten verschoben ($p < 2 \times 10^{-16}$).

adjustiert auf Alter und Geschlecht		LS			GRS		
		Stand. β	p-Wert	Erklärte Varianz	Stand. β	p-Wert	Erklärte Varianz
LIFE-Adult (n=4325)	LS1	0.086	6.19×10^{-10}	0.55%	0.058	2.92×10^{-5}	0.31%
	LS2	0.115	5.06×10^{-10}	0.26%	0.059	2.52×10^{-5}	0.31%
LIFE-Heart (n=3157)	LS1	0.175	$< 2 \times 10^{-16}$	1.84%	0.063	1.7×10^{-4}	0.38%
	LS2	0.197	$< 2 \times 10^{-16}$	1.03%	0.062	2.0×10^{-4}	0.38%

Tabelle 5: Einfluss von Genetik und Lebensstil auf Plaque-Score PS. Ergebnisse der multivariaten linearen Regression von LS, GRS, Alter und Geschlecht.

Zusammenfassung

Es gibt einen Trend zu höheren GRS-Werten in der LIFE-Heart Studie gegenüber LIFE-Adult. Dieser Trend wird signifikant, wenn man statt des Verdachtskollektivs die KHK- und MI-Fälle mit LIFE-Adult vergleicht. Das Ergebnis ist plausibel, da in einem Kollektiv von KHK-Patienten die Chance höher ist, KHK-fördernde Allele zu beobachten.

Der Effekt des GRS auf PS ist gering und in beiden Studien ähnlich. Ein möglicher Grund für den relativ schwachen Effekt auf Karotis-Plaques kann dadurch erklärt werden, dass die für den GRS verwendeten SNPs KHK-SNPs sind und dadurch das genetische Risiko für Karotis-Plaques nicht exakt darstellen.

Beide LSs haben einen signifikanten Effekt auf PS, wobei der Effekt in LIFE-Heart stärker ist. Parallel dazu zeigt sich, dass LIFE-Heart Patienten eine Tendenz zu schlechterem Lebensstil aufweisen.

Der Studien-Effekt wird durch Adjustierung auf Genetik und Lebensstil verringert, während sowohl LS1 bzw. LS2 und GRS signifikant assoziiert sind (alle p -Werte $<5 \times 10^{-8}$). Dies ist ein Hinweis darauf, dass sowohl Genetik als auch Lebensstil einen Teil des Studien-Effekts erklären, jedoch ist diese Erklärung nicht vollständig.

2.3.2 Genetische Analyse des Plaque-Score eingeschränkt auf Fälle

In der ersten Publikation [43] wurden genetische Einflüsse auf Karotis-Plaques untersucht. Insgesamt sind 36.6% der Patienten in LIFE-Heart, bzw. 45.1% in LIFE-Adult Kontrollen, d.h. sie haben einen Plaque-Score PS=0. Dadurch wird neben dem Schweregrad auch die Prävalenz eines Plaques abgebildet. Um zu analysieren, ob ähnlich wie bei der KHK genetische Unterschiede zwischen Prävalenz und Schweregrad erkennbar sind, bietet es sich an, nur an Karotis-Plaques erkrankte Teilnehmer und deren Schweregrad zu analysieren (PS=1, ..., 4).

Daher wurde analog zur ersten Publikation für beide Studien je eine GWAS zu PS durchgeführt, eingeschränkt auf Fälle, die mindestens einen Plaque aufweisen (n=2004 in LIFE-Adult und n=2585 in LIFE-Adult) und adjustiert auf Alter und Geschlecht. Der genetische Effekt wurde als additiv angenommen und wie in der Publikation wurde PS als quasi kontinuierliche Zielgröße betrachtet. Zusätzlich wurde hier der umfangreichere Score PS8 analysiert (n=1593 in LIFE-Adult und n=2505 in LIFE-Adult), um zu testen, ob die Zusatz-Information aus ICA und ECA andere Ergebnisse liefert (gleiche Adjustierung und Modellannahmen wie für PS). Anschließend wurden die SNPs gefiltert auf MAF>1% in beiden Studien, info-Score>0.5 in beiden Studien und mindestens einer der p -Wert<0.1. Insgesamt wurden für PS 1,906,215 SNPs und für PS8 1,910,620 SNPs meta-analysiert. Ein LD-Pruning wurde mit der r^2 -Grenze 0.5 durchgeführt. Die Manhattan-Plots und die Summary-Statistiken beider Phänotypen sind im Anhang gezeigt (Tab. 12, Abb. 15 und Abb. 16).

Die Ergebnisse sind in vier Abschnitte gegliedert: zum einen ein Vergleich mit Hits aus der Publikation, getrennt danach, ob die Assoziation in der Subgruppe stärker oder schwächer wurde, bezogen sowohl auf die Effektstärke als auch Signifikanz. Zum anderen eine kurze Analyse neuer Hits mit suggestiver Signifikanz ($p<1\times 10^{-6}$), sowohl für PS als auch für PS8.

Verschlechterung im Case-Only Setting

Interessanterweise konnte für PS der genomweit signifikante Hit auf 9p21 nur mit nomineller Signifikanz repliziert werden (rs9644862: $\beta=0.044$, p -Wert=0.020, s. Tab. 6). Auch der Hit auf Chromosom 10, rs2902548, war signifikant assoziiert, aber nicht auf einem genomweiten Niveau ($\beta=-0.092$, p -Wert= 5.3×10^{-4}). Die beiden Suggestiv-Hits waren ebenfalls nur nominal signifikant assoziiert (rs141249844: p -Wert= 3.6×10^{-4} ; rs1349287: p -Wert= 3.8×10^{-3}).

Verbesserung im Case-Only Setting

Dem gegenüber stehen drei Hits, deren Assoziation im kontrollfreien Modell stärker wurden und damit zu Suggestiv-Hits zählen ($p<1\times 10^{-6}$, s. Tab. 6):

- rs9398699: dieser SNP liegt in der Nähe des Genes *FABP7* auf Chromosom 6, das für das „fatty acid-binding protein typ 7“ codiert, welches vor allem im Gehirn exprimiert wird. Er ist stärker assoziiert, aber wie schon in der vorherigen Analyse nicht durch andere SNPs in der Region unterstützt. Zusätzlich ist er eine seltene Variante (MAF etwa 1%).

- rs12064043: dieser SNP liegt in der Nähe von *SERBP1* auf Chromosom 1 und ist in LD mit einem Hit aus dem ersten Paper ($r^2=0.50$, rs12089923). Allerdings ist dieser SNP nicht durch andere Varianten unterstützt.
- rs11976578: dieser SNP liegt auf Chromosom 7 in einem Intron von *PTPRN2*. In der ersten Publikation war nicht dieser SNP assoziiert, aber einer im gleichen Gen (rs185796593). In Gegensatz dazu sind sieben weitere SNPs in hohem LD ($r^2>0.5$) ebenfalls assoziiert ($p<1\times 10^{-5}$, siehe Abbildung 17 im Anhang). Das Genprodukt *PTPRN2* ist ein Autoantigen, das mit Diabetes mellitus Typ 1 assoziiert ist [49]. Zusätzlich wurde das Gen in einer anderen GWAS mit Endpunkt Karotis-Kalzifikation in Diabetespatienten beschrieben [50].

Lokus	Gen	Publikation 1				Case-only Analyse			
		SNP	MAF	beta	p-Wert	SNP	MAF	beta	p-Wert
9p21	<i>CDKN2B-AS1</i>	rs9644862	0.43	0.131	8.7E-12	rs9644862	0.48	0.044	2.0E-02
10q24	<i>SFXN2</i>	rs2902548	0.18	-0.141	2.0E-08	rs2902548	0.17	-0.092	5.3E-04
6q13	<i>KCNQ5</i>	rs141249844	0.02	0.475	5.8E-07	rs141249844	0.02	0.341	3.6E-04
3p24	<i>KCNH8</i>	rs1349287	0.06	0.208	9.5E-07	rs1349287	0.06	0.125	3.8E-03
6q22.31	<i>FABP7</i>	rs9398699	0.01	-0.425	2.1E-06	rs9398699	0.01	-0.452	3.0E-07
1p31.3	<i>SERBP1</i>	rs12089923	0.36	0.094	2.0E-06	rs12064043	0.38	0.106	5.2E-07
7q36.3	<i>PTPRN2</i>	rs185796593	0.06	0.240	1.8E-06	rs11976578	0.21	0.136	3.3E-07

Tabelle 6: Vergleich der Top-Loki aus der ersten Publikation und der Case-Only Analyse zum Phänotyp PS. Grün markiert sind p-Wert kleiner als 1×10^{-6} , rot markiert sind p-Werte zwischen 0.05 und 1×10^{-6} . Die beiden genomweit-signifikanten Hits aus der ersten Publikation erreichen nicht die suggestive Signifikanz von $p<1\times 10^{-6}$. Für die Loci 1p31.3 und 7q36.3 sind die SNPs aufgrund des priority prunings unterschiedlich. Sie sind paarweise in LD ($r^2>0.5$).

Neue Hits für PS

Es konnten keine neuen genomweit signifikanten SNPs detektiert werden. Allerdings gab es zwei neue Loci auf Chromosom 1, die mit suggestiver Signifikanz mit PS assoziiert waren (s. Tab 7). Der eine liegt in der Nähe von *NEGR1*, ist allerdings nur schwach durch andere Varianten unterstützt. *NEGR1* wurde bereits als Locus für BMI und Übergewicht beschrieben [51]. Der zweite Hit liegt vor dem *APOA2* Gen. Die Assoziation würde biologisch Sinn ergeben, da die Plaqueentwicklung stark durch den Lipidmetabolismus beeinflusst wird. Aber auch dieser Hit ist nur schwach unterstützt und hat eine kleine MAF (2%).

Neue Hits für PS8

Auch für den Phänotyp PS8 konnten keine genomweit signifikanten SNPs gefunden werden. Hier wurden allerdings drei Varianten mit suggestiver Signifikanz detektiert, unter anderem ein Hit auf Chromosom 1 im Gen *PTPN22* (rs35025140: $\beta=-0.212$, p -Wert= 6.46×10^{-7} , s. Tab. 7). Neben dem Locus-Tophit gab es 5 Varianten in niedrigem LD ($r^2<0.5$, siehe Abbildung 18 im Anhang), die den Locus unterstützen und teilweise im Nachbar-Gen *MAGI3* liegen. Neben dem Support durch andere SNPs hat rs35025140 auch eine hohe MAF (23.4%). Zusätzlich ist der Top-Hit in LD mit der Variante rs12144309 (LD $r^2=0.54$), die kürzlich in Assoziation mit KHK berichtet wurde [52].

