

Aus der  
Lübecker Interdisziplinären Plattform für Genomanalytik (LIGA)  
der Universität zu Lübeck – Sektion Medizin –  
Leitung: Prof. Dr. med Lars Bertram

---

# Die Rolle des CD33-Gens in der Prädisposition zum Morbus Alzheimer

Inaugural-Dissertation  
zur  
Erlangung der Doktorwürde  
an der Universität zu Lübeck  
-Aus der Sektion Medizin-

vorgelegt von  
Anastasia Paulssen (geb. Nikulina)  
aus Ufa, Russland

Lübeck  
2021

1. Berichterstatter: Prof. Dr. med. Lars Bertram  
2. Berichterstatter/-in: Prof. Dr. rer. nat. Hauke Busch

Tag der mündlichen Prüfung: 04.08.2022  
zum Druck genehmigt, Lübeck, den 04.08.2022

- Promotionskommission der Sektion Medizin -

## Inhaltsverzeichnis

1	Abkürzungsverzeichnis .....	1
2	Einleitung und Fragestellung.....	3
2.1	Morbus Alzheimer als Krankheitsbild .....	3
2.2	Ätiologie der AD.....	4
2.3	Genetische Assoziationsstudien der spätmanifesten Form der AD .....	5
2.4	CD33 Gen: Aufbau, Funktion und Assoziation mit der AD .....	8
2.5	Fragestellung und Ziel der Arbeit.....	13
3	Material und Methoden .....	14
3.1	Metaanalyse und systematischer Review .....	14
3.2	Methodische Vorgehensweise .....	15
3.3	Auswahlkriterien inklusive Ein- und Ausschlusskriterien .....	16
3.4	Informationsquellen und Datensuchstrategie .....	16
3.5	Auswahl der Studien und Datensammlung .....	17
3.6	Statistische Auswertung und Reduzierung von Biases in der Metaanalyse .....	18
3.7	Datengenerierung und statische Auswertung eQTL-GWAS.....	19
4	Ergebnisse.....	21
4.1	Metaanalysen genetischer Assoziationsstudien zur Rolle des CD33-Gens in der AD .....	21
4.1.1	PRISMA Flow Chart.....	21
4.1.2	Überblick Ergebnisse der eigenen Metaanalysen .....	23
4.1.3	SNP rs3865444 .....	24
4.1.4	SNP rs3826656 .....	27
4.1.5	SNP rs1697553 .....	29
4.1.6	SNP rs2455069 .....	29
4.1.7	SNP rs12459419 .....	30
4.1.8	SNP rs35112940 .....	30
4.1.9	Zusammenfassung der in dieser Arbeit berechneten Metaanalysen .....	31
4.1.10	Vergleich mit anderen Metaanalysen.....	33

4.2	Genetische Determinanten der CD33-Expression im menschlichen Gehirn.....	35
4.2.1	Ergebnisse der <i>cis</i> eQTL-Analysen zur CD33-Genexpression.....	35
4.2.2	Ergebnisse der <i>trans</i> eQTL-GWAS-Analysen zur CD33-Genexpression .....	37
4.2.3	Gen-basierte eQTL-GWAS Ergebnisse zur CD33-Genexpression im menschlichen Gehirn.....	40
5	Diskussion .....	43
5.1	Diskussion der wissenschaftlichen Relevanz der vorliegenden Dissertation .....	43
5.2	Diskussion der Ergebnisse der Metaanalysen .....	44
5.2.1	Zusammenfassung der wichtigsten Ergebnisse .....	44
5.2.2	Vergleich der erhobenen Metaanalyse-Ergebnisse mit der Literatur .....	45
5.2.3	Stärken und Schwächen dieses Teils der Arbeit.....	49
5.3	Diskussion der Ergebnisse der eQTL GWAS Analysen .....	50
5.3.1	Zusammenfassung der wichtigsten Ergebnisse .....	50
5.3.2	Stärken und Schwächen dieses Teils der Arbeit.....	52
5.4	Schlussfolgerung und Ausblick .....	53
6	Zusammenfassung (Abstract).....	55
7	Literaturverzeichnis.....	56
8	Anhang .....	66
9	Danksagung .....	79
10	Lebenslauf .....	80

## 1 Abkürzungsverzeichnis

AD	Alzheimer's Disease
AMSTAR	Assessment of Multiple SysTemAtic Reviews
APOE	Apolipoprotein E
CI	Confidence Interval
DNA	Deoxyribonucleic Acid
EAF	Effect Allele Frequency
EOAD	Early Onset Alzheimer's Disease
eQTL	Expression Quantitative Trait Loci
FUMA	Functional Mapping and Annotation of Genome-Wide Association Studies
GSA	Global Screening Array
GWAS	Genome-Wide Association Study
HWE	Hardy-Weinberg-Equilibrium
IGAP	International Genomics of Alzheimer's Project
IKMB	Institut für Klinische Molekularbiologie
ITIM	Immunoreceptor Tyrosinebased Inhibitory Motif
LD	Linkage Disequilibrium
LIGA	Lübecker Interdisziplinäre Plattform für Genomanalytik
LOAD	Late Onset Alzheimer's Disease
MAF	Minor Allele Frequency
MECIR	Methodological Expectations for the Conduct, Reporting and Updating of Systematic Reviews of Interventions
MIRNETAD	MicroRNA Dysfunction in Alzheimer's Disease
mRNA	Messenger Ribonucleic Acid
NICE	National Institute for Health and Care Excellence
OR	Odds Ratio
PCA	Principal-Component-Analyse
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QC	Quality Control
QUOROM	Quality Of Reporting Of Meta-Analyses
RNA	Ribonucleic Acid

SE	Standard Error
Siglec	Sialic acid-binding immunoglobuline-like lectins
SNP	Single Nucleotide Polymorphism
WHO	World Health Organisation

## 2 Einleitung und Fragestellung

### 2.1 Morbus Alzheimer als Krankheitsbild

“Alzheimer’s disease starts when a protein that should be folded up properly  
misfolds into a kind of demented origami”

- Gregory A. Petsko –

Mit diesem Satz beschrieb Gregory A. Petsko, Professor of Neurology an der Harvard Medical School, im Rahmen seines TED-Talks im Jahr 2008 mit informellem Sprachgebrauch die Pathogenese der Demenz vom Alzheimer-Typ [15, 87]. Unter dem Begriff „Demenz“ ist ein Syndrom meist neurodegenerativer Genese gemeint, deren Hauptsymptom eine Verschlechterung von kognitiven Fähigkeiten mit Verlust der Alltagskompetenz ist. Die Alzheimer Krankheit (auch: Morbus Alzheimer; engl. Alzheimer’s Disease; AD) ist mit bis zu 80% aller Fällen die häufigste Form der Demenz [15]. Sie ist nach dem deutschen Psychiater und Pathologen Alois Alzheimer benannt, der das klinische und pathologische Bild der Erkrankung erstmals im Jahr 1907 beschrieb [3].

Das klinische Bild der AD lässt sich grob in vier Stadien einteilen [14]. Ein mehrere Jahre dauerndes präklinisches (mit anderen Worten: präsymptomatisches) Stadium ist durch milde Gedächtnissstörung und bereits vorhandene pathologische degenerative Veränderungen in Gehirnkortex und Hippocampus gekennzeichnet, ohne dass die täglichen Aktivitäten beeinträchtigt werden. Dieses Stadium wird auch „silent stage“ genannt [30]. Als Nächstes beginnt das milde Stadium der AD mit Störung des täglichen Lebens aufgrund Konzentrations- und Gedächtnisverlustes, zeitlicher und örtlicher Desorientierung, Stimmungslabilität und Depressionsmanifestation. Das dritte Stadium ist ein bereits moderates Stadium, das durch Impulskontrollstörung und schwere Beeinträchtigung der kognitiven Fähigkeiten wie Lesen, Schreiben und Sprechen, sowie Erkennen von Familienmitgliedern gekennzeichnet ist [53, 112]. Das Spätstadium der AD führt zu massiven kognitiven und funktionellen Beeinträchtigungen wie Dysphagie, Miktionssstörungen und Abnahme von Aktivität bis zu Bettlägerigkeit. Diese fortgeschrittenen Komplikationen können zum Tod führen [14].

Aktuelle Zahlen der World Health Organisation (WHO) zeigen auf, dass die AD mittlerweile zu den zehn häufigsten Todesursachen weltweit gehört. In Europa und in den USA belegt die AD im Jahr 2019 bereits den dritten Platz der häufigsten Todesursachen [113].

Die Inzidenz steigt weiterhin in hohem Tempo - es wird geschätzt, dass bis zum Jahr 2050 die Zahl der Menschen mit der Alzheimer Krankheit 131.5 Millionen übersteigen wird [104]. Es kann daher davon ausgegangen werden, dass die AD nicht nur eine aktuelle sozioökonomische Belastung ist, sondern auch weiterhin eine große zukünftige Herausforderung für die Gesundheits- und Sozialsysteme sein wird. Diese Statistik und die alarmierenden Prognosen verdeutlichen den notwendigen Bedarf von neuen therapeutischen Strategien. Trotz erheblicher weltweiter Forschungsfortschritte gibt es zurzeit jedoch keine kausale Therapie der AD, sondern nur kurzfristig erfolgreiche Methoden zur Linderung der Symptome (z. B. durch Gabe von Acetylcholinesterase-Inhibitoren [14]). Ein erfolgversprechendes Potential wird jedoch in der Entwicklung von neuen Therapiemöglichkeiten, unter anderem dem Genom-Editing und anderen molekularen und stammzellbasierten Therapien gesehen [89].

## 2.2 Ätiologie der AD

Nach aktuellem wissenschaftlichen Forschungsstand wird die Ätiopathogenese der AD durch die abnorme Ablagerung von  $\beta$ -Amyloid und Tau-Protein erklärt [108]. Die Struktur und Funktion von Proteinen (sowie deren Dysfunktion) wird laut dem zentralen Dogma der Molekularbiologie maßgeblich von Genen bestimmt, d.h. der Abfolge von Nukleotiden in der DNA [88]. Für viele Fälle der frühmanifesten Form der AD (engl. „Early onset Alzheimer's disease“; EOAD; Erkrankungsbeginn <65 Jahre) sind seltene, aber nahezu vollständig penetrante Mutationen in den Genen Amyloid-Vorläuferprotein (*APP*), Presenilin 1 und 2 (*PSEN1*, *PSEN2*) verantwortlich. Diese monogenen Formen der AD machen allerdings mit deutlich weniger als 5% nur einen sehr geringen Anteil aller AD-Fälle aus [10]. Die häufigste Form der Erkrankung ist die spätmanifeste AD (engl. „late onset Alzheimer's disease“; LOAD; Erkrankungsbeginn  $\geq$ 65 Jahre), die auf ein komplexes Zusammenspiel von genetischen und nicht-genetischen Faktoren zurückzuführen ist. Für den überwiegenden Teil des LOAD-Risikos (einer häufig zitierten Studie zufolge ~70% [56]) sind genetische Faktoren verantwortlich, weswegen diese Form der Erkrankung häufig auch als „polygene AD“ bezeichnet wird. Der stärkste bekannte genetische Risikofaktor für diese Form ist das Epsilon 4-Allel ( $\epsilon$ 4-Allel) des Apolipoprotein-E-Gens (*APOE*) [10, 26, 96, 105]. Es wird geschätzt, dass dieses Allel die Heritabilität der AD, d.h. den Anteil der Erkrankung der durch genetische Faktoren verursacht wird, zu nahezu 50% erklärt [4].

Dies bedeutet im Umkehrschluss, dass es andere Risikogene der LOAD geben muss, die für die andere Hälfte der geschätzten Heritabilität verantwortlich sind. Die Suche nach weiteren LOAD-Riskogenen ist seit den 1980'er Jahren Gegenstand eines großen Teils der AD-Forschung und ist auch Gegenstand dieser Dissertation.

## 2.3 Genetische Assoziationsstudien der spätmanifesten Form der AD

Genetische Assoziationsstudien sind ein wichtiges Instrument zur Identifizierung von genetischen Risikofaktoren komplexer Erkrankungen. Das klassische Studiendesign zur Assoziationsprüfung ist die Fall-Kontroll-Studie, bei der die Häufigkeit einer von genetischen Markern innerhalb einer Gruppe erkrankter Probanden (Fälle) und einer zweiten, aus gesunden Personen bestehenden Gruppe (Kontrollen), miteinander verglichen wird. Eine (signifikant) erhöhte oder erniedrigte Häufigkeit eines genetischen Markers in der Fall-Gruppe deutet auf ein (statistisch) größeres oder kleineres Risiko für die Entstehung einer bestimmten Krankheit hin. Einzelnukleotid-Polymorphismen (engl. „single nucleotide polymorphism“; SNP) sind heutzutage die am häufigsten getesteten Marker in genetischen Assoziationsstudien. Ein SNP beschreibt eine Variation in der DNA-Sequenz, bei deren sich typischerweise die Allele an einer Nukleotidbase unterscheiden. Eine signifikant assoziierte Variante kann dabei auch selbst krankheitsbeeinflussende funktionelle Effekte auslösen (sog. direkte Assoziation, z. B. verursacht durch Aminosäure-austauschende Polymorphismen). Sehr viel häufiger werden aber indirekte Assoziationen beobachtet, d.h. dass einer signifikant assoziierten Variante kein unmittelbar funktionell-relevant Effekt zugeordnet werden kann. Dies ist z. B. der Fall, wenn der assoziierte SNP im Kopplungsungleichgewicht (engl. „linkage disequilibrium“; LD) mit der eigentlichen krankheitsbeeinflussenden Variante steht, diese selbst aber nicht getestet wurde [58, 109].

Des Weiteren unterscheidet man zwischen Kandidatengen-Studien, in denen eine genetische Assoziationsanalyse mit typischerweise wenigen Polymorphismen, die aufgrund bestimmter Kriterien („Kandidatengen“) ausgewählten wurden, durchgeführt werden und sog. genomweiten Assoziationsstudien (engl. „genome-wide association study“; GWAS), in denen im Extremfall das gesamte Genom eines Organismus ohne vorherige Hypothese bzgl. des Kandidatenstatus der Varianten analysiert wird [58, 101].