Ein zweiter interessanter Hit wurde auf Chromosom 6 detektiert (rs79398870: $\beta=0.327$, p -Wert= 8.45×10^{-7} , MAF=7.9%). Der SNP liegt in der Nähe des *GJA1* Gens, das für ein Gap Junction Protein kodiert (siehe Abbildung 19 im Anhang). Es gibt Hinweise, dass dieses Protein mit dazu beiträgt, atheromatöse Plaques in der ICA zu stabilisieren [53].

Ein weiterer Locus mit suggestiver Signifikanz lag auf Chromosom 16 (rs72807538: $\beta=0.537$, p -Wert= 8.42×10^{-7} , MAF=3.2%). Der Locus ist unterstützt durch zwei Varianten, allerdings erstreckt er sich über 1.25 Mb und enthält über 30 Gene (siehe Abbildung 20 im Anhang). Zusätzlich ist die MAF der SNPs niedrig (zwischen 2% und 5%).

Lokus	Gen (Distanz)	SNP	EAF	Phänotyp	beta	p-Wert
1p31.1	<i>NEGR1</i> (52 kb)	rs186841677	0.013	PS	-0.444	1.53E-07
1q23.3	<i>APOA2</i> (0.98 kb)	rs149905240	0.023	PS	0.328	6.83E-07
1p13.2	<i>PTPN22</i> (0 kb)	rs35025140	0.234	PS8	-0.212	6.46E-07
6q22.31	<i>GJA1</i> (45 kb)	rs79398870	0.079	PS8	0.327	8.45E-07
16q24.3	<i>ZNF276</i> (0 kb)	rs72807538	0.032	PS8	0.537	8.42E-07

Tabelle 7: Neu detektierte Loci in der GWAS zu PS und PS8 in Case-Only Setting mit p -Wert $< 1 \times 10^{-6}$. Für weitere SNPs und ausführlichere Annotation siehe Tab. 12 im Anhang.

Zusammenfassung

Insgesamt lässt sich feststellen, dass in der neuen Analyse bei keinem der beiden untersuchten Phänotypen ein genomweit signifikanter Hit detektiert werden konnte. Der Top-Locus der vorherigen Analyse, 9p21, war hier nur schwach mit PS assoziiert und der Effektschätzer deutlich kleiner. Dies ist ein klarer Hinweis darauf, dass nicht nur ein Powerverlust aufgrund der kleineren Fallzahl vorliegt, sondern dass dieser Locus zwar einen Einfluss auf die Prävalenz von Karotis-Plaques hat, aber nicht für den Schweregrad der Atherosklerose in der Karotis relevant ist. Dies steht im Einklang mit Zeller et al. [47], die zeigen konnten, dass der Effekt des 9p21 Locus für den Schweregrad schwächer ist als für die Prävalenz.

Im Gegensatz zu 9p21 gab es drei Loci, deren Assoziation mit PS sich in der neuen Analyse verstärkte, und zwei neue Loci, die in der vorherigen Analyse mit Kontrollen nicht gefunden wurden. Einige davon sind biologisch plausibel, wie 1q23.3 im *APOA2* Gen oder 7q36.3, der schon in Assoziation mit Karotis-Kalzifikation in Diabetespatienten beschrieben wurde. Dies ist ein weiterer Hinweis dafür, dass sich der genetische Einfluss auf Prävalenz und Schweregrad unterscheidet. Die Wahl einer Fall-Studie ist daher lohnenswert, allerdings wird eine höhere Fallzahl als in der vorliegenden Analyse benötigt, um genomweit signifikante bzw. valide Hits zu finden.

Die zusätzliche Analyse mit dem Phänotyp PS8 legt nahe, dass es unterschiedliche genetische Einflussgrößen gibt, je nach Lokalisation der Plaques. Auch hier ist allerdings eine Validierung bzw. Fallzahlsteigerung notwendig, um genomweit signifikante Varianten zu detektieren.

2.3.3 Haplotypen-Analyse des PCSK9 Locus

In der zweiten Publikation wurden vier unabhängige SNPs (paarweises $r^2 < 0.1$) im PCSK9 Locus detektiert, die mit dem PCSK9-Plasmaspiegel mit genomweiter Signifikanz assoziiert sind. In einer multivariaten Analyse waren drei davon immer noch genomweit signifikant assoziiert, sowohl in einem rein genetischen Modell (erklärte Varianz 4.3%) als auch in einem auf Alter, Geschlecht, Rauchen und Statin-Therapie adjustierten Modell (erklärte Varianz 19.4%). Da dieses Ergebnis eine allelische Heterogenität suggeriert, wurden die Haplotypen dieser drei SNPs noch einmal genauer betrachtet, um zu testen, ob es einen nicht-additiven Effekt, bzw. eine SNP-Interaktion gibt und ob die Verwendung der Haplotypen dadurch mehr Varianz des PCSK9-Spiegels erklären als die einzelnen SNPs im multivariaten Modell. Dazu wurden zuerst die häufigen Haplotypen detektiert und in einer Varianzanalyse verglichen.

Die Haplotypen der SNPs rs45448095, rs11591147 und 1:55520994 wurden mittels SHAPEIT geschätzt (SHAPEIT Version v2.r837, Genotyp-Wahrscheinlichkeitsschwelle 0.9) und auf Patienten mit PCSK9-Messung und vollständigen Kovariablen eingeschränkt (n=3290 Patienten, Kovariablen: Alter, Geschlecht, Rauchen und Statin-Therapie). Als Effekt-Allel wurde jeweils das Allel definiert, das zu niedrigeren PCSK9-Spiegeln beiträgt. Eine Übersicht der verwendeten SNPs und deren wichtigsten Kenngrößen ist in Tabelle 8 gegeben (Werte aus Publikation).

SNP_ID	Position	Referenz-Allel	Effekt-Allel	β	lg(p)	EAF	erklärte Varianz	Info-Score
rs45448095	55505447	C	T	-0.085	11.36	0.133	0.012	0.683
rs11591147	55505647	G	T	-0.315	16.71	0.012	0.018	0.581
1:55520994	55520994	GCG	G	-0.060	13.52	0.437	0.015	0.761

Tabelle 8: SNPs der PCSK9-GWAS, für die eine Haplotyp-Analyse durchgeführt wurde. Referenz- und Effekt-Allel wurden so angepasst, dass der β -Schätzer < 0 ist. Die Werte von β , lg(p), Frequenz des Effekt-Allels (EAF) und erklärter Varianz stammen aus der GWAS von Kapitel 2.2 mit allen Patienten (n=3290, adjustiert auf Alter, Geschlecht, Rauchen und Statin-Therapie).

Insgesamt wurden sechs von acht möglichen Haplotypen detektiert, wobei der häufigste Typ die Kombination „000“ ($\hat{=}$ Wildtyp) ist, also der Haplotyp ohne jegliches Effekt-Allel. Die Kombination der beiden Effekt-Allele von rs45448095 und rs11591147 wurde nicht gefunden (Haplotypen „110“ und „111“). Grund hierfür ist sowohl die geringe EAF der beiden SNPs, wodurch die erwartete Häufigkeit gering ist (\hat{n} =10.7 bei 6716 Haplotypen), als auch der verhältnismäßig schlechte Info-Score (Info-Score=0.581 für rs11591147, Info-Score=0.683 für rs45448095). Die Schwelle für die Haplotypen-Schätzung lag bei 0.9, das heißt, Genotypen mit einer niedrigeren Wahrscheinlichkeit wurden auf fehlend gesetzt. Dies ist bei SNPs mit niedrigem Info-Score häufiger der Fall. Die anderen beiden Haplotypen mit Effekt-Allel des Top-Hits rs11591147 sind selten (Haplotypen „010“ und „011“, Häufigkeit <1%). Zur Vereinfachung wurden daher diese beiden Haplotypen als „rare types“ (X1X) zusammengefasst (s. Tab. 9). Die anderen vier möglichen Haplotypen ohne rs11591147-T waren häufig.

Haplotyp	Häufigkeit	ln(PCSK9)
rare = X1X	53	5.17
101	384	5.34
100	275	5.39
001	2123	5.43
000	3745	5.47

Tabelle 9: Tabelle der detektierten Haplotypen und zugehörigen mittleren PCSK9-Werten. SNPs nach Position geordnet (rs45448095, rs11591147, 1:55520994). Alle SNPs wurden so codiert, dass „1“ jeweils das Allel mit PCSK9 reduzierendem Effekt ist (grün markiert), das andere Allel „0“. Die PCSK9-Werte wurden zuerst mit dem natürlichen Logarithmus transformiert. Anschließend wurde das arithmetische Mittel pro Haplotyp berechnet.

Jedem Haplotyp wurde ein PCSK9-Wert zugeordnet (Allel-Counting Technik). Um die Haplotypen zu sortieren, wurde der Mittelwert pro Typ bestimmt und diese anschließend aufsteigend geordnet. Wie erwartet haben die seltenen Varianten „X1X“ den niedrigsten Mittelwert, da sich hier der Effekt des Top-Hits durchsetzt. Die übrigen vier sortieren sich nach der Anzahl der Effekt-Allele, wobei der Effekt von rs45448095-T stärker ist als der von 1:55520994-G (s. Tab. 9).

Um die Gruppenunterschiede zu testen, wurde eine einfaktorielle Varianzanalyse (ANVOA) durchgeführt, adjustiert auf Alter, Geschlecht, Statin-Medikation und Raucherstatus (n=3290 Patienten, bzw. 6580 Haplotypen). Mit einem Post-hoc Test (Tukey-HSD) wurde geprüft, zwischen welchen Gruppen es zu signifikanten Mittelwertunterschieden kam. Das Ergebnis der Post-hoc Analyse ist in Tabelle 10 und Abbildung 14 zusammengefasst. Sowohl die seltenen Typen „X1X“ als auch der Wildtyp „000“ unterscheiden sich signifikant von allen anderen Gruppen (p -Werte zwischen 3.09×10^{-4} und 7.95×10^{-13}). Von den drei mittleren Gruppen unterscheiden sich nur „101“ und „001“ signifikant voneinander ($p=5.68 \times 10^{-5}$), während die Differenz von „101“ zu „100“, bzw. „100“ zu „001“ nicht signifikant ist. Auch die p -Wert Adjustierung nach Holm erzeugte das gleiche Ergebnis bezüglich der signifikanten Unterschiede.

	rare	101	100	001	000
rare	-	0.194	0.231	0.264	0.303
101	1.98E-05	-	0.036	0.070	0.109
100	3.62E-07	0.467	-	0.034	0.072
001	9.86E-11	5.68E-05	0.320	-	0.039
000	7.95E-13	4.41E-12	3.09E-04	3.47E-06	-

Tabelle 10: Ergebnisse der Tukey-HSD Analyse. Die Differenz der Gruppen-Mittelwerte sind im oberen Dreieck eingetragen, die zugehörigen p -Werte sind im unteren Dreieck notiert. Signifikante p -Werte und die zugehörigen Differenzen sind rot unterlegt.

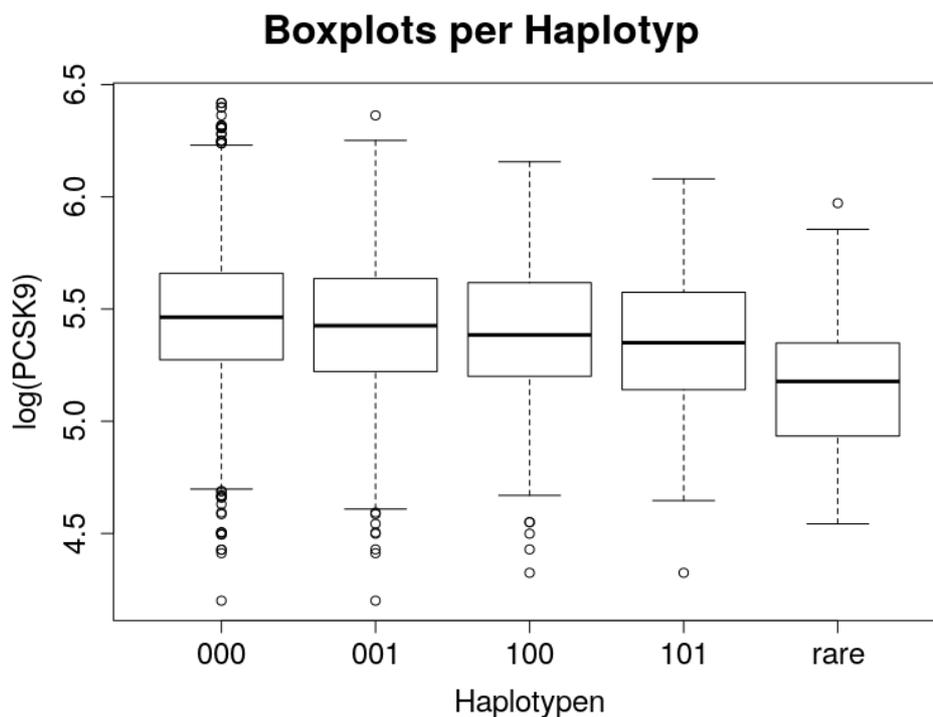


Abbildung 14: Verteilung der log-transformierten PCSK9-Werte, aufgeteilt nach den fünf Haplotypen.

Zusammenfassung

Die allelische Heterogenität im *PCSK9* Locus konnte bestätigt werden. Es wurden jedoch weder SNP-Interaktionen noch nicht-additive Effekte detektiert. In einem einfachen linearen Modell erklärten die Haplotypen 2.7% der Varianz von PCSK9 und damit weniger als die einzelnen drei SNPs in einem multivariaten Modell (4.3%).

Das Effekt-Allel von rs11591147 hat einen starken alleinstehenden Effekt, der die Effekte der anderen SNPs überlagert. Dies spiegeln auch die Effektschätzer der GWAS wider: das β von rs11591147 ist etwa viermal so groß wie das der beiden anderen SNPs. Es ist daher plausibel, die Haplotypen mit rs11591147-T in der Gruppe „X1X“ zusammenzufassen.

Die anderen beiden Varianten hatten ähnliche Beta-Schätzer. Jede einzelne Variante führt zu einer PCSK9-Reduktion, wobei es keinen signifikanten Unterschied zwischen dem Haplotyp „100“ und „001“ gibt, wohl aber zwischen den jeweiligen Gruppen und dem Wildtyp „000“. Die Kombination der beiden Varianten im Typ „101“ führt zu einer weiteren PCSK9-Reduktion, wobei der Unterschied hier zwischen „101“ und „100“ nicht signifikant ist. Grund dafür kann die verhältnismäßig geringe Fallzahl sein.

3 Schlussfolgerungen und Ausblick

Obwohl Karotis-Plaques wichtige Risikofaktoren der KHK und des Schlaganfalls sind [4, 54], ist die Anzahl bekannter genetischer Assoziationen gering. Ursache dafür ist unter anderem die Heterogenität der Messmethoden bzw. der Definitionen sowie die bisher geringe Fallzahl entsprechender genomweiter Studien. Im Rahmen dieser Arbeit wurden zwei Studien verwendet, für die die Karotis standardisiert vermessen und ein Plaques-Score PS definiert wurde. Dieser Score hat drei Vorteile: Er kann nicht-invasiv bestimmt und je nach Fragestellung auf die Plaqueprävalenz reduziert oder auf den feinstufigeren Score PS8 erweitert werden. Zusätzlich kann er durch Ausschluss von gesunden Probanden den Schweregrad der Belastung mit Karotis-Plaques darstellen. Mit diesem Score wurden verschiedene Analysen zur Suszeptibilität für Karotis-Plaques bzw. Schwere der Belastung durchgeführt.

Die Diskussion der Ergebnisse dieser Arbeit erfolgte bereits ausführlich in den jeweiligen Publikationen. Hier werden kurz die eingangs gestellten Fragen aufgegriffen und beantwortet. Abschließend wird ein Ausblick für künftig geplante Projekte gegeben.

Assoziation genetischer Varianten mit Karotis-Plaques

Zur Detektion genetischer Varianten wurde die erste große Meta-GWAS zu Karotis-Plaque Belastung mit dem Phänotyp PS ausgeführt, die statt der reinen Prävalenz die Anzahl der Plaques betrachtet. Hier konnten zwei genomweite Hits beschrieben werden. Zusätzlich gab es einige genetische Varianten, die mit Karotis-Plaques zumindest suggestiv signifikant assoziiert sind. Dazu gehörten auch bekannte KHK-Loki. Im Vergleich der Studiendesigns konnte man sehen, dass es unterschiedliche Marker für die Plaque-Prävalenz und den Schweregrad gibt und dass der Effekt von der Lokalisation der Plaques abhängt.

Hier kann eine erste Parallele zur KHK gezogen werden: die Heterogenität der SNPs weist auf viele unabhängige Loki mit jeweils kleinen Effekten hin, die in Summe das Risiko für Karotis-Plaques darstellen. Mit steigender Fallzahl werden hier sicherlich noch weitere Loki hinzukommen.

Effekt der KHK-Loki auf Karotis-Plaques

In den Enrichment-Analysen wurde gezeigt, dass ein nicht-zufälliger Anteil der KHK-Loki mit PS signifikant assoziiert ist. Die Effekte der Loki auf PS waren schwächer als die auf KHK, aber es konnte eine signifikante Korrelation der Effekte nachgewiesen werden. Dieser Überlapp an ähnlichen Effekten gleicher Loki deutet auf einen gemeinsamen genetischen Pathomechanismus von KHK und Karotis-Plaques hin. Die KHK-Loki können daher als potentielle Karotis-Loki betrachtet werden, die bislang aufgrund mangelnder Power noch nicht detektiert werden konnten.

Kausale Wirkung von PCSK9 auf Atherosklerose

Es wurde eine GWAS zum PCSK9-Plasmaspiegel durchgeführt, wobei erstmals genomweit signifikante Hits detektiert werden konnten. Zusätzlich fungierten die vier neu beschriebenen, unabhängigen Varianten im *PCSK9* Locus als Instrumente in einer Mendelschen Randomisierung, womit erstmalig der PCSK9-Plasmaspiegel als Risikofaktor mit kausaler Wirkung auf Karotis- und Koronargefäße identifiziert wurde.

Es zeigt sich, dass PCSK9 als Genprodukt des KHK-Lokus *PCSK9* einen kausalen Effekt auf Atherosklerose in der Karotis hat, obwohl dieser Locus noch nicht als Karotis-Locus beschrieben wurde und auch in der Enrichment-Analyse nicht signifikant mit PS assoziiert war. Ein plausibler Grund dafür ist die relativ geringe Fallzahl. Dies steht im Einklang mit der Enrichment-Analyse, dass KHK-Loci potentielle Karotis-Loci darstellen. Basierend auf diesen Ergebnissen kann man vermuten, dass es neben dem Überlapp genetischer Loci auch einen Überlapp kausaler Zusammenhänge gibt. Diese Annahme muss noch in unabhängigen Studien mit weiteren Loci und Kausalketten geprüft werden.

Ausblick

In der vorliegenden Arbeit konnte festgestellt werden, dass es unterschiedliche genetische Faktoren gibt, die die Suszeptibilität für Karotis-Plaques beeinflussen, und dass die Stärke der detektierten SNPs davon abhängt, welchen Phänotypen man analysiert. Die Analysen waren deshalb so aussagekräftig, da in beiden verwendeten Studien die gleiche Messmethode gemäß einem einheitlichen standardisierten Vorgehen angewandt wurde.

Zusätzlich konnte eine Anreicherung von KHK-Loci festgestellt werden, die auch mit PS assoziiert sind. Dies weist auf gemeinsame Pathomechanismen hin. Mehrere biologisch interessante neue Varianten erreichten in den dargestellten Analysen suggestive Signifikanz. Diese müssen in weiteren, unabhängigen Studien validiert bzw. in größeren Meta-Analysen bestätigt werden.

Um die kausalen Beziehungen weiter zu analysieren, ist eine Meta-Analyse zu PCSK9-Spiegeln mit der LURIC-Studie [55] aktuell in Planung. Des Weiteren ist ein Meta-GWAS zu mehreren Sexual-Hormonen und Steroiden geplant, die Aufschluss über den kausalen Einfluss von Hormonen bringen soll.

Neben der Validierung der gefundenen Varianten ist ein weiterer offener Themenkomplex der genetische Einfluss auf die Plaquefläche bzw. den Verschlussgrad der Karotis und die Plaqueentwicklung. In der LIFE-Adult Studie werden aktuell die Plaqueflächen in der Karotis in etwa 1200 Probanden erfasst. Zusätzlich soll im Follow-up die Karotissonographie wiederholt werden. Die dadurch gewonnenen Längsschnitt-Daten erlauben Aussagen über das Plaquewachstum in einem Zeitraum von etwa 10 Jahren.