Experimentell war der Kandidatengenansatz für lange Zeit (~1980-2006) technisch die einzige Möglichkeit der Durchführung einer genetischen Assoziationsstudie. In diesem, jetzt häufig als „Pre-GWAS-Ära“ bezeichneten Zeitabschnitt, wurden im Alzheimer-Feld bereits tausende solcher Kandidatengen-Studien veröffentlicht, die einen Zusammenhang zwischen der AD und buchstäblich hunderten von mutmaßlichen Risikoallelen in verschiedenen Genen festgestellt oder ausgeschlossen haben. Um diese Ergebnisse besser verfolgen zu können und eine bessere übergreifende Interpretation der veröffentlichten Daten zu ermöglichen, wurde von der Arbeitsgruppe von Prof. Bertram im Jahr 2005 eine öffentlich zugängliche Datenbank für genetische Assoziationsstudien der AD erstellt (AlzGene-Datenbank), die alle seinerzeit publizierten Assoziationsstudien in

einen Katalog erfasst hat und die unter der URL <http://www.alzgene.org> erreichbar ist [2]. Diese Datenbank stellte nicht nur ein leistungsfähiges Instrument zur Verbesserung des Verständnisses der AD-Genetik dar, sondern diente auch für die potenzielle Suche nach den Genkandidaten bei anderen polygenen Erkrankungen. In 2007 wurden die Ergebnisse einer ersten systematischen Meta-Analyse der in AlzGene erfassten Kandidatengen-Studien in der Zeitschrift *Nature Genetics* veröffentlicht [9]. Neben der Bestätigung der Assoziation mit dem bereits oben erwähnten ε4-Allel im *APOE*-Gen hat diese Meta-Analyse insgesamt 24 Polymorphismen in 13 weiteren Kandidatengenen als studienweit signifikant assoziiert ermittelt, d.h. in *ACE*, *CHRN2*, *CST3*, *ESR1*, *GAPDHS*, *IDE*, *MTHFR*, *NCSTN*, *PRNP*, *PSEN1*, *TF*, *TFAM* und *TNF* [9]. Bemerkenswert ist, dass die nachfolgend durchgeführten GWAS nur sehr wenige dieser AD-Kandidatengene aus der „Pre-GWAS-Ära“ validiert haben [10].

Seit dem Jahr 2007 wurden in ca. 60 GWAS viele weitere AD-assozierte Polymorphismen in anderen Risikogenen entdeckt. Der mit Abstand am stärksten – sowohl basierend auf der Effektstärke als auch in Bezug auf die statistische Evidenz - assoziierte Polymorphismus aller AD-GWAS ist das schon oben beschriebene ε4-Allel im *APOE*-Gen. Die zum Zeitpunkt der Schriftlegung aktuellsten und von der Stichprobenanzahl größten AD-GWAS sind die Studien von Jansen et al. [48] und Kunkle et al. [54], die im Jahr 2019 ebenfalls in der Zeitschrift *Nature Genetics* publiziert wurden. Laut der Einschätzung von Bertram und Tanzi [10] zeigen die meisten Ergebnisse dieser beiden Studien eine gute Übereinstimmung miteinander. Insgesamt wurden 32 Risikoloci für die LOAD identifiziert, 27 davon mit „häufigen“ Polymorphismen, d.h. Varianten, in denen das seltene Allel eine Frequenz von  $\geq 1\%$  in der Normalbevölkerung hat (s. Tabelle 1), und 5 weitere mit seltenen Varianten (Frequenz <1%; s. Tabelle 2).

Chr	Pos	LeadSNP	A1 vs. A2	MAF†	P-value Jansen et al.	P-value Kunkle et al.	AD effect	Nearest gene	AD pathway	Potential link to AD pathogenesis
1	161155392	rs4575098	A vs. G	0.240	2.05E-10	2.34E-02*	risk	ADAMTS4	none	Neuroprotection; Extracellular Matrix Protease
1	207786828	rs2093760	A vs. G	0.225	1.10E-18	1.66E-15*	risk	CR1	immune	Innate Immunity; Neuroinflammation
2	127891427	rs4663105	A vs. C	0.412	3.38E-44	2.16E-26*	risk	BIN1	lipid	Cellular Protein Trafficking
2	233981912	rs10933431	G vs. C	0.240	8.92E-10	3.42E-09**	protection	INPPD5	none	Autophagy; Viral Infection
4	11026028	rs6448453	A vs. G	0.228	1.93E-09	4.90E-05*	risk	CLNK	none	Innate Immunity; Neuroinflammation
6	32583357	rs9469112	T vs. A	0.153	8.41E-11	2.32E-07**	protection	HLA-DRB1	immune	Adaptive Immunity
6	47432637	rs9381563	C vs. T	0.344	2.52E-10	3.57E-10**	risk	CD2AP	none	Blood Brain Barrier; $\beta$ -Transcytosis
7	99971834	rs4727449rs18A	A vs. G	0.323	2.22E-15	1.22E-09**	protection	ZCWPW1	none	Innate Immunity; Neuroinflammation
7	143108158	rs7810606	T vs. C	0.425	3.59E-11	1.13E-06**	protection	EPHA1	none	Signal Transduction
8	27464929	rs28834970/rs	A vs. G	0.390	2.61E-19	5.60E-23**	protection	CLU/PTK2B	immune; lipid; tau	$\beta$ - clearance/Signal Transduction
10	11717397	rs11257238	C vs. T	0.382	1.26E-08	2.61E-07**	risk	ECHDC3	none	Lipid Metabolism
11†	47380340	rs3740688	G vs. T	0.458	4.50E-05	5.46E-13**	protection	SPI1/CELF1	immune	Potential false-positive result (not replicated in UKB)
11	59958380	rs2081545	A vs. C	0.342	1.55E-15	5.35E-17**	protection	MS4A6A	immune	Innate Immunity; Neuroinflammation
11	85776544	rs867611	G vs. A	0.342	2.19E-18	3.41E-19**	protection	PICALM	APP	Blood Brain Barrier; $\beta$ -Transcytosis
11	121435587	rs11218343	C vs. T	0.035	1.09E-11	2.88E-12**	protection	SORL1	lipid; APP	Cellular Protein Trafficking
14†	53391680	rs17125924	G vs. A	0.099	5.26E-06	1.42E-09**	protection	FERMT2	n.a.	Potential false-positive result (not replicated in UKB)
14	92938855	rs12590654	A vs. G	0.347	1.65E-10	8.73E-09*	protection	SLC24A4	none	Calcium Homeostasis
15	59022615	rs442495	C vs. T	0.334	1.31E-09	2.51E-07**	protection	ADAM10	immune	Shedase; APP Processing
15	63569902	rs117618017	T vs. C	0.132	3.35E-08	2.38E-04*	risk	APH1B	none	$\gamma$ -secretase; APP Processing
16†	19808163	rs7185636	C vs. T	0.156	1.40E-01	2.4E-08***	protection	IQCK	n.a.	Likely false-positive result (not replicated in Jansen)
16	31133100	rs59735493	A vs. G	0.324	3.98E-08	7.42E-03*	protection	KAT8	none	Transcriptional Regulation
16†	79355857	rs62039712	G vs. A	0.094	7.66E-01	3.70E08*	risk	WWOX	n.a.	Likely false-positive result (not replicated in Jansen)
17	5138980	rs113260531	A vs. G	0.118	9.16E-10	3.70E-04**	risk	SCIMP	none	Innate Immunity; Neuroinflammation
17	47450775	rs28394864	A vs. G	0.471	1.87E-08	4.85E-03*	risk	AB13	none	Innate Immunity; Neuroinflammation
17	61538148	rs138190086	A vs. G	0.017	2.65E-04	5.30E-09***	risk	ACE	immune	$\beta$ - degradation; blood pressure regulation
18	56189459	rs76726049	C vs. T	0.011	3.30E-08	1.76E-01*	risk	ALPK2	none	Signal Transduction
19	1039323	rs111278892	G vs. C	0.165	7.93E-11	1.10E-07*	risk	ABC47	lipid; APP	Lipid Metabolism; Innate Immunity
19	45411941	rs429358	C vs. T	0.155	<1E-900	1.17E-881*	risk	APOE	lipid; APP; tau	$\beta$ - clearance/Lipid Metabolism
19	46241841	rs76320948	T vs. C	0.059	4.64E-08	1.22E-04*	risk	AC074212.3	none	?
19	51727962	rs3865444	A vs. C	0.336	6.34E-09	5.27E-06**	protection	CD33	none	Innate Immunity; Neuroinflammation
20	54998544	rs6014724	G vs. A	0.089	6.56E-10	3.65E-07*	protection	CASS4	none	Signal Transduction
21†	28156856	rs2830500	A vs. C	0.336	1.65E-02	2.60E-08***	protection	ADAMTS1	n.a.	Likely false-positive result (not replicated in Jansen)

Tabelle 1: Übersicht von häufigen LOAD-Varianten [10]

Chr	Pos	LeadSNP	A1 vs. A2	MAF†	P-value Jansen	P-value Kunkle	AD effect	Nearest gene	AD pathway	Potential link to AD pathogenesis
3	57226150	rs184384746	T vs. C	0.002	1.24E-08	n.a.	risk	HESX1	none	Homeobox Gene; Development
6	41129252	rs75932628	T vs. C	0.002	2.95E-15	2.95E-12*	risk	TREM2	immune system response	Innate Immunity; neuroinflammation
7	145950029	rs114360492	T vs. C	0.0003	2.10E-09	n.a.	risk	CNTNAP2	none	Neuronal Development
16#	81942028	rs72824905	G vs C	0.01	2.11E-03	7.92E-03*	protection	PLCG2	none	Microglial activation; neuroinflammation
17#	47297297	rs616338	T vs. C	0.01	7.81E-07	n.a.	risk	AB13	none	Microglial activation; neuroinflammation

Tabelle 2: Übersicht von seltenen LOAD-Varianten [10]

Zu den wichtigsten GWAS-Erkenntnissen (bisher) gehört ein verbessertes Verständnis der molekularen Pathogenese der AD. Eine unerwartete aber über die bisherigen GWAS hinweg relativ konsistente Erkenntnis war, dass viele GWAS-Gene, die signifikante Assoziationen mit dem AD-Risiko zeigen, einen Bezug zur menschlichen Immunantwort haben [10]. Auch die Ergebnisse von Genexpressionsstudien bestätigen diesen Befund, weil z. B. die GWAS-Riskogene besonders hoch in immunsystemrelevanten Geweben wie Milz, Leber, Leukozyten, aber vor allem in den Mikrogliazellen exprimiert sind [48]. Folgenden AD-Riskogenen, die via GWAS identifiziert wurden, wird laut [6, 10] ein Bezug zum Immunsystem / Immunantwort zugeschrieben (s. auch Abbildung 1): ABCA7, AB13, CD2AP, CD33, CLU, CR1, EPHA1, HLA-DRB5/DRB1, INPP5D, MS4A6A/MS4A6E, PLCG2, TREM2, TYROBP, SPI1, PLGCG2 und CD33. Für diese Arbeit besonders interessant ist die Assoziation mit dem Gen CD33, das im Nachfolgenden näher beschrieben werden soll.

Eine Übersicht aller AD-bezogenen Riskogene und deren Funktionen ist in der Abbildung 1 schematisch dargestellt.

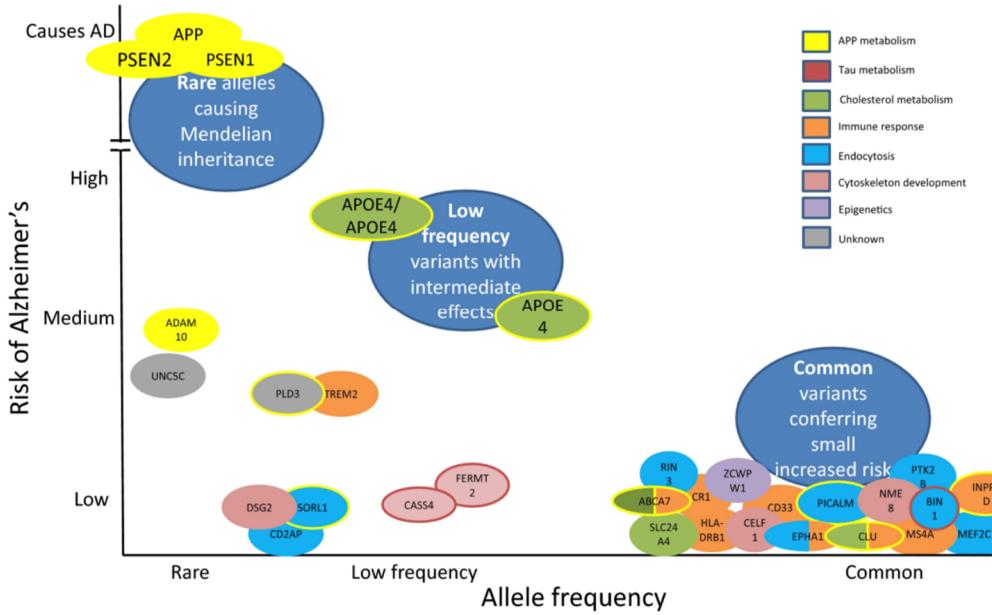


Abbildung 1: AD-Gene-Overview [56]

Legende: Interne Farbe entspricht der Funktion. Gelb eingekreiste Gene beeinflussen den Stoffwechsel des Amyloid-Vorläuferproteins; rot eingekreiste Gene beeinflussen den Tau-Stoffwechsel

## 2.4 CD33 Gen: Aufbau, Funktion und Assoziation mit der AD

Das CD33-Gen (engl. „cluster of differentiation 33“) liegt auf dem langen Arm vom Chromosom 19 (19q13.33) und spielt eine wichtige Rolle in der Regulation der Immunzellfunktionen der sog. „angeborenen Immunantwort“. Das Gen kodiert das „myeloid cell surface antigen CD33“-Protein, ein Zelloberflächen-Rezeptor aus der Familie der Siglec-Proteine (sialic acid-binding immunoglobuline-like lectins) und wird daher auch als Siglec-3 bezeichnet [33]. Die Expression des kodierten Proteins findet am stärksten an hämatopoetischen und phagozytischen Zellen und an Mikrogliazellen im Gehirn statt [5, 120]. Mikroglia sind im Gehirn und Rückenmark angesiedelte Makrophagen, die eine zentrale Rolle in der ersten (d.h. angeborenen) Immunantwort spielen [115].

Erste Hinweise auf eine genetische Assoziation zwischen Polymorphismen im CD33-Gen und dem LOAD-Risiko wurden in einer familien-basierten GWAS der Arbeitsgruppe von Prof. Bertram bereits im Jahr 2008 beschrieben (mit SNP rs3826656) [8]. Ähnliche Befunde, d.h. eine genomweit signifikante Assoziation mit dem LOAD-Risiko, wurden nachfolgend in mehreren Fall-Kontroll-GWAS anderer Arbeitsgruppen berichtet (v.a. mit SNP rs3865444) [45, 73]. Eine Metaanalyse aus den beiden Studien mit insgesamt 18.762 Fällen und 29.827 Kontrollen zeigte erneut eine signifikante Assoziation zwischen diesem SNP und AD mit einem OR von 0,91 (P-Wert =  $1.6 \times 10^{-9}$ , 95%-CI: 0.88 – 0.9) [73]. Eine weitere GWAS von Lambert et al. aus dem Jahr 2013 konnte diese Entdeckung mit einem OR von 0,99 und

dem P-Wert von 0,69 allerdings nicht bestätigen, obwohl es keine größeren Unterschiede in den eingeschlossenen Datensätzen oder in den verwendeten analytischen Ansätzen gab [55]. Die aktuellste und derzeit größte GWAS zum AD-Risiko, die im Jahr 2019 von Jansen et al. veröffentlicht wurde [48], zeigte erneut eine genomweite Signifikanz für die Assoziation zwischen rs3865444 und dem Risiko der LOAD (P-Wert von  $6.3 \times 10^{-9}$ ). In dieser Metaanalyse wurden 71.880 Fälle und 383.378 Kontrollen eingeschlossen. Um die statistische Aussagekraft zu erhöhen, wurde in dieser GWAS die „AD-by-Proxy“ Klassifizierung verwendet, womit bereits eine vorhandene elterliche AD Diagnose als Kriterium genügte, um in die Gruppe der AD-Fälle eingeschlossen zu werden. Zusammenfassend lassen die derzeit verfügbaren GWAS auf einen genomweit signifikanten Zusammenhang zwischen dem CD33-Gen und dem AD-Risiko schließen [35, 41].

Die meisten bis dato durchgeführten GWAS beziehen sich ausschließlich auf Probanden von nordeuropäischer („weißer“) Abstammung. Andere ethnische Gruppen wurden v.a. im Kontext von Kandidatengenstudien infolge der primären GWAS-Ergebnisse untesucht. So gibt es z. B. zahlreiche genetische Assoziationsstudien in Studienpopulationen ostasiatischer Abstammung (z. B. aus China oder Japan), die auf eine Assoziation der am häufigsten untersuchten CD33 SNPs getestet haben [68, 106, 118]. Weitere Assoziationsstudien gab es auch in anderen Abstammungsgruppen, z. B. in der Gruppe mit weiß-hispanischer Herkunft [83, 91] oder in der Gruppe mit afrikanischer Abstammung [64]. Qualitativ beurteilt deuten diese Studien darauf hin, dass CD33 auch in anderen Abstammungsgruppen eine Assoziation mit dem AD-Risiko aufweist, allerdings erscheint dies aufgrund der teilweise widersprüchlichen Datenlage nicht ganz so eindeutig wie bei den Studien mit nordeuropäischer Abstammung. Um hierzu eine quantitative Einschätzung zu erlangen, wurden anhand der publizierten Daten bereits mehrere Metaanalysen zum Thema durchgeführt. Exemplarisch seien hier die zwei neuesten Metaanalysen aus dem Jahr 2018 aufgeführt und näher beschrieben, nämlich die Studien von Jiang et al. [50] und von Moreno-Grau et al. [72].

Die umfangreichste Metaanalyse zur Assoziation zwischen dem Gen CD33 und AD wurde von Jiang et al. [50] erfasst und in der Zeitschrift *Annals of Translational Medicine* publiziert. Die Autoren haben 17 Studien (inkl. 3 GWAS) mit insgesamt 127.435 Teilnehmern in die Metaanalyse zu dem SNP rs3865444 eingeschlossen und 4 Studien mit 1.810 Probanden in die Metaanalyse zu dem SNP rs3826656. Zudem wurde eine Subgruppenanalyse von verschiedenen Populationen durchgeführt. Die Autoren haben eine signifikante Assoziation zwischen dem A-Allel des SNPs rs3865444 und einem geringeren AD-Risiko festgestellt (OR = 0.94; 95%-CI: 0.90 – 0.98). In der Subgruppenanalyse konnte dieser Effekt lediglich

in der Gruppe mit europäischer Abstammung bestätigt werden ( $OR = 0.92$ ; 95%-CI: 0.90 – 0.94). In der ostasiatischen Population zeigte sich dieser Zusammenhang nicht signifikant ( $OR = 0.87$ ; 95%-CI: 0.65 – 1.17). Des Weiteren zeigten die Ergebnisse dieser Studie keine signifikante Assoziation zwischen dem SNP rs3826656 und dem AD-Risiko ( $OR = 0.94$ ; 95%-CI: 0.62 – 1.41) [50]. Nach einer detaillierten Auseinandersetzung im Rahmen dieser Dissertation mit der Publikation von Jiang et al. wurde ersichtlich, dass diese Metaanalyse mehrere methodischen Fehler aufweist, auf die ich in der Diskussion (Kapitel 5.2) noch genauer eingehen werde.

Eine weitere Metaanalyse zu diesem Thema wurde von Moreno-Grau et al. [72] in der Zeitschrift *Oncotarget* publiziert. Diese Studie hat sich hauptsächlich mit dem SNP rs3865444 des *CD33*-Gens beschäftigt. In der Metaanalyse wurden nicht nur die Daten aus einer eigenen Fall-Kontroll-Studie eingeschlossen, sondern auch die von 13 weiteren Studien inkl. einer GWAS. Die gesamte Stichprobengröße umfasste 90.913 Personen. Die Ergebnisse dieser Metaanalyse haben mit dem o.g. SNP rs3865444 ebenfalls keine signifikante Assoziation mit dem AD Risiko gezeigt ( $OR = 0.98$ , 95%-CI: 0.93 – 1.04, P-Wert = 0.48) [72]. In der Diskussion unten (Kapitel 5.2.2) wird auch auf die Qualität dieser Studie noch im Detail eingegangen.

Basierend auf den ersten GWAS-Ergebnissen, die auf eine Involvierung von *CD33* in der AD hindeuteten, haben zahlreiche Arbeitsgruppen auch funktionell-genetische Experimente durchgeführt und veröffentlicht. Malik et al. fand z. B. im Jahr 2013 heraus, dass die gesamte *CD33*-mRNA im Gehirn von Alzheimer-Erkrankten um etwa 25% erhöht ist und dass die *CD33*-Expression durch das seltene Allel (A-Allel) des SNP rs3865444 etwas verringert ist [65]. Dasselbe Allel wurde in den oben beschriebenen GWAS mit einem verminderten Risiko der AD assoziiert, so dass basierend auf den Daten von Malik et al. ein möglicher Mechanismus der *CD33*-Assoziation in der Genexpression begründet liegen könnte. Eine andere Genexpressionsstudie von Gricuic et al. bestätigte die Verringerung sowohl der *CD33*-Expression als auch der  $\beta$ -Amyloid-Belastung im Gehirn [41]. Aufgrund der bisher beschriebenen Effektrichtungen könnte also eine verminderte *CD33*-Expression protektiv (bzw. eine höhere Expression schädlich) in Bezug auf das AD-Risiko wirken. Eine graphische Darstellung der Funktion des schützenden seltenen Allels von SNP rs386544 ist in der Abbildung 2 schematisch dargestellt. Die höheren *CD33*-Expressionswerte im Gehirn führen demnach zu weniger AD-typischen neuropathologischen Veränderungen in Gehirnen von Trägern des protektiven A-Allels [41].

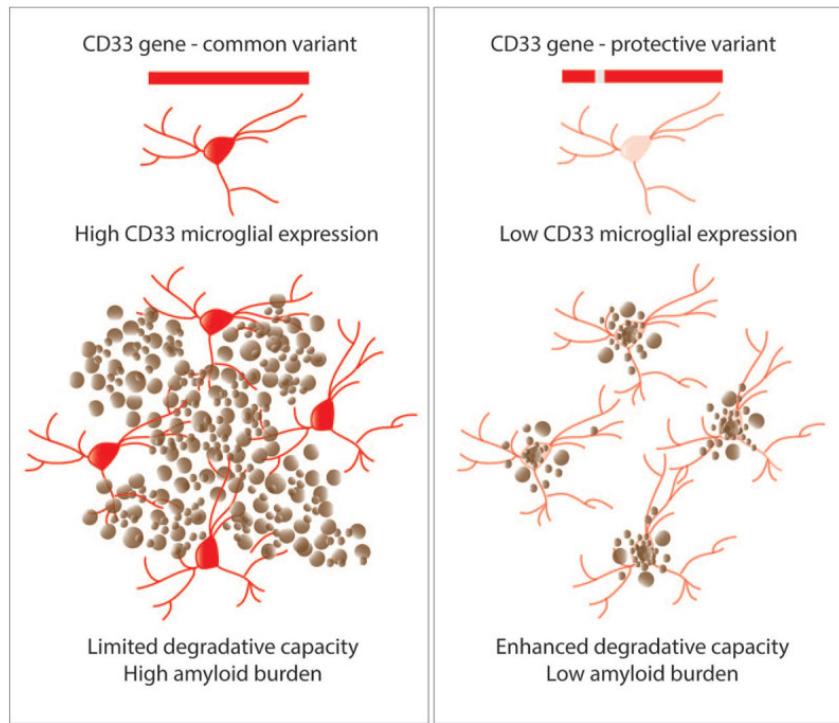


Abbildung 2: Schematische Darstellung der Funktion des schützenden Minor-(A-)Allels des CD33-SNP rs3865444 [41]

*CD33* gehört zu der Immunglobulin-Superfamilie, es hat somit eine Immunglobulin-Domäne und die gleiche Faltungs-Struktur [120]. *CD33* umfasst sieben kodierende Exone und enthält eine V-Typ-Immunglobulin-ähnliche Domäne (V-Ig), mehrere extrazelluläre C2-Ig-Domänen und ein zytosolisches Immunrezeptor-Tyrosin-basiertes inhibitorisches Motiv (engl. „immunoreceptor tyrosinebased inhibitory motif“; ITIM), sowie eine ITIM-ähnliche Sequenz [35, 120]. Die V-Ig-Domäne ist erforderlich für die Hemmung der  $\beta$ -Amyloid-Aufnahme in den Mikrogliazellen [40]. Mithilfe der ITIM-Domäne spielt der *CD33*-Rezeptor eine inhibierende Rolle und führt zu einer Hemmung von vielen zellulären Prozessen der angeborenen Immunantwort wie Phagozytose, Apoptose oder Zytokinfreisetzung [41]. Das seltener A-Allel des *CD33*-SNP rs3865444 erhöht die Wahrscheinlichkeit der alternativen mRNA-Prozessierung zugunsten von Transkripten ohne Exon 2, das V-Ig kodiert [41]. Das wiederum führt zu einer verringerten Expression von *CD33* in voller Länge und einer erhöhten Expression von *CD33*- $\Delta$ V-Ig-Isoform (*D2-CD33*-Isoform), die laut Zhao et al. [120] und Griciuc et al. [41] mit einem verringerten Alzheimer-Risiko einhergeht. Des Weiteren stellten Estus et al. die Hypothese auf, dass die *D2-CD33*-Isoform die Aktivierung von Mikroglia durch ähnliche Mechanismen wie *TREM2* (engl. „triggering receptor expressed on myeloid cells 2“), einem anderen immunsystemrelevanten LOAD-Risikogen [61], auslöst. *CD33* und *TREM2* gehören zu den Immunrezeptoren der Zellmembran, die mit dem Ko-Rezeptor DAP12 interagieren [35]. *TREM2* fördert hierbei die zelluläre Phagozytose und Chemotaxis direkt durch den Ko-Rezeptor DAP12 [41].

Die Phosphorylierung der ITIM-Domäne von CD33 führt indirekt über intrazelluläre Signalmoleküle (Phosphatasen SHP1 und SHP2) zur Inhibierung der zellulären Signalübertragung [41]. Das funktionelle Zusammenspiel zwischen dem CD33, D2-CD33 und TREM2 Rezeptoren ist in der Abbildung 3 schematisch dargestellt.

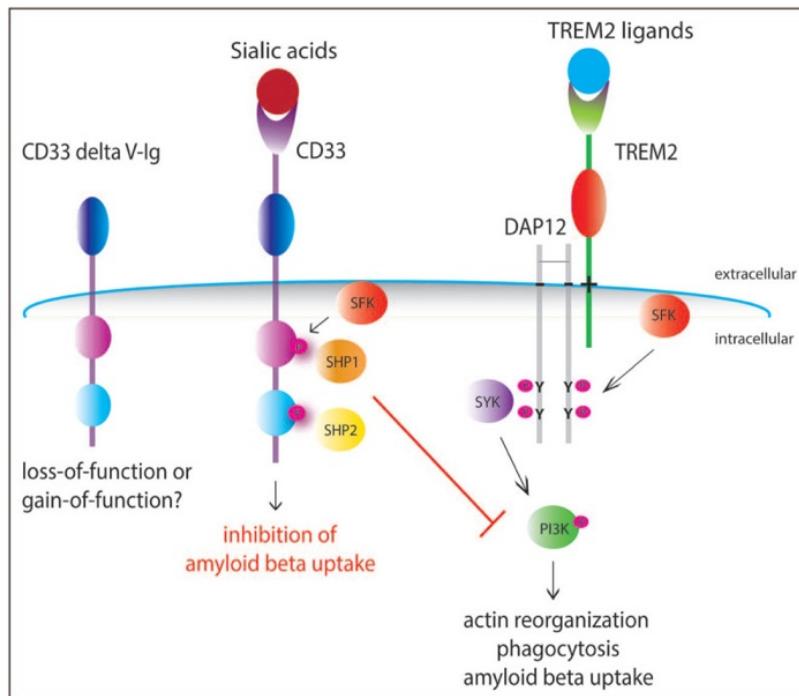


Abbildung 3: Zusammenspiel zwischen den Mikroglia-Rezeptoren CD33 und TREM2 [41]

Laut dieser Hypothese führt also die Verringerung der hemmenden Wirkung von CD33 und/oder die Erhöhung der TREM2-Aktivität zu einer Steigerung der Funktion der Mikrogliazellen. Das wiederum könnte zu einem Rückgang der Ablagerungen von  $\beta$ -Amyloid und Tau-Protein führen und somit zur Reduktion der AD-relevanten neuropathologischen Veränderungen [41].

Generell könnte ein besseres Verständnis der ätiopathogenetischen Mechanismen, ob und wie das CD33-Gen die LOAD beeinflusst, einen weiteren Beitrag zur pharmakologischen Therapieentwicklung leisten [6]. In diesem Zusammenhang wurde die Vermutung aufgestellt, dass die Reduzierung des Einflusses des CD33-Proteins durch pharmakologische Mittel, wie z. B. Antikörper-Therapien, das Risiko und den Schweregrad der AD verringern könnten [35].