Nachwort

*„Für die genetisch Überlegenen ist der Erfolg leichter zu erreichen,
aber keinesfalls garantiert,
schließlich gibt es kein Gen für das Schicksal.“*

Gattaca. Andrew Niccol. US 1997. TC 00:23:06 [56]

Im Science-Fiction Film „Gattaca“ [56] können Embryos mittels In-Vitro-Befruchtung und Gentechnik so modifiziert werden, dass sie keinerlei Veranlagungen für alle erdenklichen Krankheiten in ihrem Genom tragen. Durch diese Optimierung haben sie später höhere Erfolgchancen, sowohl privat als auch im Beruf. Trotzdem ist ihnen der Erfolg nicht garantiert, was im Film anhand der natürlich gezeugten Protagonisten dargestellt wird.

Mit der Genschere CRISPR-Cas9 [57] ist die Wissenschaft der im Film beschriebenen Gentechnik einen Schritt näher gekommen, aber wie hier gezeigt wurde, ist die alleinige Optimierung der KHK-Loki keine Garantie für ein gesundes, atherosklerosefreies Leben, da das nichtgenetische Risiko einen großen Anteil am Krankheitsausbruch und –verlauf hat. Es lassen sich lediglich die Chancen auf ein gesundes Leben verbessern.

4 Zusammenfassung der Arbeit

Dissertation zur Erlangung des akademischen Grades:

Dr. rer. nat.

Titel:

Molekulargenetische Faktoren der Suszeptibilität für Karotis-Plaques

Eingereicht von:

M. Sc. Janne Pott

Angefertigt am:

Institut für medizinische Informatik, Statistik und Epidemiologie der medizinischen Fakultät der Universität Leipzig

Betreut von:

Prof. Dr. Markus Scholz

Datum der Einreichung:

Mai 2018

Biologischer Hintergrund

Subklinische atherosklerotische Veränderungen der Karotis wie Karotis-Plaques stellen einen unabhängigen Risikofaktor der koronaren Herzerkrankung (KHK) und des Schlaganfalls dar. Sowohl KHK als auch Karotis-Plaques sind Formen der Atherosklerose und ein gemeinsamer genetischer Pathomechanismus wird vermutet. Ein Vorteil der Ermittlung des Karotisstatus ist die Möglichkeit der nicht-invasiven Untersuchung mittels Ultraschall. Ein Nachteil ist ein fehlender Goldstandard zur Karotis-Messung. Als Konsequenz ist die Phänotyp-Definition zwischen Studien sehr heterogen. Dies ist ein Grund für die bislang geringe Zahl von identifizierten genetischen Markern, die Plaquentstehung in der Karotis beeinflussen.

Ein Risikofaktor der Atherosklerose ist die Störung des Lipidstoffwechsels. Das Proprotein Konvertase Subtilisin/Kexin Typ 9 (PCSK9) spielt hier eine wichtige regulatorische Rolle, indem es die Anzahl von LDL-Rezeptoren verringert. Dadurch steigt mit einem hohen PCSK9-Spiegel das Risiko für kardiovaskuläre Ereignisse. Trotz seiner klinischen Relevanz wurden noch keine genomweit signifikanten Varianten beschrieben, die einen Einfluss auf den PCSK9-

Plasmaspiegel haben. Zudem gibt es keine Studien zu kausalen Effekten von PCSK9 auf Atherosklerose in der Karotis.

Fragestellungen

Das Hauptziel dieser kumulativen Dissertation ist die Klärung der Frage, in wieweit Genetik einen Einfluss auf die Anfälligkeit für Karotis-Plaques hat. Folgende drei Fragestellungen wurden dazu bearbeitet:

- 1) Welche genetischen Marker sind mit Karotis-Plaques assoziiert?
- 2) Inwiefern gibt es Überschneidungen zwischen Loki, die mit KHK bzw. Karotis-Plaques assoziieren?
- 3) Gibt es eine kausale Wirkung von PCSK9 auf Karotis-Plaques?

Methoden

In der Studie LIFE-Adult und im KHK-Verdachtskollektiv der LIFE-Heart Studie wurden Karotis-Sonographien nach gleichem Protokoll durchgeführt. Darauf aufbauend wurde ein Karotis-Plaque-Score PS definiert, der die Summe der Plaques in *Arteria carotis communis* (CCA) und *Sinus caroticus* (Bulbus) darstellt. Zusätzlich wurde die Intima-Media-Dicke der Karotis (cIMT) bestimmt und ein weiterer Score (PS8), der zusätzlich auch die *A. carotis interna* (ICA) und *A. carotis externa* (ECA) umfasst.

Der Score PS wurde in je zwei genomweite Assoziationsstudien (GWAS) in beiden Kohorten als Phänotyp analysiert, einmal in einem Fall-Kontroll-Ansatz (PS von 0 bis 4, n=4037 bzw. n=3152 in LIFE-Adult bzw. LIFE-Heart) und einmal in einem Fallansatz (PS von 1 bis 4, n=2004 bzw. n=2585 in LIFE-Adult bzw. LIFE-Heart), um nur die Schwere der Atherosklerose zu analysieren. Für beide Ansätze wurden die Ergebnisse in einer Meta-Analyse zusammengefasst. Das Fall-Setting wurde für PS8 wiederholt.

Um eine Anreicherung von KHK-SNPs festzustellen, die auch mit PS assoziiert sind, wurden Enrichment-Analysen mit 58 bekannten KHK-Loki durchgeführt. Neben der Einzel-Analyse der SNPs wurde zusätzlich ein genetischer Risiko-Score (GRS) aufgestellt. Auch die Assoziation von Varianten in hohem LD mit den KHK-Loki wurden berücksichtigt und mittels Permutationstest auf Überzufälligkeit signifikanter Assoziationen mit PS getestet.

Die Kausalität von PCSK9 auf Karotis-Plaques wurde durch Mendelsche Randomisierung überprüft. Dafür wurden zunächst mittels einer GWAS zu den PCSK9-Plasmaspiegeln in der LIFE-Heart Studie geeignete Instrumente gesucht (n=3290). Die vier detektierten genomweit signifikanten und unabhängigen Varianten wurden auf allelische Heterogenität getestet und anschließend als Instrumente in der Mendelschen Randomisierung verwendet.

Ergebnisse

Die Ergebnisse der beiden Meta-GWAS unterschieden sich deutlich. Im Fall-Kontroll-Setting wurde die Prävalenz mitbetrachtet, und es konnten zwei genomweit signifikante Hits beschrieben werden ($p < 5 \times 10^{-8}$). Der Top-Hit liegt im Locus 9p21 (rs9644862, $\beta = 0.131$, $p = 8.7 \times 10^{-12}$), der bekannt ist für seine Assoziation mit KHK. In der zweiten GWAS ohne Kontrollen war diese Variante nur schwach assoziiert und hatte einen deutlich schwächeren Effekt ($\beta = 0.044$, $p = 0.020$).

In der Fall-Meta-GWAS, die nur den Schweregrad der Belastung mit Karotis-Plaques darstellt, konnten keine genomweit signifikanten Varianten detektiert werden, dafür aber fünf mit suggestiver Signifikanz ($p < 1 \times 10^{-6}$), die in der Fall-Kontroll-Analyse nicht oder nur schwach assoziiert waren. In einer zusätzlichen GWAS zum Phänotyp PS8 wurden weitere Varianten mit suggestiver Signifikanz detektiert.

Die Enrichment-Analyse von KHK-Loki ergab überzufällig viele nominal signifikante Assoziationen mit PS und konkordanter Effektrichtung ($OR = 4.6$, $p = 3.2 \times 10^{-3}$). Dies konnte im Permutationstest und durch die Analyse des GRS bestätigt werden (Permutationstest: $p = 1.0 \times 10^{-8}$, lineare Regression von GRS auf PS: $p = 2.0 \times 10^{-6}$). In einem Vergleich zwischen LIFE-Adult und LIFE-Heart konnte zudem gezeigt werden, dass der GRS in LIFE-Heart tendenziell größer ist ($p = 0.068$). Zusätzlich wurde hier ein signifikanter Trend hin zu schlechteren Lebensstilen in der LIFE-Heart festgestellt (Armitage Trend-Test, $p < 2 \times 10^{-16}$).

In der GWAS zum PCSK9-Plasma-Spiegel wurden zwei genomweit signifikante Loci detektiert. Einer davon war der PCSK9-Genlocus, in dem vier unabhängige SNPs (LD $r^2 < 0.1$) mit genomweiter Signifikanz detektiert wurden. Der Top-Hit rs11591147 kodiert die Missense-Mutation R46L ($\beta = -0.315$, $p = 1.9 \times 10^{-17}$). In einem multivariaten Modell unter Berücksichtigung nichtgenetischer Risikofaktoren erklärten die vier PCSK9-SNPs zusammen 4.4% der Varianz. In einer Haplotypen-Analyse konnte eine allelische Heterogenität bestätigt werden. Jedoch erklären die Haplotypen weniger Varianz als das multivariate Modell (2.7%).

Mittels Mendelscher Randomisierung konnte festgestellt werden, dass PCSK9 nicht nur einen kausalen Effekt auf KHK-Prävalenz ($\beta = 0.960$, $p = 0.048$) bzw. -Schweregrad ($\beta = 1.223$, $p = 0.005$) hat, sondern auch auf Karotis-Plaques ($\beta = 0.605$, $p = 0.023$). Ein signifikanter Effekt auf cIMT konnte nicht gezeigt werden ($\beta = 0.041$, $p = 0.061$). In einem Kontroll-Setting ohne Patienten mit Statin-Medikation verstärkte sich der Effekt auf die beiden Karotis-Phänotypen (PS: $\beta = 0.769$, $p = 0.008$; cIMT: $\beta = 0.045$, $p = 0.054$).

Schlussfolgerung

Der Plaque-Score PS ist ein Phänotyp, der nicht-invasiv bestimmt werden kann und je nach Fragestellung modifiziert werden kann: Er kann um die *A. carotis interna* (ICA) und *A. carotis externa* (ECA) erweitert werden, wodurch er feinstufiger wird und das Risiko für Schlaganfall besser abbildet, oder auf die reine Plaque-Prävalenz ($PS > 0$) reduziert werden. Durch

Ausschluss von Kontrollen kann damit zusätzlich auch der Schweregrad der Karotis-Plaques-Belastung untersucht werden.

Anhand dieser Scores konnte gezeigt werden, dass das Auftreten von Karotis-Plaques genetischen Einflüssen unterliegt, die je nach untersuchtem Phänotyp und Studiendesign unterschiedlich stark ausgeprägt sind. Da viele der Loci jedoch noch keine genomweite Signifikanz erreichten, ist eine Validierung mit höherer Fallzahl und in unabhängigen Studien notwendig.

Des Weiteren konnte eine überzufällige Anreicherung von KHK-Loci festgestellt werden, die ebenfalls einen Einfluss auf die Suszeptibilität auf Karotis-Plaques haben. Dies lässt gemeinsame genetische Pathomechanismen vermuten. Die KHK-Loci können daher als potentielle Karotis-Loci betrachtet werden, die bislang aufgrund mangelnder Power noch nicht detektiert werden konnten.