## 2.5 Fragestellung und Ziel der Arbeit

Wie oben ausgeführt, weisen die bisher zur Rolle des CD33-Gens in Bezug auf das AD-Risiko durchgeführten genetischen Assoziationsstudien unterschiedliche, teilweise widersprüchliche Effekte auf. Es ist daher von Interesse die aktuelle Datenlage zur Assoziationsevidenz zwischen Polymorphismen im CD33-Gen in publizierten Fall-Kontroll-Studien systematisch und quantitativ aufzuarbeiten und nach dem heutigen Stand der Evidenz zu beurteilen. Des Weiteren weist die Expression des CD33-Gens im Gehirn von AD-Patienten, die das protektive Allel eines der AD-assoziierten SNPs (rs3865444) tragen, in einigen Studien geringere Werte auf. Dies könnte zu einem vermehrten Abbau des schädlichen A $\beta$ 42-Proteins führen und somit das in den genetischen Assoziationsstudien beobachtete verminderte Krankheitsrisiko erklären. Diese Vorarbeiten bedürfen allerdings noch einer unabhängigen Replikation und Ausweitung, um z. B. andere genetische Faktoren, die die CD33-Genexpression im Gehirn regulieren, zu identifizieren. Beide Punkte werden in der vorliegenden Arbeit in den folgenden zwei Fragestellungen bearbeitet:

**1. Welche genetischen Polymorphismen des CD33-Gens beeinflussen das Risiko der AD unter systematischer Betrachtung der derzeit verfügbaren Literatur?**

*Zur Beantwortung dieser Fragestellungen wurde ein systematischer Review der verfügbaren Literatur durchgeführt und die exzerpierten Daten anhand von Metaanalysen quantitativ ausgewertet. Hierbei wurde insbesondere auch der Frage nachgegangen, ob Unterschiede in den Risikoeffekten nachweisbar sind, wenn die Studienpopulationen nach ethnischer Abstammung getrennt analysiert werden.*

**2. Welche genetischen Polymorphismen zeigen eine Assoziation mit der Expression des CD33-Gens im Gehirn von AD Patienten und Kontrollpersonen?**

*Zur Beantwortung dieser Fragestellung wurden genomweite Genotypdaten mit CD33-Gehirnexpressionsdaten zusammengeführt und im Kontext einer sog. eQTL-GWAS ausgewertet. Bei der Auswertung wurde ein besonderes Augenmerk auf die Frage gelegt, ob es eine Überlappung mit den meta-analysierten AD Risiko-SNPs aus der ersten Fragestellung gibt.*

### 3 Material und Methoden

#### 3.1 Metaanalyse und systematischer Review

Zur Beantwortung der ersten Fragestellung meiner Dissertation wurde ein systematischer Review der Literatur mit anschließenden Metaanalysen als Verfahren genutzt.

Unter einem systematischen Review wird eine Übersichtsarbeit verstanden, welche versucht zu einer konkreten Fragestellung den aktuellen Forschungsstand zu identifizieren, Primärquellen zu selektieren und zu bewerten, sowie die zugehörigen relevanten Daten zu sammeln [1, 25, 70]. Hierfür wurden die Primärquellen (hier: genetische Assoziationsstudien) gesucht, identifiziert, gelesen und im Falle der entsprechenden Relevanz die Daten in eine Tabelle extrahiert (s. Anhang 1).

Unter einer Metaanalyse wird die systematische Zusammenfassung und anschließende Darstellung von Ergebnissen einzelner Primärquellen verstanden. Hierfür werden durch statistische Methoden die einzelnen Ergebnisse der inkludierten Studien – unter Berücksichtigung ihres jeweiligen Größeneffektes – quantitativ in ein mathematisches Gesamtergebnis synthetisiert. Die entsprechenden Auswertungen werden häufig in Form von sog. Forrest-Plots abgebildet. [12, 38, 70]

Aufgrund der Objektivität beim Analysieren und Kombinieren verschiedener Ergebnisse werden Metaanalysen als wissenschaftliche Methodik mit einer der höchsten qualitativen Beweiskraft angesehen [1].

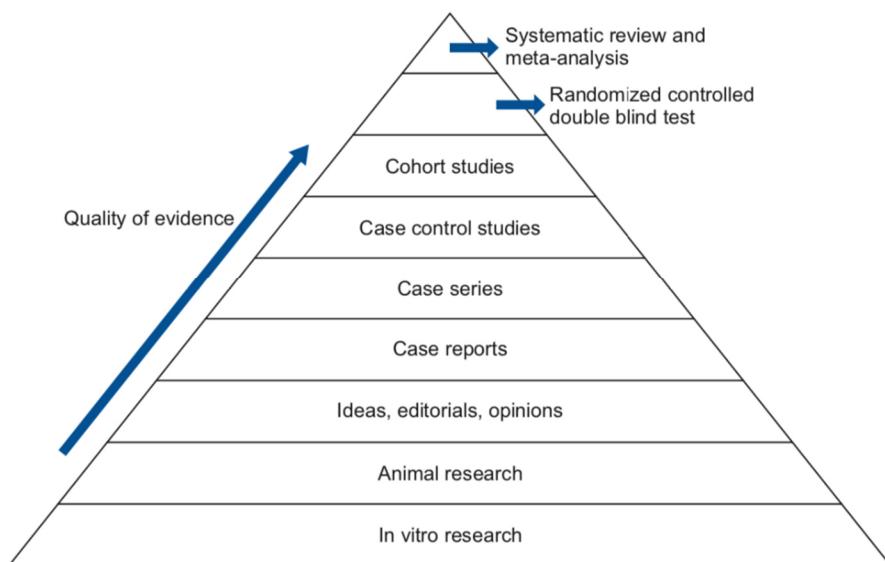


Abbildung 4: Qualitative Beweiskraft unterschiedlicher wissenschaftlichen Methoden [1]

### 3.2 Methodische Vorgehensweise

Es wurde nach etablierten wissenschaftlichen Vorgehensweisen und Leitlinien zur Durchführung und Auswertungen von systematischen Reviews und Metaanalysen gesucht, um sich an diesen in der vorliegenden Arbeit zu orientieren.

Mittlerweile beschäftigt sich eine Vielzahl an Organisationen mit zugehörigen Datenbanken, Leitlinien (engl. guidelines) und Checklisten mit der Durchführung von systematischen Reviews. In diesem Zusammenhang sind u.a. das Cochrane Handbook for Systematic Reviews of Interventions [43], das Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [84], das National Institute for Health and Care Excellence (NICE) [79], und das NHS Centre for Reviews and Dissemination [117] genannt. Die in publizierten systematischen Reviews am häufigsten aufgeführten Referenzen sind das PRISMA Statement und das Cochrane Handbook [52]. In der vorliegenden Arbeit habe ich mich dementsprechend auch an dem PRISMA Statement und dem Cochrane Handbook orientiert, um eine methodisch qualitativ hochwertige Durchführung des systematischen Reviews sowie der Auswertung der extrahierten Daten mittels Metaanalyse sicherzustellen.

Cochrane ist eine in dem Vereinten Königreich ansässige internationale Non-Profit Organisation, welche 1993 gegründet worden ist und mit Mitgliedern aus über 130 Ländern das Ziel verfolgt, innerhalb der Gesundheitsbranche hoch-qualitative Informationen bereitzustellen, um entsprechende fundierte Entscheidungen zu. Hierfür werden insbesondere auch systematische Reviews erstellt und deren Daten anschließend in Form von Metaanalysen zusammengefasst. Das Cochrane Handbook for Systematic Reviews of Interventions beschreibt neben den Grundprinzipien, die Vorgehensweise zur Erstellung systematischen Reviews und Metaanalysen, um diese innerhalb dem Cochrane Netzwerk zu veröffentlichen [43, 86].

PRISMA ist ein im Jahr 2009 überarbeiteter empfehlender Leitfaden, basierend auf dem zuvor durch ein internationales Netzwerk an Forschern im Gesundheitsbereich entwickelten QUOROM Statement („Quality Of Reporting Of Meta-Analyses“), welches vorrangig das Ziel verfolgt, die Qualität von Metaanalysen randomisierter kontrollierten Studien zu verbessern. Das PRISMA Statement beinhaltet eine PRISMA Checkliste mit 27 Punkten und einem PRISMA Flow Diagram, welches in vier Phasen aufgeteilt ist [70, 84, 86].

Der PRISMA Statement Leitfaden und der Leitfaden in Form des Cochrane Handbook for Systematic Reviews of Interventions sind miteinander kompatibel. Das von Cochrane entwickelte „MECIR“ Protokoll (engl. methodological expectations for the conduct, reporting and updating of systematic reviews of interventions) zitiert z. B. an unterschiedlichen Stellen

direkt die Empfehlungen des PRISMA-Statements und hat deren Vorgehensweise entsprechend eingearbeitet, um die Konformität zum PRISMA-Statement sicherzustellen [43, 70, 86]. Mit Orientierung an dem Cochrane Handbook und dem PRISMA Statement leitete sich daher auch die weitere methodische Vorgehensweise dieses Teils meiner Dissertationsarbeit ab. Des Weiteren habe ich die PRISMA Checklist von 2020 [70, 84] ausgefüllt (s. Anhang 2).

### 3.3 Auswahlkriterien inklusive Ein- und Ausschlusskriterien

Um zu bestimmen ob identifizierte Studien im späteren Auswahlprozess berücksichtigt werden wurden zunächst Ein- und Ausschlusskriterien festgelegt.

Die Einschlusskriterien zur Auswahl der Studien wurden wie folgt festgelegt:

- Genetische Assoziationsstudien mit direktem Bezug zur AD
- Studien in denen das CD33-Gen explizit benannt wird
- Studie englischsprachig

Die Ausschlusskriterien zur Auswahl der Studien wurden wie folgt festgelegt:

- Studien ohne Angaben zu Fall- und Kontrollgruppen
- Studien mit geringen Fall- und Kontrollgruppen (< 100 Teilnehmern)
- Studien zu denen keine Volltext-Publikation gefunden wurde

Eine detaillierte Gruppierung der Studien ist nicht erforderlich gewesen.

### 3.4 Informationsquellen und Datensuchstrategie

Zur Identifizierung der relevanten Studien wurde eine Literaturrecherche über das Online-Portal „PubMed“ des US-amerikanischen „National Center for Biotechnology Information“ (NCBI) durchgeführt [74]. Zusätzlich wurden in so identifizierten Studien die aufgeführten Referenzen geprüft, um ggf. weitere relevante Studien zu identifizieren. Die letzte Auswertung der Suchergebnisse bei PubMed hat am 31.12.2020 stattgefunden. Das Datum **31.12.2020** stellt somit den aktuellsten Stand der vorliegenden Arbeit dar.

Als Datensuchstrategie erfolgte auf PubMed die Suche mit dem folgenden „keywords string“: „CD33 AND ALZ“. Die Suchkriterien wurden bewusst sehr generell gefasst, um möglichst viele relevante Studien zu identifizieren. Hierbei wurde in Kauf genommen, dass die allermeisten der unter diesem String identifizierten Studien *keine* genetischen Assoziationsstudien zu CD33-Polymorphismen darstellten.

### 3.5 Auswahl der Studien und Datensammlung

Die Beurteilung, ob eine identifizierte Studie die Einschlusskriterien erfüllt bzw. ob vorhandene Ausschlusskriterien dazu führen, dass eine Studie nicht berücksichtigt wird erfolgte manuell und durch mich. Anschließend erfolgte eine unabhängige Nachkontrolle der durchgeführten Studienauswahl und eingeschlossenen Daten durch eine fachkundige Mitarbeiterin der Arbeitsgruppe (Frau Olena Ohlei), sowie stichprobenhaft durch Prof. Bertram selbst.

Es wurden keine Werkzeuge zur Automatisierung des Auswahlprozesses genutzt, weder bei der initial manuell durchgeführten Studienauswahl noch bei der Nachkontrolle. Den in die Auswahl einbezogenen Studien wurden die Informationen manuell entnommen und tabellarisch in einer Microsoft-Excel-Datei gespeichert, s. Anhang 1.

Um Heterogenität zwischen den in die Metaanalyse eingeschlossenen Studien zu minimieren und eine einheitliche, standardisierte Vergleichbarkeit zu schaffen erfolgte mit den vorhandenen Daten zunächst eine Plausibilitätsprüfung der OR (Odds Ratio). In den Ergebnissen der eingeschlossenen Studien, die nicht nach dem additiven genetischen Transmissionsmodell kalkuliert worden sind (sondern z. B. unter Annahme eines autosomal-dominanten oder –rezessiven Modells), wurde die Berechnung der OR und SE eigenständig durchgeführt, um die Ergebnisse nach additivem Modell darzustellen. In den Fällen, wo einzelnen Daten zu Allelen nicht vorhanden waren erfolgte eine Berechnung der Allelenfrequenzen aus den MAF (engl. „minor allele frequency“), sowie reportierten Zahlen zu verwendeten Fällen und Kontrollen der jeweiligen Publikation. Generell wurden, wenn verfügbar, Daten aus unadjustierten Assoziationsanalysen entnommen, d.h. Daten ohne Anpassung an Alter, Geschlecht und / oder APOE Status, um die Vergleichbarkeit der Ergebnisse zwischen den Studien zu erhöhen.

Die Berechnung der Allelfrequenzen (z. B. aus Genotyphäufigkeiten) erfolgte mit der Software Excel und die Berechnung der OR und SE basierend auf den vorhandenen Allelenfrequenzen erfolgte mit der Software RevMan 5.4.1 der Cochrane Organisation [23].

### 3.6 Statistische Auswertung und Reduzierung von Biases in der Metaanalyse

Für die Berechnung der nachfolgend durchgeführten Metaanalysen wurde die Software RStudio in der Version 1.2.5019 von dem Unternehmen RStudio, Inc. genutzt. Dies erfolgte v.a. unter Anwendung des „R package meta“ in der Version 4.15-1 basierend auf den Konzepten und Modellen zur Metaanalyse von Schwarzer et al. (2015) [94]. Des Weiteren wurden mit der Skriptsprache R Code die Metaanalyse-Ergebnisse in Form von Forrest-Plots visualisiert, um z. B. die OR mit 95% Konfidenzintervall abzubilden.

Als statistisches Modell für die Metaanalysen wurde vorrangig das sog. „random effects“ Modell anstelle eines „fixed effect“ Modells verwendet [12]. Fixed-effect-Metaanalysen wurden nur dann berechnet, wenn die folgenden zwei Kriterien erfüllt waren: 1) alle eingeschlossenen Studien bezogen sich auf dieselbe Abstammungsgruppe und 2) es gab keinen Hinweis auf eine Heterogenität größer 50% zwischen den einzelnen Studienergebnissen. Letztere wurde mittels der  $I^2$ -Metrik bewertet [44].  $I^2$  quantifiziert den prozentualen Anteil der Gesamtvariation über die eingeschlossenen Studien hinweg, der auf Heterogenität zurückzuführen ist.  $I^2$  wird aus der Q-Statistik (Cochran's Heterogenitätsstatistik) berechnet und üblicherweise wie folgt interpretiert:  $I^2$ -Werte bis 25% werden als „geringe“ Evidenz für Heterogenität angesehen, Werte ab 75% als „starke“ Heterogenität und Werte dazwischen als „moderat“ [44]. Allerdings wird  $I^2$  nicht als Teststatistik betrachtet, sondern nur als Maß [90]. Eine signifikante Heterogenität besteht bei einem nominellen Signifikanzniveau von 5 % [90]. Die verwendeten R-Skripte zur Durchführung der Metaanalysen ist in Anhang 3 beigefügt.

Biases (Verzerrungsrisiken) entstehen gemäß der Cochrane Collaboration und dem AWMF Institut für Medizinisches Wissensmanagement entweder in der einzelnen individuellen Studie oder im Ergebnis der durchgeführten Metaanalyse, indem beispielsweise einzelne Studien bewusst nicht berücksichtigt werden, dessen Ergebnisse nicht dem Interesse des Autors der Metaanalyse entsprechen. Zur Reduzierung derartiger Verzerrungen in den hier durchgeführten Metaanalysen wurde die AMSTAR Checkliste verwendet [43, 93, 97], s. Anhang 4.

Das Hardy-Weinberg-Equilibrium (HWE) wurde verwendet, um die eine generelle Qualitätsüberprüfung der eingeschlossenen Studien zu leisten, da Abweichungen vom HWE auf technische Probleme der verwendeten Genotypisierungsassays hinweisen können [47]. In diesem Zusammenhang sei bereits bemerkt, dass die Genotypeverteilungen in allen eingeschlossenen Studien in keinem Fall eine Abweichung vom HWE zeigten.

Ein weiterer wichtiger Grund für Verzerrungen von Metaanalyseergebnissen ist das Vorliegen eines sog. „Publication Bias“ [12]. Dieser Bias entsteht z. B. dann, wenn nicht alle durchgeführten Studien veröffentlicht werden, insbesondere die Studien mit einem negativen (im Sinne von: nicht statistisch signifikanten) Ergebnis. Der Hintergrund kann der Glaube sein, dass Studien mit schwachen und nicht signifikanten Effekt als nicht wertvoll erachtet werden, oder eigenen Ergebnissen zuwiderlaufen. Um einen möglichen Publication Bias zu identifizieren werden häufig sog. Funnel-Plots erstellt. Sind die in diesen Plots dargestellten „Trichter“ (engl. Funnel) nicht symmetrisch, könnte dies auf einen Publication Bias hinweisen [12, 24, 43, 93]. Da die Bewertung der Symmetrie der Funnel-Plots subjektiv ist, wird zur Objektivierung der Datenlage häufig der von Egger et al. vorgeschlagene „Egger-Test“ (eine bestimmte Art der Regressionsanalyse zur Bestimmung der Assymetrie in Funnel-Plots) durchgeführt. Liegt der p-Wert dieses Egger-Tests unter <0,05 kann dies auf das Vorliegen eines signifikanten Publikations-Bias hinweisen [32, 94, 95]. In der vorliegenden Arbeit wurde für jede durchgeführte Metaanalyse ein Funnel-Plot erstellt und bewertet, s. Anhang 5. Zur Objektivierung einer möglichen Assymetrie wurde auch der Egger Test durchgeführt, s. Anhang 6.

### 3.7 Datengenerierung und statische Auswertung eQTL-GWAS

*Humane Proben.* Für die eQTL-GWAS-Analysen konnte ich auf entsprechende Hochdurchsatzdaten der Arbeitsgruppe von Prof. Bertram zurückgreifen, die bisher mit dieser Fragestellung noch nicht ausgewertet wurden. Bei den verwendeten Proben handelte es sich um DNA- bzw. RNA-Proben, die aus *postmortem* gesammelten Gehirnproben (entorhinaler Kortex) von insgesamt 90 AD-Fällen und 90 Kontrollen extrahiert wurden. Die Proben wurden von Kooperationspartnern der Oxford BrainBank gesammelt und an die Arbeitsgruppe im Rahmen des „MiRNetAD“-Projekts transferiert und dort molekular analysiert [29, 103].

Die Erlaubnis zur Analyse der Biomaterialien (DNA- bzw. RNA-Proben) wurde durch die Ethikkommission der Universität zu Lübeck (Aktenzeichen 19-392A vom 22.09.2019) bestätigt, s. Anhang 7

*Genotypisierungsdaten.* Für die genoweite SNP-Genotypisierung der 180 Studienteilnehmer wurde der „Global Screening Array“ (GSA) mit „shared custom content“ (Illumina, Inc.) benutzt, der insgesamt 696.375 SNPs je DNA-Probe enthält. Die molekularen Daten wurden am Institut für Klinische Molekularbiologie (IKMB) generiert und durch Frau Olena Ohlei aus der Arbeitsgruppe LIGA einer initialen Qualitätskontrolle (engl.

Quality Control; QC) und nachfolgend Genotypimputation unterzogen (s. [46] für eine detaillierte Darstellung der Methoden).

Die „post-imputation“ QC wurde dann von mir auf dem „Omics“ Rechencluster der Universität zu Lübeck in der Linux-Umgebung mittels der PLINK v2 Software [11] vorgenommen. Auf der SNP-Ebene umfasste dies den Ausschluss von SNPs mit folgenden Kriterien: 1) Imputations-Qualitäts-Score  $r^2 < 0.7$ ; 2) MAF  $< 1\%$ ; 3) Genotypisierungseffizient  $< 0.98\%$ , 4) HWE-Deviation  $p < 5e-6$  (in Kontrollen). Auf der Sample-Ebene erfolgte der Ausschluss von Probanden mit folgenden Kriterien: 1) Personen mit Diskrepanzen in genetischen vs. reportierten Geschlecht; 2) Duplizierte Samples / MZ-Zwillinge; 3) Verwandschaftsverhältnisse  $>$  vierten Grades ( $\hat{\pi}_{i,i} > 0.05$ ); 4) Proben mit implausibler Heterozygotie ( $\text{mean } (\text{het}) \pm 6 \times \text{SD}$ ). Die Erfassung der Populationszugehörigkeit erfolgte mittels Principal-Component-Analyse (PCA). Nach der post-Imputations-QC verblieben 7.417.011 SNPs von 177 Probanden (davon 87 Fälle und 90 Kontrollen) zur eQTL-GWAS-Analyse.

*Genexpressionsdaten.* In denselben Gehirnproben wurden transkriptom-weite Genexpressionsprofile mittels RNA-Sequenzierung generiert, die als abhängige Variable („expression“) in den eQTL-Analysen benutzt wurden. Die molekularen Daten wurden ebenfalls am IKMB generiert und durch einen Mitarbeiter der Arbeitsgruppe LIGA (Herrn Marcel Schilling) prozessiert und qualitätskontrolliert. Sämtliche experimentellen und bioinformatischen Schritte sind in [29] beschrieben. Für die hier durchgeführten eQTL-GWAS-Analysen wurden mir die normalisierten Expressionsdaten aller Transkripte zum CD33-Gen (Ensembl-ID: ENSG00000105383) [33] überlassen und diese als abhängige Variable in den linearen Regressionsmodellen der eQTL-GWAS verwendet.

*Statistische Analysen.* Die eigentliche eQTL-GWAS-Analysen erfolgten unter Verwendung eines linearen Regressionsmodells unter Verwendung von imputierten SNPs-Dosages (als unabhängigen Variablen) nach dem additiven Transmissionsmodell unter Einbeziehung des Geschlechts und der drei ersten populationsspezifischen Hauptkomponenten aus der oben beschriebenen PCA. Um einen Einfluss des Diagnosestatus zu kontrollieren wurde dieser ebenfalls als Kovariable eingeschlossen. Alle eQTL-GWAS-Analysen erfolgten mittels der PLINK v2 Software [11]. Die Post-GWAS Analysen (d.h. im Kontext dieser Arbeit: Gen-basierte GWAS und Auswertung bestimmter Annotationen der assoziierten SNPs und Gene) wurden mittels des „Functional Mapping and Annotation of Genome-Wide Association Studies“ (FUMA; [111]) Tools durchgeführt.

## 4 Ergebnisse

### 4.1 Metaanalysen genetischer Assoziationsstudien zur Rolle des CD33-Gens in der AD

#### 4.1.1 PRISMA Flow Chart

Die Suche nach den in Kapitel 3.4. beschriebenen Verfahren ergab 146 Studien, die bis zum 31.12.2020 in der PubMed-Datenbank aufgeführt waren. Im Anschluss wurde in den so identifizierten Studien noch manuell nach den Verweisen auf andere relevante Publikationen gesucht und in der Auswahl berücksichtigt. Zusätzlich wurden hierdurch 3 weiteren Studien mit aufgenommen. Insgesamt wurden somit 149 Studien identifiziert.

Die identifizierten Studien wurden anschließend auf die definierten Ein- und Ausschlusskriterien überprüft. Alle Publikationen waren auf Englisch erfasst und von allen lag mir eine Volltextversion vor.

Aus dieser Vorauswahl wurden nach dem Screening der Abstrakte 110 Studien ausgeschlossen, da diese keine genetischen Assoziationsstudien *per se* darstellten. Die 39 verbliebenen Studien wurden im Volltext gesichtet. Hiervon wurden 10 weitere Publikationen ausgeschlossen: 3 Metaanalysen und 2 Übersichtsarbeiten. Die anderen 5 ausgeschlossenen Studien haben sich entweder mit einer anderen Thematik beschäftigt oder haben SNPs lediglich in der Nähe von CD33 identifiziert. Am Ende dieses Prozesses verblieben 29 Studien, deren Daten für die Metaanalysen dieses Teils meiner Dissertation verwendet wurden.

Aus diesen 29 Studien wurden dann die relevanten Daten zur Assoziationsevidenz entnommen und in eine Tabelle überführt. Bei diesem Prozess stellte sich heraus, dass bei drei Studien nicht ausreichend Daten vorhanden waren (z. B. fehlten Angaben zur OR, CI oder SE); diese Studien wurden daraufhin ebenfalls ausgeschlossen. Ferner wurden die 2 GWAS von Lambert et al. [55] und von Naj et al. [73] aufgrund der vollständigen Überlappung der Datensätze mit der GWAS von Jansen et al. [48] ausgeschlossen. Des Weiteren wurden die Datensätze „GERAD1“ und „EADI1“ aus der Studie von Hollingworth et al. [45] und die Datensätze „Mayo2“ („Jacksonville“, „Rochester“ und „Autopsy“) und „ARUK“ aus der Studie von Carrasquillo et al. [20] aussortiert, weil diese bereits in den Datensätzen „IGAP“ und „ADSP“ der GWAS von Jansen et al. [48] zusammengeführt wurden.

Alle 24 in die Metaanalysen eingeschlossenen genetischen Assoziationsstudien sind vom Studiendesign Fall-Kontroll-Studien, bei zwei Studien handelt es sich um GWAS [45] [48]. Für den SNP rs3865444 wurden 21 Studien (22 Datensätze) in die Metaanalyse eingeschlossen [20, 22, 28, 31, 45, 48, 49, 51, 59, 64, 66–68, 71, 72, 83, 91, 102, 106, 110,

119], für SNP rs3826656 6 Studien [48, 51, 59, 66, 71, 118]. Jeweils 3 Studien wurden für die SNPs rs1697553 [22, 45, 48], rs2455069 [45, 48, 92], rs12459419 [48, 51, 92] und rs35112940 [48, 85, 92] einbezogen. Weitere Details zu den eingeschlossenen Studien sind in den Abschnitten zu den jeweilig metaanalysierten SNPs zu finden.

Zur visuellen Darstellung der Vorgehensweise der Studienauswahl wurde sich an „PRISMA Flow Diagram 2020“ orientiert [70, 84]. Das Flussdiagramm ist in der Abbildung 5 dargestellt.

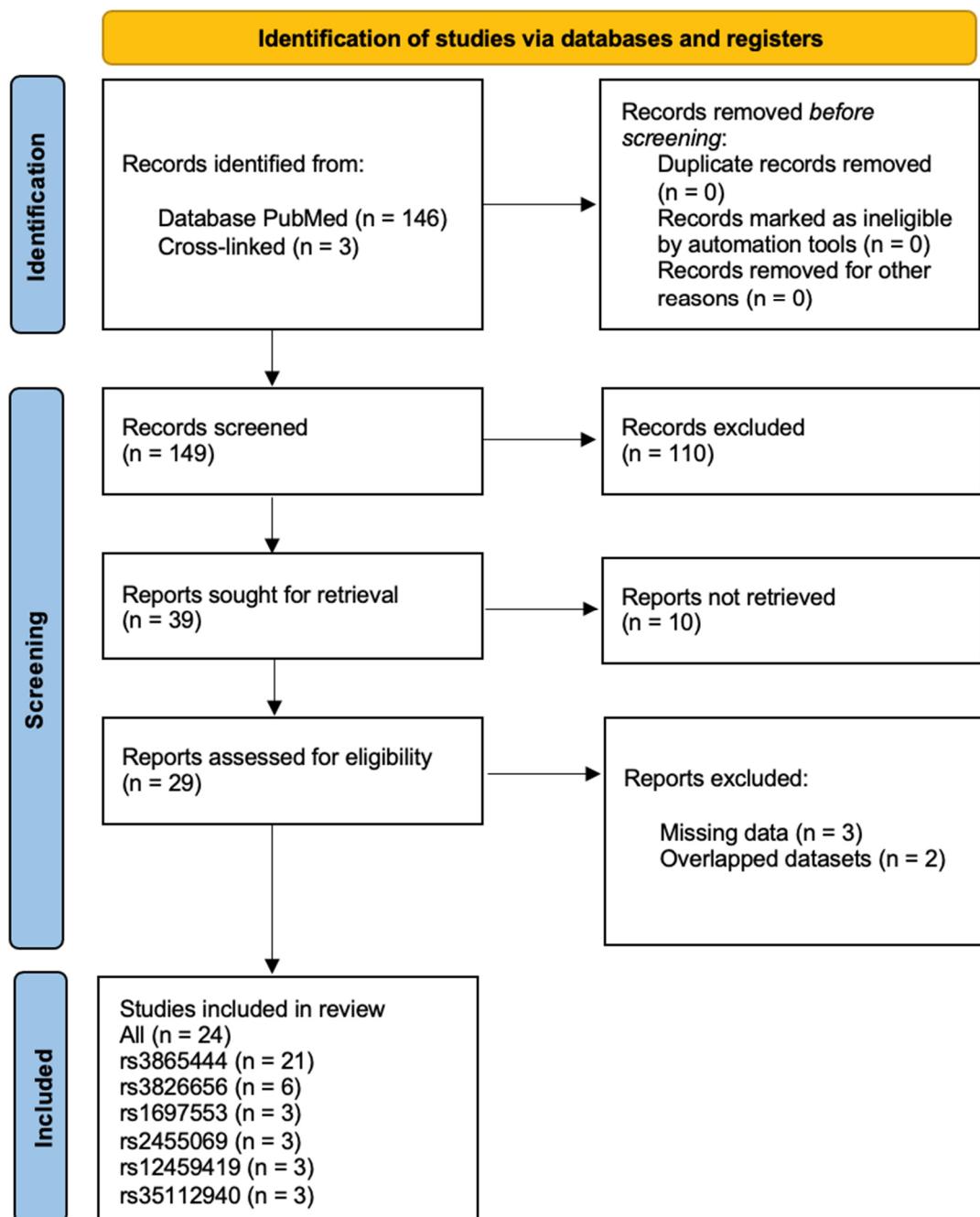


Abbildung 5: Flussdiagramm basierend auf dem PRISMA flow chart 2020

#### 4.1.2 Überblick Ergebnisse der eigenen Metaanalysen

Insgesamt wurden in dieser Arbeit sechs CD33-Polymorphismen (rs3865444, rs3826656, rs1697553, rs2455069, rs12459419, rs35112940) metaanalyisiert. Wenn möglich, erfolgte im jeweiligen SNP eine separate Subgruppen-Analyse nach Stratifizierung für ethnische Abstammung.