Neben den oben beschriebenen Assoziationen konnte gezeigt werden, dass PCSK9 als Genprodukt eines KHK-Risiko-Locus einen kausalen Effekt auf Karotis-Plaques hat. Es ist daher plausibel, PCSK9 auch als Risiko-Locus für Karotis-Plaques zu betrachten, der aufgrund niedriger Fallzahlen noch nicht mit genomweiter Signifikanz detektiert werden konnte.

5 Literaturverzeichnis

1. Bärtsch A. *Zum Glück - loslassen*. 1st ed. Frauenfeld: Liebig; 2007.
2. Kuhlmann I, Chin D, Rimbach G. *Prävention kardiovaskulärer Erkrankungen und Atherosklerose*. Wiesbaden: Springer Fachmedien Wiesbaden; 2014.
3. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47:1121–30. doi:10.1038/ng.3396.
4. Weissgerber A, Scholz M, Teren A, Sandri M, Teupser D, Gielen S, et al. The value of noncoronary atherosclerosis for identifying coronary artery disease: results of the Leipzig LIFE Heart Study. *Clin Res Cardiol*. 2016;105:172–81. doi:10.1007/s00392-015-0900-x.
5. Bis JC, Kavousi M, Franceschini N, Isaacs A, Abecasis GR, Schminke U, et al. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet*. 2011;43:940–7. doi:10.1038/ng.920.
6. Pan S. Molecular mechanisms responsible for the atheroprotective effects of laminar shear stress. *Antioxid Redox Signal*. 2009;11:1669–82. doi:10.1089/ars.2009.2487.
7. Lusis AJ. Atherosclerosis. *Nature*. 2000;407:233–41. doi:10.1038/35025203.
8. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. *Circ Res*. 2016;118:564–78. doi:10.1161/CIRCRESAHA.115.306566.
9. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med*. 2016;375:2349–58. doi:10.1056/NEJMoa1605086.
10. Jeon H, Blacklow SC. Structure and physiologic function of the low-density lipoprotein receptor. *Annu Rev Biochem*. 2005;74:535–62. doi:10.1146/annurev.biochem.74.082803.133354.
11. Catapano AL, Graham I, Backer G de, Wiklund O, Chapman MJ, Drexel H, et al. 2016 ESC/EAS Guidelines for the Management of Dyslipidaemias. *Eur Heart J*. 2016;37:2999–3058. doi:10.1093/eurheartj/ehw272.
12. Taylor F, Huffman MD, Macedo AF, Moore THM, Burke M, Davey Smith G, et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2013:CD004816. doi:10.1002/14651858.CD004816.pub5.
13. Bergeron N, Phan BAP, Ding Y, Fong A, Krauss RM. Proprotein convertase subtilisin/kexin type 9 inhibition: a new therapeutic mechanism for reducing cardiovascular disease risk. *Circulation*. 2015;132:1648–66. doi:10.1161/CIRCULATIONAHA.115.016080.
14. Robinson JG, Farnier M, Krempf M, Bergeron J, Luc G, Averna M, et al. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. *N Engl J Med*. 2015;372:1489–99. doi:10.1056/NEJMoa1501031.
15. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med*. 2017;376:1713–22. doi:10.1056/NEJMoa1615664.
16. Alberts B, Hunt T, Wilson J. *Molecular biology of the cell*. 5th ed. Abingdon: Garland Science; 2008.
17. World Health Organization. Top 10 causes of death. 2015. http://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/. Accessed 30 Apr 2018.

18. Loeffler M, Engel C, Ahnert P, Alfermann D, Arelin K, Baber R, et al. The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health*. 2015;15:691. doi:10.1186/s12889-015-1983-z.
19. Beutner F, Teupser D, Gielen S, Holdt LM, Scholz M, Boudriot E, et al. Rationale and design of the Leipzig (LIFE) Heart Study: phenotyping and cardiovascular characteristics of patients with coronary artery disease. *PLoS ONE*. 2011;6:e29070. doi:10.1371/journal.pone.0029070.
20. Beutner F, Henger S, Heise C. Karotissonographie: Standard Operating Procedure (SOP) 2011.
21. Stein JH, Korcarz CE, Hurst RT, Lonn E, Kendall CB, Mohler ER, et al. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Endorsed by the Society for Vascular Medicine. *J Am Soc Echocardiogr*. 2008;21:93-111; quiz 189-90. doi:10.1016/j.echo.2007.11.011.
22. Blausen.com staff. Medical gallery of Blausen Medical 2014. *Wiki J Med* 2014. doi:10.15347/wjm/2014.010.
23. Bommas-Ebert U, Teubner P, Voß R, Krahn V. *Kurzlehrbuch Anatomie und Embryologie ; [mit Muskeltrainer]*. 3rd ed. Stuttgart [u.a.], Stuttgart [u.a.]: Thieme; 2011.
24. Affymetrix I. *Axiom Genotyping Solution Data Analysis Guide*.
25. Wang J. An Estimator for Pairwise Relatedness Using Molecular Markers. *Genetic Society of America*. 2002:1203–15.
26. König IR, Loley C, Erdmann J, Ziegler A. How to include chromosome X in your genome-wide association study. *Genet Epidemiol*. 2014;38:97–103. doi:10.1002/gepi.21782.
27. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74. doi:10.1038/nature15393.
28. Delaneau O, Howie B, Cox AJ, Zagury J-F, Marchini J. Haplotype estimation using sequencing reads. *Am J Hum Genet*. 2013;93:687–96. doi:10.1016/j.ajhg.2013.09.002.
29. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5:e1000529. doi:10.1371/journal.pgen.1000529.
30. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39:906–13. doi:10.1038/ng2088.
31. Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, et al. Ensembl 2017. *Nucleic Acids Res*. 2017;45:D635-D642. doi:10.1093/nar/gkw1104.
32. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017;45:D896-D901. doi:10.1093/nar/gkw1133.
33. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648–60. doi:10.1126/science.1262110.
34. Yu G, Wang L-G, Yan G-R, He Q-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*. 2015;31:608–9. doi:10.1093/bioinformatics/btu684.
35. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet*. 2013;14:379–89. doi:10.1038/nrg3472.
36. Sillesen H, Muntendam P, Adourian A, Entekin R, Garcia M, Falk E, Fuster V. Carotid plaque burden as a measure of subclinical atherosclerosis: comparison with other tests for subclinical arterial disease in the High Risk Plaque BioImage study. *JACC Cardiovasc Imaging*. 2012;5:681–9. doi:10.1016/j.jcmg.2012.03.013.

37. Burgess S, Thompson SG. Mendelian randomization: Statistical methods for using genes to estimate causal estimation. Boca Raton: Chapman & Hall/CRC; 2015.
38. Wehby GL, Ohsfeldt RL, Murray JC. 'Mendelian randomization' equals instrumental variable analysis with genetic instruments. *Stat Med.* 2008;27:2745–9. doi:10.1002/sim.3255.
39. Mendel G. Versuche über Pflanzenhybriden. *Verhandlungen des Naturforschenden Vereines in Brünn.* 1866:3–47.
40. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology.* 2000;29:722–9. doi:10.1093/ije/29.4.722.
41. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. *J Clin Epidemiol.* 2016;69:208–16. doi:10.1016/j.jclinepi.2015.08.001.
42. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med.* 2016;35:1880–906. doi:10.1002/sim.6835.
43. Pott J, Burkhardt R, Beutner F, Horn K, Teren A, Kirsten H, et al. Genome-wide meta-analysis identifies novel loci of plaque burden in carotid artery. *Atherosclerosis.* 2017;259:32–40. doi:10.1016/j.atherosclerosis.2017.02.018.
44. Pott J, Schlegel V, Teren A, Horn K, Kirsten H, Bluecher C, et al. Genetic Regulation of PCSK9 (Proprotein Convertase Subtilisin/Kexin Type 9) Plasma Levels and Its Impact on Atherosclerotic Vascular Disease Phenotypes. *Circ Genom Precis Med.* 2018;11:e001992. doi:10.1161/CIRCGEN.117.001992.
45. Dron JS, Hegele RA. Complexity of mechanisms among human proprotein convertase subtilisin-kexin type 9 variants. *Curr Opin Lipidol.* 2017;28:161–9. doi:10.1097/MOL.0000000000000386.
46. Dewpura T, Raymond A, Hamelin J, Seidah NG, Mbikay M, Chrétien M, Mayne J. PCSK9 is phosphorylated by a Golgi casein kinase-like kinase ex vivo and circulates as a phosphoprotein in humans. *FEBS J.* 2008;275:3480–93. doi:10.1111/j.1742-4658.2008.06495.x.
47. Zeller T, Seiffert M, Müller C, Scholz M, Schäffer A, Ojeda F, et al. Genome-Wide Association Analysis for Severity of Coronary Artery Disease Using the Gensini Scoring System. *Front Cardiovasc Med.* 2017;4:57. doi:10.3389/fcvm.2017.00057.
48. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc.* 2003;35:1381–95. doi:10.1249/01.MSS.0000078924.61453.FB.
49. Li Q, Borovitskaya AE, DeSilva MG, Wasserfall C, Maclaren NK, Notkins AL, Lan MS. Autoantigens in insulin-dependent diabetes mellitus: molecular cloning and characterization of human IA-2 beta. *Proc Assoc Am Physicians.* 1997;109:429–39.
50. Divers J, Palmer ND, Langefeld CD, Brown WM, Lu L, Hicks PJ, et al. Genome-wide association study of coronary artery calcified atherosclerotic plaque in African Americans with type 2 diabetes. *BMC Genet.* 2017;18:105. doi:10.1186/s12863-017-0572-9.
51. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518:197–206. doi:10.1038/nature14177.
52. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res.* 2018;122:433–43. doi:10.1161/CIRCRESAHA.117.312086.
53. Nakase T, Ishikawa T, Miyata H. Protective effects of connexins in atheromatous plaques in patients of carotid artery stenosis. *Neuropathology.* 2017;37:97–104. doi:10.1111/neup.12345.

54. Chambless LE, Folsom AR, Clegg LX, Sharrett AR, Shahar E, Nieto FJ, et al. Carotid wall thickness is predictive of incident clinical stroke: the Atherosclerosis Risk in Communities (ARIC) study. *Am J Epidemiol.* 2000;151:478–87.
55. Winkelmann BR, März W, Boehm BO, Zotz R, Hager J, Hellstern P, Senges J. Rationale and design of the LURIC study--a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics.* 2001;2:S1-73.
doi:10.1517/14622416.2.1.S1.
56. Niccol A. *Gattaca: Gattaca*; 1997.
57. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012;337:816–21.
doi:10.1126/science.1225829.

6 Anlagen

In diesem Kapitel sind die Manhattan-Plots, die ausführliche Tabelle zur Meta-Analyse und die RA-Plots aus Kapitel 2.3.2 aufgeführt.

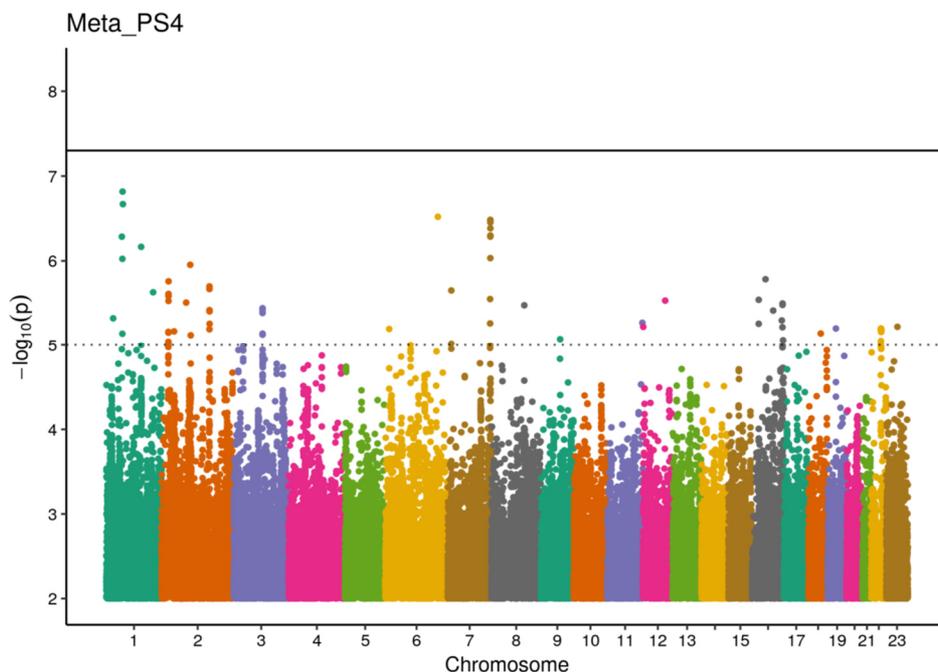


Abbildung 15: Manhattan-Plot der Meta-Analyse zum Phänotyp PS (s. Kapitel 2.3.2). Die durchgezogene Linie markiert die genomweite Signifikanz ($p < 5 \times 10^{-8}$), die gestrichelte Linie entspricht $p < 1 \times 10^{-5}$. Es gibt keine genomweit signifikanten Hits, aber einige SNPs mit suggestiver Signifikanz ($p < 1 \times 10^{-6}$).

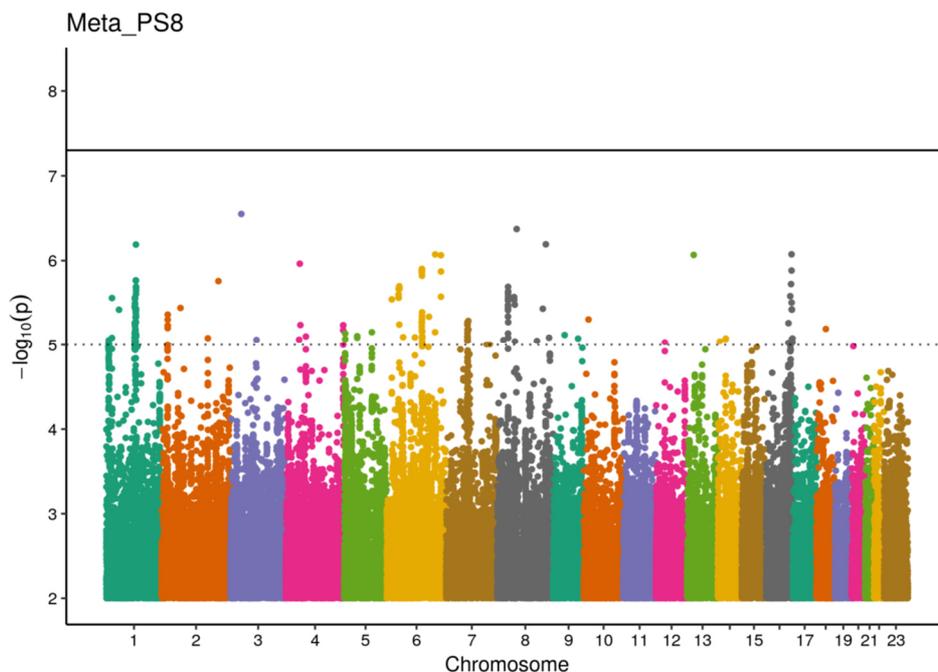


Abbildung 16: Manhattan-Plot der Meta-Analyse zum Phänotyp PS8 (s. Kapitel 2.3.2). Die durchgezogene Linie markiert die genomweite Signifikanz ($p < 5 \times 10^{-8}$), die gestrichelte Linie entspricht $p < 1 \times 10^{-5}$. Es gibt keine genomweit signifikanten Hits, aber einige SNPs mit suggestiver Signifikanz ($p < 1 \times 10^{-6}$).

Anlagen

SNP	Chr:Pos	Effekt-Allel	EAF	benachbarte Gene	Phäno- typ	N	info	β	SE	p-Wert	tau	Q
rs186841677:72800505:A:G	1:72800505	G	0.013	NEGR1; GDI2P2	PS	4589	0.869	-0.444	0.085	1.53E-07	0.13	2.09
rs190069629:37967657:A:G	3:37967657	G	0.013	CTDSP1; MIR26A1; VILL; PLCD1	PS8	4098	0.965	0.780	0.152	2.82E-07	0.00	0.73
rs9398699:123076113:C:A	6:123076113	A	0.989	FABP7; PKIB; SMPDL3A; ATP5LP2	PS	4589	1	-0.452	0.088	3.02E-07	0.00	0.75
rs11976578:158070432:G:A	7:158070432	A	0.785	PTPRN2	PS	4589	0.685	0.136	0.027	3.28E-07	0.00	0.28
rs34821062:42780832:A:AT	8:42780832	AT	0.144	HOOK3; RNF170; MIR4469; THAP1	PS8	4098	0.785	0.276	0.055	4.24E-07	0.00	0.96
rs12064043:67992621:C:T	1:67992621	T	0.382	RP4-763G1.2; SERBP1; IL12RB2; GADD45A	PS	4589	0.843	0.106	0.021	5.18E-07	0.00	0.20
rs7464710:120624386:C:G	8:120624386	G	0.250	ENPP2; TAF2; NOV; DSCC1	PS8	4098	0.963	0.206	0.041	6.42E-07	0.10	3.25
rs35025140:114382521:G:GA	1:114382521	GA	0.234	PTPN22; AP4B1-AS1; RSBN1; BCL2L15	PS8	4098	0.965	-0.212	0.043	6.46E-07	0.00	0.39
rs149905240:161194396:G:C	1:161194396	C	0.023	APOA2; TOMM40L; AL590714.1; MIR5187; FCER1G; NR1I3; NDUFS2; ADAMTS4; PCP4L1; PPOX; B4GALT3	PS	4589	0.928	0.328	0.066	6.83E-07	0.00	0.82
rs72807538:89791343:T:A	16:89791343	A	0.032	ZNF276; VPS9D1; FANCA; SPATA2L; CDK10	PS8	4098	0.837	0.537	0.109	8.42E-07	0.11	1.36
rs79398870:121712292:T:C	6:121712292	C	0.079	GJA1; C6orf170	PS8	4098	0.994	0.327	0.066	8.45E-07	0.00	0.03
rs2657109:43701037:G:A	13:43701037	A	0.980	DNAJC15; LINC00400; ENOX1; EPSTI1	PS8	4098	0.754	-0.687	0.139	8.57E-07	0.21	1.78
rs368336153:150441282:G:A	6:150441282	A	0.011	PPP1R14C; ULBP3; PHBP1; RAET1M	PS8	4098	0.924	0.858	0.174	8.64E-07	0.00	0.05

Tabelle 11: Summary-Statistiken der Meta-Analyse von PS und PS8 im Case-Only Setting, die einen p-Wert $< 1 \times 10^{-6}$ hatten. EAF steht für Effekt-Allel-Frequenz. Die benachbarten Gene innerhalb ± 250 kb sind angegeben. Der gegebene info-Score ist der minimale Score aus beiden Studien. Tau entspricht der Wurzel der Varianz zwischen den beiden Studien, Q steht für Cochrans Q und ist ebenfalls ein Maß für die Heterogenität.

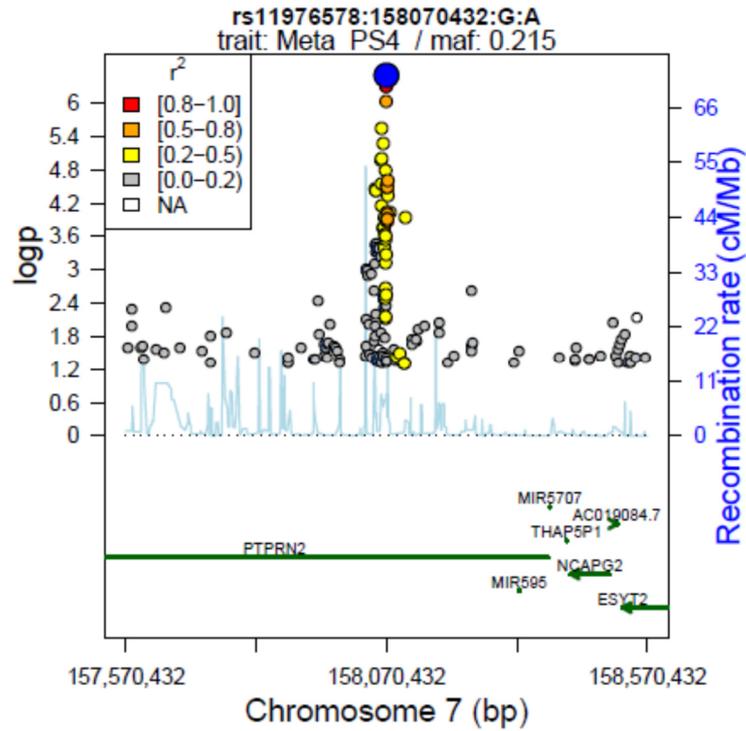


Abbildung 17: RA-Plot von rs11976578. Die SNPs wurden gemäß ihrer Position auf dem Chromosom und ihrem p -Wert geplottet. Die Färbung gibt das LD bezüglich der Top-Variante (blau) an. Der Top-SNP liegt im Gen PTPRN2, welches in einer anderen GWAS mit Endpunkt Karotis-Kalzifikation in Diabetes Patienten beschrieben wurde [50].