Die Stichprobengrößen variierten hierbei je nach SNP von 1.193 zu 476.187. Bei der Analyse der separaten Untergruppen nach ethnischer Abstammung variierten die Stichprobengrößen zwischen 845 (rs3865444 – „hispanic whites“) und 467.369 (rs3865444 – „north european“).

Insgesamt zeigten alle sechs metaanalysierten CD33-Polymorphismen Evidenz für eine nominal signifikante Assoziation mit dem AD-Risiko in mindestens einer der durchgeföhrten Analysen (s. Tabelle 3): drei SNPs zeigten dabei einen signifikant protektiven Effekt für AD (rs3865444, rs12459419 und rs35112940), während drei SNPs mit erhöhtem Risiko für die AD assoziiert waren (rs3826656, rs1697553 und rs2455069).

Population	N Studien	N Probanden		Metaanalyse			Heterogenität	
		Fälle	Kontrolle	OR	95%-CI	P-Wert	$I^2$	P-Wert
<b>rs3865444 A vs. C</b>								
All <sup>a</sup>	21	80784	395403	0.92	0.77 - 1.10	0.36	99%	0
North European	9	76632	390737	0.99	0.98 - 0.99	3.41e-09	24%	0.22
East Asian <sup>a</sup>	8	3180	3588	1.01	0.81 - 1.27	0.93	86%	< 0.01
White Hispanic <sup>a</sup>	2	354	491	0.88	0.63 - 1.23	0.46	53%	0.15
<b>rs3826656 G vs. A</b>								
All <sup>a</sup>	6	73168	384825	1.11	0.95 - 1.29	0.20	73%	< 0.01
East Asian	4	1008	1090	1.25	1.10 - 1.42	8.20e-04	7%	0.36
<b>rs1697553 G vs. A</b>								
All	3	72879	384903	1.01	1.01 - 1.01	5.28e-07	0%	0.53
<b>rs2455069 G vs. A</b>								
North European	3	72921	385025	1.01	1.01 - 1.01	6.87e-07	0%	0.53
<b>rs12459419 T vs. C</b>								
All	3	72441	384372	0.99	0.98 - 0.99	6.48e-09	0%	0.87
<b>rs35112940 A vs. G</b>								
North European	3	72344	384107	0.99	0.98 - 0.99	1.14e-05	33%	0.23

Tabelle 3: Überblick der Ergebnisse der hier durchgeföhrten Metaanalysen; <sup>a</sup> Verwendung des „random effects“ Modell aufgrund mindestens „moderater“ Heterogenität (beurteilt anhand  $I^2$ -Methrik; s. Methoden)

In den nachfolgenden Abschnitten wird auf die jeweiligen sechs SNPs detaillierter eingegangen und die Besonderheiten der Resultate beschrieben.

#### 4.1.3 SNP rs3865444

Dieser Polymorphismus des *CD33*-Gens zog in der von mir im Rahmen dieser Arbeit gesichteten Literatur die meiste Aufmerksamkeit der Forschenden auf sich. Die publizierten GWAS und mehrere Kandidatengen-Studien haben bisher widersprüchliche Ergebnisse erbracht, was die Assoziation zwischen dem SNP rs3865444 und dem LOAD-Risiko betrifft (s. Kapitel 2.4). Die bisher publizierten Metaanalysen zu diesem Thema haben ebenfalls bislang keinen einheitlichen Konsens erzielt (s. Kapitel 4.1.10).

In die Metaanalyse für diesen *CD33*-Polymorphismus wurde die in dieser Arbeit größte Anzahl (21 Publikationen mit 22 Datasets) an genetischen Assoziationsstudien (inkl. 2 GWAS) für LOAD mit der umfangreichsten Stichprobengröße (476.187 Probanden) eingeschlossen.

Abbildung 6 zeigt den Forrest-Plot der eingeschlossenen 21 Studien zu diesem SNP. Diese Metaanalyse erfolgte zunächst ohne Subgruppen-Analyse nach Populationen und zur Auswertung wurde aufgrund der großen Heterogenität der eingeschlossenen Studien ( $I^2 = 99\%$ ) das „random effects“ Modell angewendet.

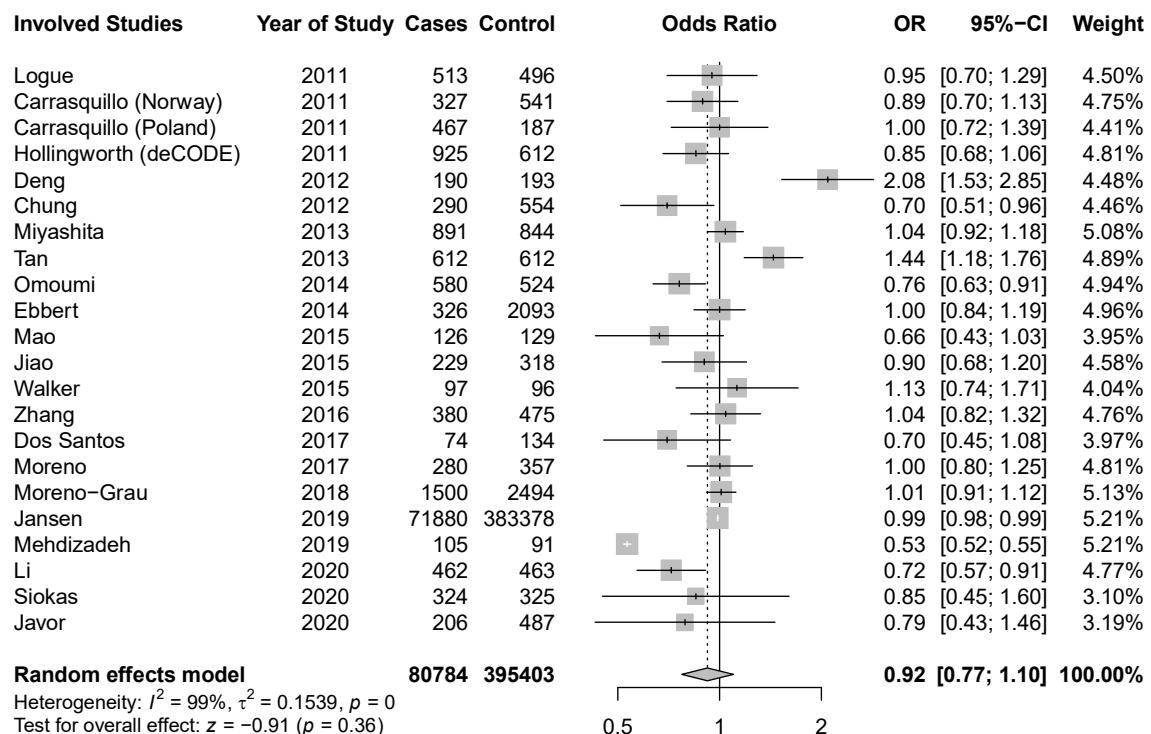


Abbildung 6: Forrest-Plot für den SNP rs3865444 (random effects Modell; A vs. C) nach Einstellung aller ethnischen Abstammungsgruppen

Als Ergebnis zeigt sich, dass das Minor-(A)-Allel von SNP rs3865444 mit einem reduzierten LOAD-Risiko verbunden ist und somit einen protektiven Effekt hat. Das Ergebnis dieser

Metaanalyse war allerdings nicht statistisch signifikant ( $OR = 0.92$ , 95%-CI:  $0.77 – 1.10$ ; P-Wert = 0.36).

Bei Betrachtung des dazugehörigen Funnel-Plots (Abbildung 7) zeigt sich, dass dieser symmetrisch ist. Der Egger-Test bestätigt dies mit einem P-Wert von 0.4156 als objektivem Nachweis diese Einschätzung. Es besteht somit kein Hinweis auf einen „publication bias“.

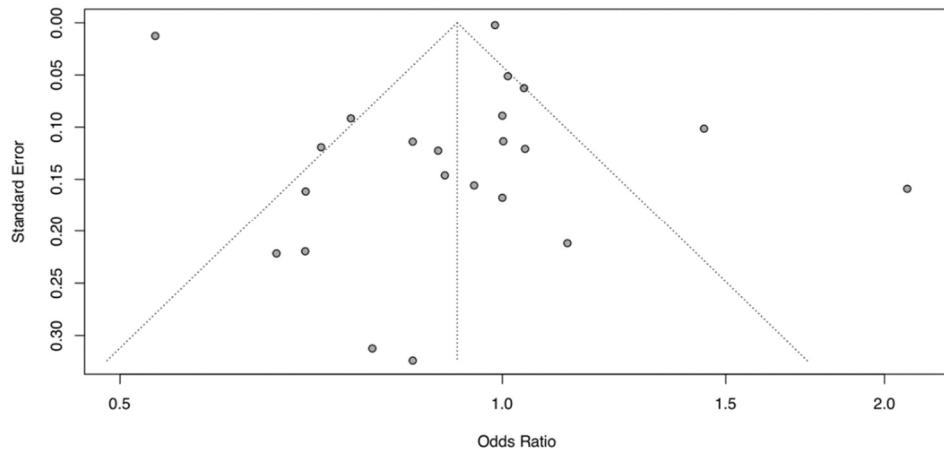


Abbildung 7: Funnel-Plot für den SNP rs3865444 nach Einschluss aller ethnischen Abstammungsgruppen

Im Anschluss wurden die Studien noch einmal unter Berücksichtigung ihrer ethnischen Abstammungen getrennt analysiert, d.h. nach nordeuropäischer Herkunft (engl. „north european“), nach ostasiatischer Herkunft (engl. „east-asian“) und nach hispanischer Herkunft (engl. „white hispanic“). Nur in der Subgruppen-Analyse der Nordeuropäer wurde eine signifikante Assoziation des SNPs rs3865444 (Minor-(A)-Allel) mit einem geringeren AD-Risiko festgestellt ( $OR = 0.99$ , 95%-CI:  $0.98 – 0.99$ ; P-Wert = 3.41e-09). Der Forrest-Plot ist in Abbildung 8 dargestellt. Die zu dem SNP rs3865444 gehörigen Funnel-Plots sind im Anhang 5 dieser Arbeit abgebildet.

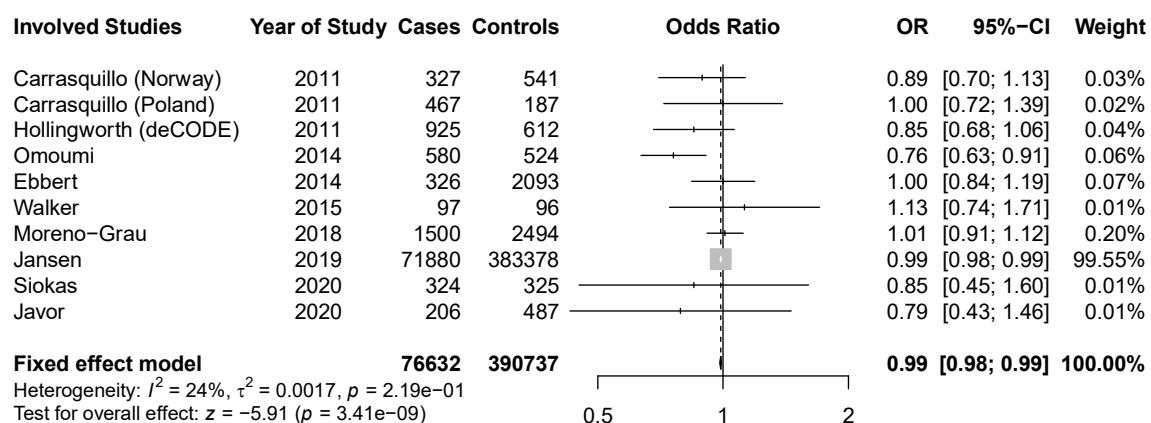


Abbildung 8: Forrest-Plot für den SNP rs3865444 (fixed effects Modell; A vs. C); nur nordeuropäische Abstammung

Auch wenn die Heterogenität der Metaanalyse in der Gruppe der Nordeuropäer niedrig ist ( $I^2 = 24\%$ ), erkennt man, dass das Signal hauptsächlich durch die Daten aus der GWAS von Jansen et al. bedingt ist (Gewichtung: 99,55%). Das Wiederholen der Analyse nach Weglassen der GWAS von Jansen et al. zeigt zwar eine Erhöhung der Effektstärke (OR = 0.94), jedoch verringert sich die Signifikanz (P-Wert = 0.08), s. Abbildung 9.