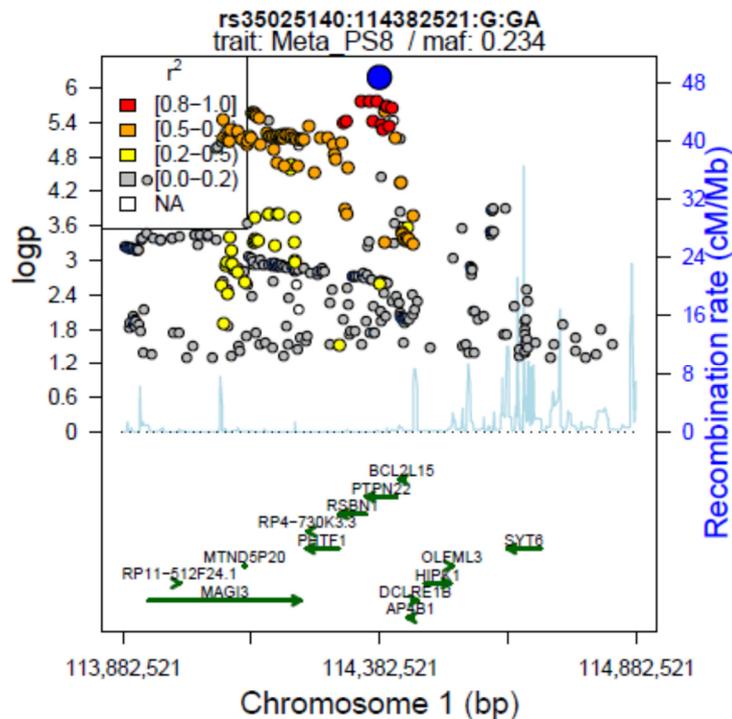


Abbildung 18: RA-Plot von rs35025140. Die SNPs wurden gemäß ihrer Position auf dem Chromosom und ihrem p -Wert geplottet. Die Färbung gibt das LD bezüglich der Top-Variante (blau) an. Der Top-SNP liegt im Gen PTPN22 und weitere stark assoziierte Varianten im Gen MAGI3. Das Gen RSBN1 wurde kürzlich als Locus für KHK beschrieben [52].

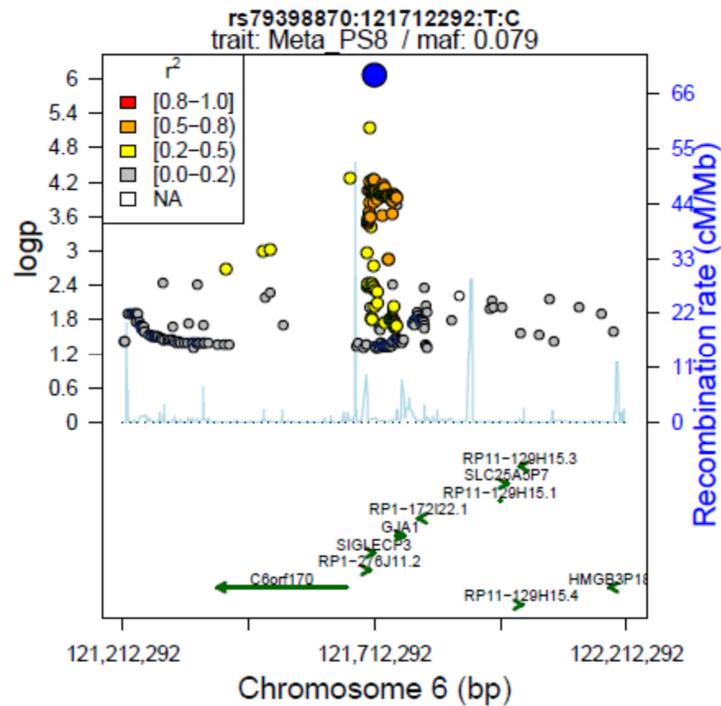


Abbildung 19: RA-Plot von rs79398870. Die SNPs wurden gemäß ihrer Position auf dem Chromosom und ihrem p -Wert geplottet. Die Färbung gibt das LD bezüglich der Top-Variante (blau) an. Der Top-SNP liegt im Gen GJA1. Es gibt Hinweise, dass dieses Protein mit dazu beiträgt atheromatöse Plaques in der ICA zu stabilisieren [53].

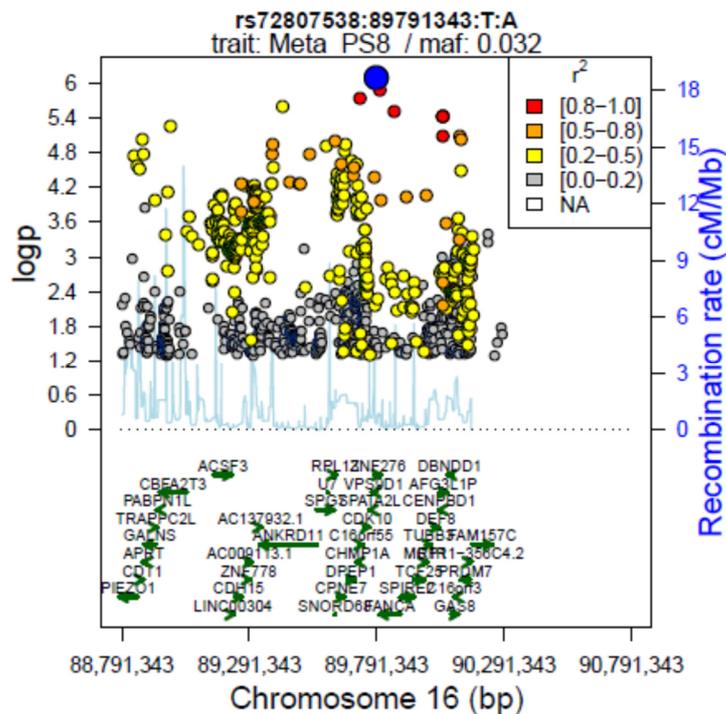


Abbildung 20: RA-Plot von rs72807538. Die SNPs wurden gemäß ihrer Position auf dem Chromosom und ihrem p -Wert geplottet. Die Färbung gibt das LD bezüglich der Top-Variante (blau) an. Diese Region ist sehr genreich.

Darstellung des eigenen Beitrags

Darstellung des eigenen Beitrags zur Publikation:

Titel: Genome-wide meta-analysis identifies novel loci of plaque burden in carotid artery

Journal: Atherosclerosis

Autoren: Janne Pott, Ralph Burkhardt , Frank Beutner, Katrin Horn, Andrej Teren, Holger Kirsten, Lesca Holdt, Gerhard Schuler, Daniel Teupser, Markus Löffler, Joachim Thiery, Markus Scholz

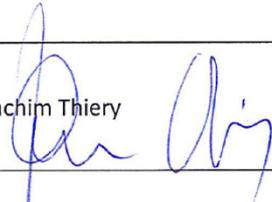
- Beitrag:**
- Qualitätskontrolle der Genotypisierung von LIFE-Heart und LIFE-Adult
 - Imputation der Genetikdaten von LIFE-Heart
 - Definition verschiedener Karotis-Gefäß-Scores
 - Durchführung der genomweiten Assoziationsanalysen (GWAS) in LIFE-Adult und LIFE-Heart
 - Durchführung der Meta-GWAS
 - Analyse und Interpretation der Meta-GWAS-Ergebnisse
 - KHK-Kandidaten-Analyse
 - Durchführung und Interpretation einer Enrichment-Analyse bezüglich KHK und Schlaganfall Loki
 - Schreiben des Manuskripts (gemeinsam mit Markus Scholz)

Co-Autor	Unterschrift	Co-Autor	Unterschrift
Ralph Burkhardt		Frank Beutner	
Katrin Horn		Andrej Teren	
Holger Kirsten		Lesca Holdt	
Gerhard Schuler		Daniel Teupser	
Markus Löffler		Joachim Thiery	
Markus Scholz			

Darstellung des eigenen Beitrags, Janne Pott
Molekulargenetische Faktoren der Suszeptibilität für Karotis-Plaques

Darstellung des eigenen Beitrags zur Publikation:

- Titel: **Genome-wide meta-analysis identifies novel loci of plaque burden in carotid artery**
- Journal: *Atherosclerosis*
- Autoren: Janne Pott, Ralph Burkhardt, Frank Beutner, ^{Katrin}~~Katrin~~ Horn, Andrej Teren, Holger Kirsten, Lesca Holdt, Gerhard Schuler, Daniel Teupser, Markus Löffler, Joachim Thiery, Markus Scholz
- Beitrag:
- Qualitätskontrolle der Genotypisierung von LIFE-Heart und LIFE-Adult
 - Imputation der Genetikdaten von LIFE-Heart
 - Definition verschiedener Karotis-Gefäß-Scores
 - Durchführung der genomweiten Assoziationsanalysen (GWAS) in LIFE-Adult und LIFE-Heart
 - Durchführung der Meta-GWAS
 - Analyse und Interpretation der Meta-GWAS-Ergebnisse
 - KHK-Kandidaten-Analyse
 - Durchführung und Interpretation einer Enrichment-Analyse bezüglich KHK und Schlaganfall Loci
 - Schreiben des Manuskripts (gemeinsam mit Markus Scholz)

Co-Autor	Unterschrift	Co-Autor	Unterschrift
Ralph Burkhardt		Frank Beutner	
Katrin Horn		Andrej Teren	
Holger Kirsten		Lesca Holdt	
Gerhard Schuler		Daniel Teupser	
Markus Löffler		Joachim Thiery	
Markus Scholz			

Darstellung des eigenen Beitrags zur Publikation:

Titel: Genetic regulation of PCSK9 plasma levels and its impact on atherosclerotic vascular disease phenotypes

Journal: *Circulation: Genomic and Precision Medicine*
(bei Einreichung: *Circulation: Cardiovascular Genetics*)

Autoren: Janne Pott, Valentin Schlegel, Andrej Teren, Katrin Horn, Holger Kirsten, Christina Blücher, Jürgen Kratzsch, Markus Löffler, Joachim Thiery, Ralph Burkhardt, Markus Scholz

- Beitrag:**
- Qualitätskontrolle der Genotypisierung von LIFE-Heart und LIFE-Adult
 - Imputation der Genetikdaten von LIFE-Heart
 - Analyse von möglichen Einflussgrößen auf PCSK9
 - Interaktionsanalyse von Alter und Geschlecht auf PCSK9
 - Korrelationsanalyse von PCSK9 Plasma Levels zu Lipid-Parametern
 - Durchführung der genomweiten Assoziationsanalysen (GWAS)
 - Analyse und Interpretation der GWAS-Ergebnisse
 - Analyse der Assoziation von PCSK9 zu verschiedenen Gefäß-Phänotypen
 - Durchführung und Interpretation der Mendelschen Randomisierung
 - Schreiben des Manuskripts (gemeinsam mit Markus Scholz und Ralph Burkhardt)

Co-Autor	Unterschrift	Co-Autor	Unterschrift
Valentin Schlegel		Andrej Teren	
Katrin Horn		Holger Kirsten	
Christina Bluecher		Jürgen Kratzsch	
Markus Löffler		Joachim Thiery	
Ralph Burkhardt		Markus Scholz	

Darstellung des eigenen Beitrags, Janne Pott
Molekulargenetische Faktoren der Suszeptibilität für Karotis-Plaques

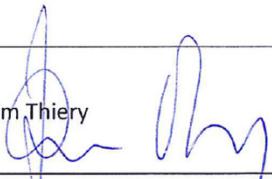
Darstellung des eigenen Beitrags zur Publikation:

Titel: Genetic regulation of PCSK9 plasma levels and its impact on atherosclerotic vascular disease phenotypes

Journal: Circulation: Cardiovascular Genetics

Autoren: Janne Pott, Valentin Schlegel, Andrej Teren, ~~Karin~~^{Katrin} Horn, Holger Kirsten, Christina Blücher, Jürgen Kratzsch, Markus Löffler, Joachim Thiery, Ralph Burkhardt, Markus Scholz

- Beitrag:
- Qualitätskontrolle der Genotypisierung von LIFE-Heart und LIFE-Adult
 - Imputation der Genetikdaten von LIFE-Heart
 - Analyse von möglichen Einflussgrößen auf PCSK9
 - Interaktionsanalyse von Alter und Geschlecht auf PCSK9
 - Korrelationsanalyse von PCSK9 Plasma Levels zu Lipid-Parametern
 - Durchführung der genomweiten Assoziationsanalysen (GWAS)
 - Analyse und Interpretation der GWAS-Ergebnisse
 - Analyse der Assoziation von PCSK9 zu verschiedenen Gefäß-Phänotypen
 - Durchführung und Interpretation der Mendelschen Randomisierung
 - Schreiben des Manuskripts (gemeinsam mit Markus Scholz und Ralph Burkhardt)

Co-Autor	Unterschrift	Co-Autor	Unterschrift
Valentin Schlegel		Andrej Teren	
Katrin Horn		Holger Kirsten	
Christina Bluecher		Jürgen Kratzsch	
Markus Löffler		Joachim Thiery	
Ralph Burkhardt		Markus Scholz	

Selbstständigkeitserklärung

Erklärung über die eigenständige Abfassung der Arbeit

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar eine Vergütung oder geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt an der Entstehung der vorliegenden Arbeit beteiligt waren. Die aktuellen gesetzlichen Vorgaben in Bezug auf die Zulassung der klinischen Studien, die Bestimmungen des Tierschutzgesetzes, die Bestimmungen des Gentechnikgesetzes und die allgemeinen Datenschutzbestimmungen wurden eingehalten. Ich versichere, dass ich die Regelungen der Satzung der Universität Leipzig zur Sicherung guter wissenschaftlicher Praxis kenne und eingehalten habe.

.....
Datum

.....
Unterschrift

Lebenslauf

Akademische und schulische Ausbildung

- Seit 07/2015 **Promotion im Rahmen einer ESF-geförderten Nachwuchsforschergruppe, Institut für medizinische Informatik, Statistik und Epidemiologie (IMISE), Universität Leipzig**
- 10/2012 – 05/2015 **Studium der Biomathematik, Ernst-Moritz-Arndt Universität (EMAU) Greifswald**
- Abschlussnote: 1.5
 - Abschluss: M.Sc.
 - Thema der Abschlussarbeit: *Phylogenetic decisive and non-decisive taxon sets - analysis and comparison*
- 10/2009 – 09/2012 **Studium der Integrated Life Sciences, Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg**
- Abschlussnote: 2.3
 - Abschluss: B.Sc.
 - Thema der Abschlussarbeit: *BAX-BIM-Interaction Studied by Molecular Dynamics Simulation*
- 09/2000 – 06/2009 **Abitur, Marie-Therese-Gymnasium, Erlangen**
- Abschlussnote: 2.0
 - Schwerpunkte: Mathematik, Chemie

Beruflicher Werdegang

- Seit 07/2015 **Wissenschaftliche Mitarbeiterin, Institut für medizinische Informatik, Statistik und Epidemiologie (IMISE), Leipzig**
- Aufgaben:**
- Aufbereitung molekularer und epidemiologischer Daten
 - Genetische Assoziationsstudien zu Lipid- und Gefäßphänotypen
 - Analyse von Kausalketten
 - Lehrtätigkeit im Modul „*Statistische Aspekte der Analyse molekularbiologischer und genetischer Daten*“
 - Betreuung von Masteranden
- 10/2013 – 01/2014 **Studentische Hilfskraft, Institut für Mathematik und Informatik, Greifswald**
- 10/2012 – 01/2013 **Studentische Hilfskraft, Institut für Mathematik und Informatik, Greifswald**
- Leitung der Übung zur Vorlesung “Gewöhnliche Differentialgleichungen”
- 04/2012 – 08/2012 **Studentische Hilfskraft, Department Mathematik, Erlangen**
- Leitung der Übung zur Vorlesung “Mathematische Verfahren in der Bioinformatik”

Weiterbildung

Seit 10/2017

Teilnahme an Kursen / Modulen des Hochschuldidaktischen Zentrum Sachsens (HDS)

Ziel: Sächsisches Hochschuldidaktik-Zertifikat (Abschluss 03/2019)

Modul 1: Grundlagen des hochschuldidaktischen Handelns

Modul 2: Workshops zu verschiedenen Themen

- Beraten und Begleiten
- Evaluieren
- Lehren und Lernen
- Prüfen und Bewerten

Weitere Kenntnisse

Kenntnisgrade: + Grundkenntnisse ++ Fortgeschritten +++ Sehr gut

Gebiet	Teilgebiet	Kenntnisgrad
Sprachen	Deutsch	+++
	Englisch	++
Betriebssysteme	Windows	+++
	Linux/UNIX	++
Anwendung	MS Office (Word, Excel, PowerPoint)	+++
	Latex	+++
	R	+++
	C++	++
	SPSS	++
	MATLAB	+

Ort, Datum

Janne Pott

Wissenschaftliche Erfahrung

Lehr- und Betreuungserfahrung während der Promotion

- | | |
|--|---|
| 10/2017 – 02/2018
10/2016 – 02/2017 | Lehrtätigkeit im Modul „Statistische Aspekte der Analyse molekularbiologischer und genetischer Daten“ <ul style="list-style-type: none">• Erstellung der Übungsaufgaben und Leitung der Übung• Konzeption und Moderation des Seminars• Entwurf der Klausuraufgaben und Klausuraufsicht |
| 07/2017 – 03/2018
06/2016 – 08/2017 | Betreuung von zwei Masteranden <ul style="list-style-type: none">• Einführung in Genetische Assoziationsstudien• Ansprechpartner für weitere Analysen• Unterstützung bei der Gliederung der Arbeit |
| Seit 04/2017 | Statistische Beratung zweier Medizin-Doktoranden <ul style="list-style-type: none">• Ansprechpartner für statistische Analysen• Unterstützung bei der Gliederung der Arbeit |

Publikationen

Pott J, Burkhardt R, Beutner F, et al. Genome-wide meta-analysis identifies novel loci of plaque burden in carotid artery. *Atherosclerosis*. 2017;259:32–40.
doi:10.1016/j.atherosclerosis.2017.02.018.

Pott J, Schlegel V, Teren A, et al. Genetic regulation of PCSK9 plasma levels and its impact on atherosclerotic vascular disease phenotypes. *Circulation: Genomic and Precision Medicine*. 2018

Franceschini N, Giambartolomei C, ... , **Pott J**, et al. Genome-wide association study of carotid intima-media thickness and plaque identifies novel loci for atherosclerosis and cardiovascular outcomes, and downstream regulatory effects in vascular tissue. Eingereicht bei *Nature Genetics*. 2018

Vorträge und Workshops

- | | |
|----------------|--|
| 13.-15.11.2017 | Workshop "Mendelian Randomization Course" , MRC Biostatistics Unit, University of Cambridge, in Cambridge, UK |
| 04.-05.05.2017 | Workshop „Biometrische Aspekte der Genomanalyse“ , in Heidelberg
Vortragsthema: "Genetic regulation of PCSK9 plasma levels and its impact on cardio-vascular disease" |

Danksagung

Hier möchte ich mich bei all jenen bedanken, die mich während der Erstellung der Promotionsarbeit unterstützt haben:

An erster Stelle möchte ich mich bei Prof. Markus Scholz für die engagierte Betreuung und Unterstützung bei der Anfertigung dieser Arbeit bedanken.

Des Weiteren gilt mein Dank allen Mitarbeitern des IMISE für das freundliche Arbeitsumfeld und der ESF-Nachwuchsforschergruppe, insbesondere Prof. Markus Löffler und Prof. Ralph Burkhardt, für die anregenden Diskussionen. Ebenso bedanke ich mich bei meinen Kollegen im Zimmer 208 für die Toleranz meiner Selbstgespräche und ihrer Diskussionsbereitschaft, und der Arbeitsgruppe Genetische Statistik, speziell bei Dr. Holger Kirsten und Katrin Horn, für deren umfangreiche fachliche Unterstützung.

Zusätzlich möchte ich mich bei allen Teilnehmern der LIFE-Adult und der LIFE-Heart Studie für die Bereitstellung ihrer Daten bedanken. Mein Dank gilt auch den Studienassistenten Kay Olisher und Annegret Unger, Sylvia Henger und dem Team vom LIFE-Datenmanagement für die gute Zusammenarbeit.

Abschließend möchte ich mich bei meiner Familie bedanken: bei meinen Eltern, ohne die ich dieses Studium nie begonnen hätte und die mich immer unterstützt haben; bei meinen beiden Schwestern für ihre gemeinsame Unterschiedlichkeit, die mich lehrt, dass Genetik nicht alles im Leben ist; und bei meinen Großeltern, Neffen und Nichte für die spaßigen Ablenkungen und den kritischen Fragen nach dem Sinn und Zweck der Genetik.