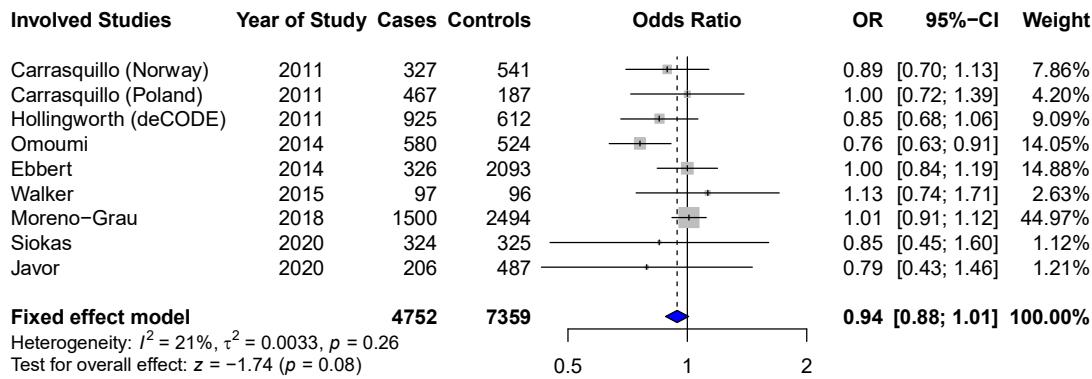


Abbildung 9: Forrest-Plot für den SNP rs3865444 (fixed effects Modell; A vs. C); nur nordeuropäische Abstammung nach Ausschluss der GWAS von Jansen et al. [48]

Bei der Betrachtung der Ergebnisse der Subanalyse in der Gruppe der ostasiatischen Abstammung (Abbildung 10) ist besonders auffällig, dass die Effektrichtung der Summary-OR konträr zu der Effektrichtung in der nordeuropäischen Population ist (OR = 1.01 vs. OR = 0.99). Somit zeigt der SNP rs3865444 in der ostasiatischen Herkunft keinen protektiven Effekt für LOAD, sondern einen (sehr leichten) Risiko-erhöhenden Effekt. Diese Konstellation könnte dadurch erklärt werden, dass der SNP rs3865444 sich in der ostasiatischen Kohorte in ausgeprägtem LD mit einem der anderen CD33-Riskoloci rs3826656 befindet (s. Kapitel 4.1.9). Das Ergebnis der Metaanalyse ist allerdings statistisch nicht signifikant (P-Wert = 0.93). Trotz der Betrachtung von Studien mit ausschließlich ostasiatischer Abstammung ist die Heterogenität der Effekstärkenschätzer der eingeschlossenen Studien hoch ( $I^2 = 86\%$ ); dies ist der Grund für die Anwendung des „random effects“ Modells.

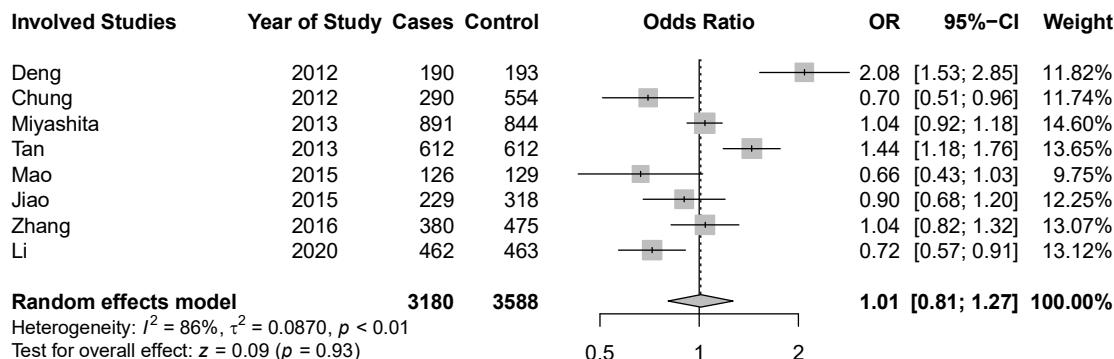


Abbildung 10: Forrest-Plot für den SNP rs3865444 (random effects Modell; A vs. C); nur ostasiatische Abstammung

Die Subgruppen-Analyse der hispanischen Population zeigte eine Assoziation des Minor-(A)-Allels des SNPs rs3865444 mit einem geringeren LOAD-Risiko (OR = 0.88, 95%-CI: 0.63 – 1.23). Die Effektrichtung passt zu der in der nordeuropäischen Population. Das Ergebnis ist allerdings statistisch auch nicht signifikant (P-Wert = 0.46). Man muss dabei jedoch berücksichtigen, dass in dieser Analyse die kleinste Anzahl an Primärstudien eingeschlossen wurde: mit 354 AD-Fällen und 491 Kontrollen umfasste die Stichprobengröße dieser Metaanalyse somit die wenigsten Studienteilnehmer. Aufgrund der relativ hohen Heterogenität der eingeschlossenen Primärstudien ( $I^2 = 53\%$ ) wurde auch hier das „random effects“ Modell zur Auswertung angewendet.

Der entsprechende Forrest-Plot ist in der Abbildung 11 dargestellt.

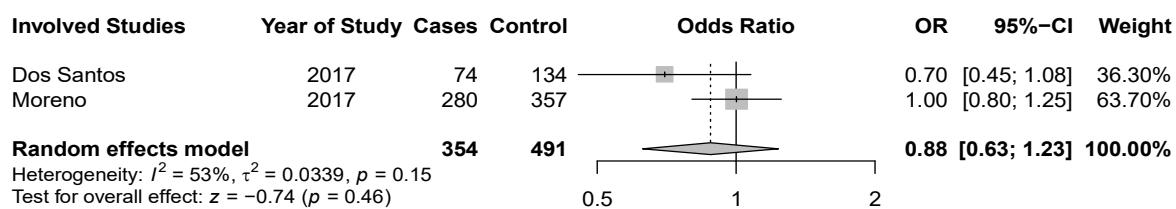


Abbildung 11: Forrest-Plot für den SNP rs3865444 (random effects Modell; A vs. C); nur hispanische Abstammung

#### 4.1.4 SNP rs3826656

Der allererste Report über eine mögliche genetische Assoziation zwischen dem CD33-Gen und dem LOAD-Risiko basierte auf SNP rs3826656 und wurde in einer familien-basierten GWAS der Arbeitsgruppe im Jahr 2008 publiziert [8].

Erwähnenswert ist es, dass sich für diesen Polymorphismus die Allelfrequenzen zwischen verschiedenen Populationen stark unterscheiden, eine Tatsache, die bei den anderen hier untersuchten SNPs nicht der Fall war. Z. B. wird in der nordeuropäischen Abstammungsgruppe das A-Allel als Major-Allel aufgeführt ( $G=0.24$ ,  $A=0.76$  [80]), in der Gruppe der ostasiatischen Herkunft ist es aber andersherum ( $G=0.7$ ,  $A=0.3$  [80]). Die Allelfrequenzen in der hispanischen Kohorte unterscheiden sich hierbei nicht wesentlich von denen der Nordeuropäer ( $G= 0.22$ ,  $A=0.78$  [80]).

Das Ergebnis dieser Metaanalyse über vier unabhängige Studien zeigt, dass in der Gruppe mit ostasiatischer Abstammung das Major-(G)-Allel von SNP rs3826656 nominal signifikant mit einem erhöhtem LOAD-Risiko assoziiert ist (OR = 1.25, 95%-CI: 1.10 – 1.42; P-Wert = 8.2e-04). Der Forrest-Plot ist in der Abbildung 12 dargestellt. Der Heterogenitätswert dieser Analyse ist niedrig ( $I^2 = 7\%$ ), daher wurde hier das „fixed effect“ Modell zur Auswertung angewendet.

Der entsprechende Funnel-Plot ist in der Abbildung 13 zu finden. Der Egger-Test bestätigt mit einem P-Wert von 0.2273, dass kein Hinweis auf einen „publication bias“ besteht, obwohl die Teststärke hierfür aufgrund der kleinen Stichprobengröße nur gering ist.

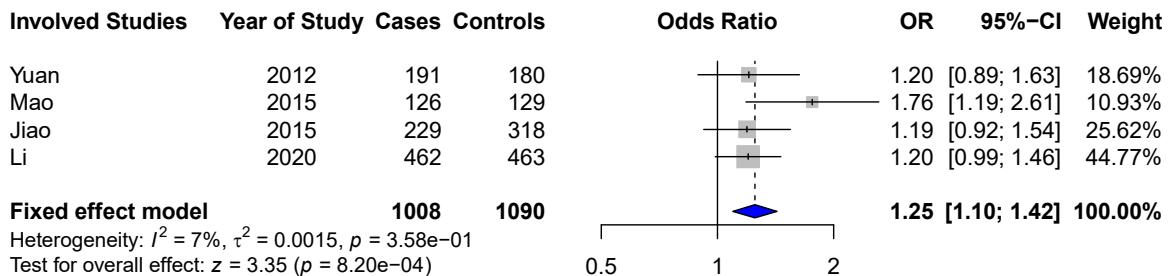


Abbildung 12: Forrest-Plot für den SNP rs3826656 (fixed effects Modell; G vs. A); nur ostasiatische Abstammung

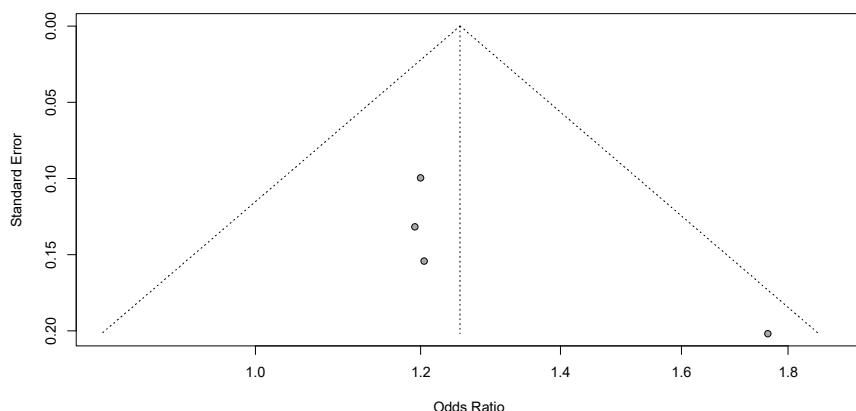


Abbildung 13: Funnel-Plot für den SNP rs3826656; nur ostasiatische Abstammung

Wenn die Daten aus allen Studien über alle Abstammungsgruppen hinweg in einer Analyse kombiniert werden, zeigt das G-Allel ebenfalls ein erhöhtes Risiko ( $OR = 1.11$ ; 95%-CI: 0.95 – 1.29), allerdings ist dieser Effekt nicht statistisch signifikant (P-Wert = 0.2). In der Abbildung 14 ist der dazugehöriger Forrest-Plot dargestellt. Erwartungsgemäß zeigt die Heterogenitätsanalyse für diese kombinierten Analysen einen vergleichsweise hohen Wert ( $I^2 = 73\%$ ).

Weil es nicht genügend unabhängige Datensätze gab ( $n=1$ ), war eine Subanalyse in der nord-europäischen Population nicht möglich. Die Ergebnisse der neuesten Metaanalyse von Jansen et al. ergaben allerdings mit einem Null-Effekt keine Hinweise auf eine Assoziation des SNPs rs3826656 mit AD in der Gruppe der Nordeuropäer ( $OR = 1.0$ ; 95%-CI: 1.0 – 1.01 [48]).

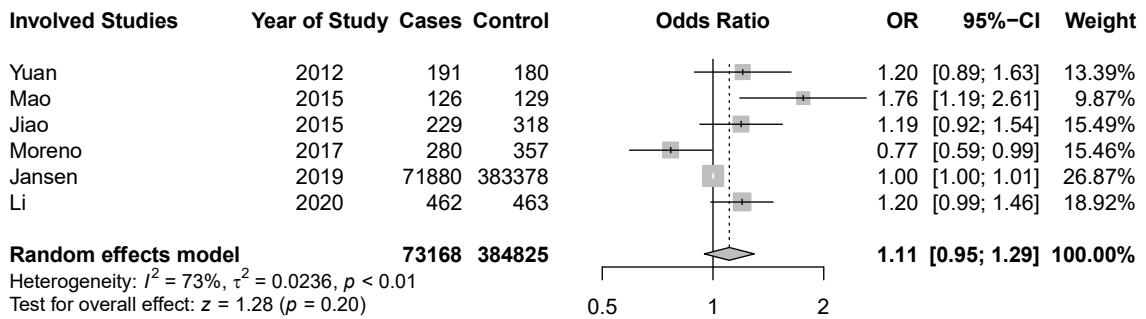


Abbildung 14: Forrest-Plot für den SNP rs3826656 (random effects Modell; G vs. A); nach Einschluss aller ethnischen Abstammungsgruppen

#### 4.1.5 SNP rs1697553

Die Analyse des SNP rs1697553 zeigt, dass das Minor-(G)-Allel eine signifikante Assoziation mit erhöhtem LOAD-Risiko besitzt, wenngleich die OR nur unwesentlich von 1 abweicht (OR = 1.01; 95%-CI: 1.01 - 1.01; P-Wert = 5.28e-07). Der vergleichsweise kleine P-Wert dieser Metaanalyse liegt begründet in der Größe der Stichprobe der GWAS von Jansen et al. (n=455.258). In Abbildung 15 ist der dazugehörige Forrest-Plot dargestellt. Der entsprechende Funnel-Plot ist in Anhang 5 dargestellt. Die weiteren Subanalysen nach ethnischer Abstammung waren in diesem Fall nicht angezeigt, da es nur eine einzige Primärstudie zur ostasiatischen Population gab und diese nur für 0,01% des Gesamteffektes verantwortlich war.

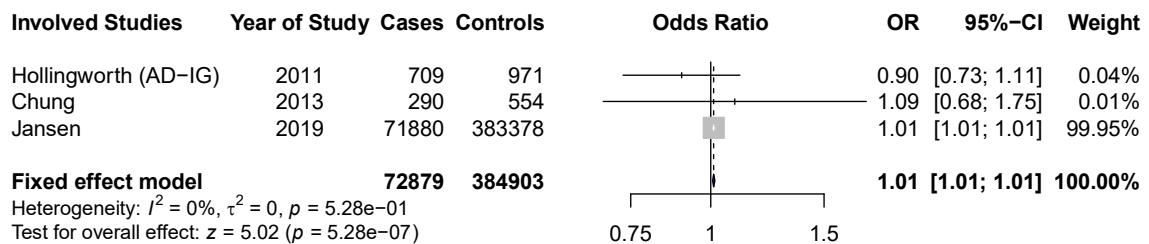


Abbildung 15: Forrest-Plot für den SNP rs1697553 (fixed effects Modell; G vs. A); nach Einschluss aller ethnischen Abstammungsgruppen

#### 4.1.6 SNP rs2455069

Die Metaanalyse des SNP rs2455069 Polymorphismus zeigte ähnliche Ergebnisse wie SNP rs1697553: es wurde eine signifikante Assoziation des Minor-(G)-Allels mit dem LOAD-Risiko in der nordeuropäischen Population festgestellt (OR = 1.01; 95%-CI: 1.01 – 1.01; P-Wert: 6.87e-07). Der Grund, weshalb die Ergebnisse der beiden SNPs (rs2455069 und rs1697553) sich stark ähneln, liegt darin begründet, dass diese Polymorphismen sich im vollständigen LD befinden (s. Kapitel 4.1.9), also die Genotypverteilungen sich nahezu perfekt entsprechen. Der Forrest-Plot zu dem SNP rs2455069 ist in der Abbildung 16 dargestellt. Der dazugehörige Funnel-Plot ist im Anhang 5 zu finden.

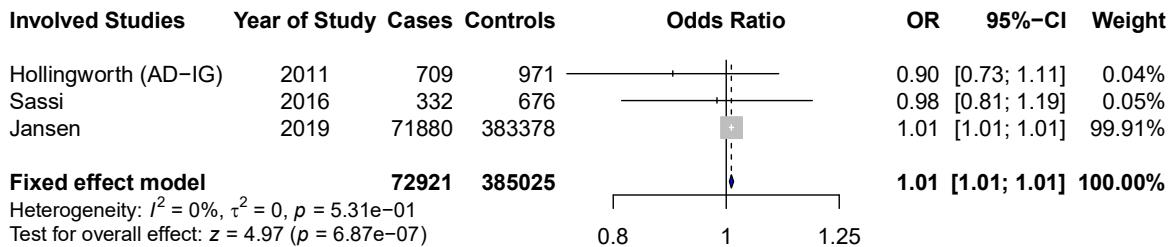


Abbildung 16: Forrest-Plot für den SNP rs2455069 (fixed effects Modell; G vs. A); nur nordeuropäische Abstammung

#### 4.1.7 SNP rs12459419

Betrachtet man die Ergebnisse dieser Metaanalyse ( $OR = 0.99$ ; 95%-CI:  $0.98 - 0.99$ ; P-Wert =  $6.48e-09$ ), zeigt sich, dass das Minor-(T)-Allel des SNPs rs12459419 einen signifikant protektiven Effekt auf das AD-Risiko hat. Man muss dabei jedoch berücksichtigen, dass der CD33-Polymorphismus SNP rs12459419 sich im vollständigen LD mit dem SNP rs3865444 befindet (s. Kapitel 4.1.9). Wie auch in anderen Metaanalysen dieser Arbeit, wurde das meiste Gewicht der GWAS von Jansen et al. [48] zugeordnet. Der Forrest-Plot ist in der Abbildung 17 dargestellt. Im Anhang 5 ist der dazugehörige Funnel-Plot zu sehen.

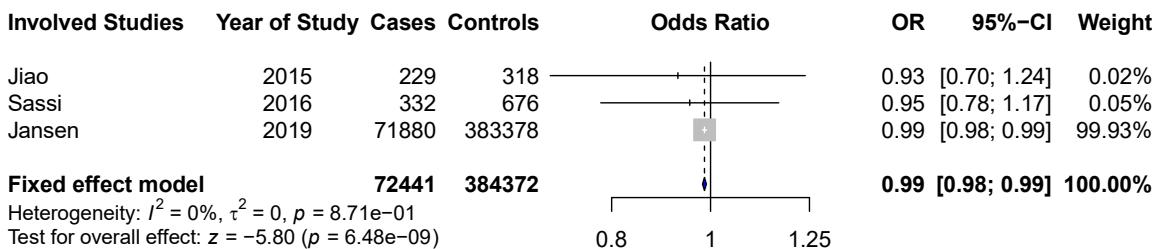


Abbildung 17: Forrest-Plot für den SNP rs12459419 (fixed effects Modell; T vs. C); nach Einschluss aller ethnischen Abstammungsgruppen

#### 4.1.8 SNP rs35112940

Die Ergebnisse ( $OR = 0.99$ ; 95%-CI:  $0.98 - 0.99$ ; P-Wert =  $1.14e-05$ ) der Metaanalysen für diesen SNP zeigen, dass das Minor-(A)-Allel signifikant mit einem geringeren LOAD-Risiko assoziiert ist. Man sollte jedoch bedenken, dass dieser SNP sich in starkem LD mit den Polymorphismen rs3865444 und rs12459419 befindet (s. Kapitel 4.1.9) und die Ergebnisse sich für alle drei Polymorphismen also bedingen. In der Abbildung 18 ist der dazugehörige Forrest-Plot dargestellt. Auch an dieser Stelle ist das Signalgewicht – wie auch schon bei den anderen hier metaanalysierten SNPs – stark durch die GWAS von Jansen et al. beeinflusst. Der Funnel-Plot ist im Anhang 5 abgebildet.

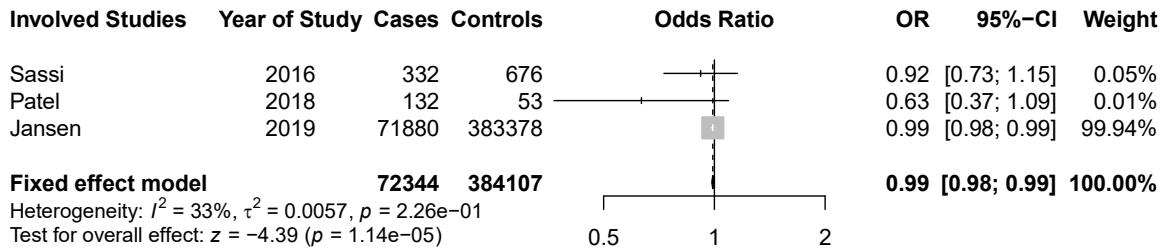


Abbildung 18: Forrest-Plot für den SNP rs35112940 (fixed effects Modell; A vs. G); nur nordeuropäische Abstammung

#### 4.1.9 Zusammenfassung der in dieser Arbeit berechneten Metaanalysen

Zusammenfassend lässt sich feststellen, dass alle sechs im Rahmen dieser Arbeit metaanalysierten CD33-Polymorphismen mindestens in einer der durchgeföhrten Subanalysen eine mindestens nominal signifikante Assoziation mit der AD zeigen (s. Tabelle 3). Die teilweise sehr ähnlichen Metaanalysen-Ergebnisse erklären sich durch die Tatsache, dass ein Teil der metaanalysierten SNPs in den hier betrachteten Abstammungsgruppen miteinander korrelieren (Tabelle 4).

Die Tabelle 4 stellt die Korrelation der oben beschriebenen SNPs untereinander in den nordeuropäischen und ostasiatischen Populationen mittels LD-Berechnung ( $r^2$ -Metrik) dar. Unter LD versteht man die nicht zufällige Assoziation von Allelen, die mit hoher Wahrscheinlichkeit gemeinsam vererbt werden.

	rs3826656	rs1697553	rs3865444	rs12459419	rs2455069	rs35112940
<b>rs3826656</b>		0.206	0.111	0.111	0.206	0.058
<b>rs1697553</b>	0.246		0.405	0.405	<b>1.0</b>	0.27
<b>rs3865444</b>	0.475	0.025		<b>1.0</b>	0.405	<b>0.626</b>
<b>rs12459419</b>	0.454	0.026	<b>0.968</b>		0.405	<b>0.626</b>
<b>rs2455069</b>	0.246	<b>1.0</b>	0.025	0.026		0.27
<b>rs35112940</b>	n.a	n.a	n.a	n.a	n.a	

Tabelle 4: LD-Statistik der sechs behandelten AD-Risiko-SNPs der Metaanalysen aus Kapitel 4.1

Legende: oberhalb der Diagonale und hellblau – für die nordeuropäische Population, unterhalb der Diagonale und dunkelblau – für die ostasiatische Population. „n.a.“ Keine Daten verfügbar.

Die LD-Matrix zeigt, dass die *CD33*-Polymorphismen rs3865444 und rs12459419 eine komplette Assoziation in der nordeuropäischen Population ( $r^2 = 1$ ) und eine sehr starke Assoziation in der ostasiatischen Population ( $r^2 = 0.968$ ) aufweisen. Das heisst, dass die jeweiligen Allele und Genotypen nahezu immer gemeinsam vererbt werden und sich daher die Assoziationsergebnisse entsprechen (müssen).

Des Weiteren befinden sich die beiden SNPs in einem starken LD mit dem SNP rs35112940 in der Gruppe der nordeuropäischen Abstammung ( $r^2 = 0.626$ ). Dementsprechend wurden auch für alle drei Polymorphismen ein signifikant protektiver Effekt auf das AD-Risiko (OR = 0.99) festgestellt (in der nordeuropäischen Population des SNPs rs3865444 und rs35112940, und in der kombinierten Analyse der nordeuropäischen und ostasiatischen Abstammungsgruppen des SNPs rs12459419).

Der SNP rs3826656 befindet sich in den Datensätzen mit ostasiatischer Abstammung nur in einem nur mittelgradig ausgeprägten LD mit dem SNP rs3865444 ( $r^2 = 0.475$ ) und mit dem SNP rs12459419 ( $r^2 = 0.454$ ). In der nordeuropäischen Population besteht zwischen diesen Varianten allerdings kein nennenswertes LD ( $r^2 = 0.111$ ). Die Ergebnisse der Metaanalysen zeigen, dass der SNP rs3826656 nur in der ostasiatischen Population signifikant mit einem erhöhten AD-Risiko verbunden ist (OR = 1.25). Lediglich in der ostasiatischen Kohorte wurde bei dem SNP rs3865444 eine ähnliche Effektrichtung festgestellt (OR = 1.01). Der Unterschied in den Effektstärken könnte am nur mittelgradig ausgeprägten LD ( $r^2 = 0.475$ ) und/oder an den unterschiedlichen Fallzahlen (3180 vs. 1008) in den Metaanalysen liegen.

Die SNPs rs2455069 und rs1697553 befinden sich im vollständigen Kopplungsungleichgewicht in der nordeuropäischen und ostasiatischen Population ( $r^2 = 1$ ). Diese zwei Polymorphismen wurden daher in den durchgeführten Metaanalysen als signifikante AD-Riskoloci mit dem gleichen Ergebnis (OR = 1.01) identifiziert.

#### 4.1.10 Vergleich mit anderen Metaanalysen

In Tabelle 5 ist eine Zusammenfassung der Ergebnisse zur Assoziation von CD33-Polymorphismen und dem AD-Risiko aus der vorliegenden Arbeit, sowie aus den früher publizierten Metaanalysen dargestellt [7, 50, 60, 66, 72, 102]. Die Studien wurden anhand der Stichprobengröße, Anzahl der in den Metaanalysen eingeschlossenen Publikationen / Datensätze und der Metaanalyse-Ergebnissen (OR, 95%-CI und P-Wert) verglichen. Die publizierten Metaanalysen haben sich hauptsächlich mit den CD33-Polymorphismen rs3865444 und rs3826656 beschäftigt. Auf die Qualität der hier dargestellten Publikationen wird noch einmal detailliert in der Diskussion eingegangen. Generell ist anzumerken, dass die Metaanalysen dieser Dissertation in allen Fällen die größeren Datensätze umfassen und somit (derzeit) als aktuellste und präziseste Schätzung der Effektstärken der analysierten Polymorphismen aufgefasst werden kann.

Fasst man die Ergebnisse der Metaanalysen zusammen, zeigt sich, dass die Autoren (inklusive der Autorin der vorliegenden Dissertation) zu dem Schluss kommen, dass das Minor-(A)-Allel des SNPs rs3865444 einen signifikant protektiven Effekt für LOAD zeigt, wenn man alle Populationen gemeinsam analysiert. Das gleiche Resultat findet man in den Subanalysen der nordeuropäischen Population. Der direkte Vergleich der Ergebnisse der Metaanalysen in der Gruppe der ostasiatischen Abstammung demonstriert allerdings widersprüchliche Ergebnisse. Diese Arbeit, sowie die von Li et al. [60] und Mao et al. [66] identifizieren das Minor-(A)-Allel des SNPs rs3865444 als LOAD-Riskoloci, wenn man nur Stichproben mit ostasiatischer Herkunft betrachtet (OR-Variation von 1.01 bis 1.22). Die Analysen von Bao et al. [7] und Jiang et al. [50] zeigten aber einen konträren Effekt (OR = 0.86 und 0.87 entsprechend). Das statistische Signifikanzniveau variierte in den gemeinsamen Metaanalysen aller Populationen des SNPs rs3865444 dabei von P-Wert=0.003 zu P-Wert=0.95 und erfüllte in keinem Fall das Kriterium der genomweiten Signifikanz (P-Wert < 5E-08). Ähnliche Ergebnisse findet man in den Subanalysen der ostasiatischen Population mit P-Wert-Variation von 0.276 bis 0.93. Von den Subanalysen der nordeuropäischen Population haben zwei - die vorliegende Arbeit und die Metaanalyse von Li et al. [60] – ein (mindestens nominales) statistisches Signifikanzniveau erreicht (s. Tabelle 5).

Betrachtet man die Ergebnisse der Metaanalysen zur Assoziation vom SNP rs3826656 und AD-Risiko in der gemischten Stichprobe, lässt sich feststellen, dass auch hier eine Unstimmigkeit vorliegt. Die Resultate der vorliegenden Arbeit legen nahe, dass das G-Allel des CD33-Polymorphismus rs3826656 mit einem erhöhtem LOAD-Risiko verbunden ist (OR = 1.11; P-Wert = 0.20), wogegen die Studie von Jiang et al. [50] einen signifikant

protektiven Effekt dieses Allels aufwies ( $OR = 0.94$ ;  $P$ -Wert < 0.01). Zwei weitere Metaanalysen (die vorliegende Arbeit und die Studie von Mao et al. [66]) zeigen, dass das Major-(G)-Allel mit dem SNP rs3826656 in der ostasiatischen Population statistisch signifikant das AD-Risiko erhöht ( $OR = 1.25$  und  $1.39$ , respektive). Der Grund für die Unstimmigkeit könnte zum Teil an der unterschiedlichen Anzahl der in die Analysen eingeschlossenen Primärstudien und Stichprobengrößen, sowie an Fehlern in der Datenextraktion und -analyse liegen. Eine qualitative Beurteilung jeder der hier erwähnten Metaanalysen ist ausführlich in der Diskussion dargestellt.

Autor, Publikationsjahr	N Studien		N Probanden	Metaanalysen-Ergebnisse		
	Datensätze	Publikationen		OR	95%-CI	P-Wert
<b>rs3865444 A vs. C - All</b>						
Current Study	22	21	476187	0.92	0.77 - 1.10	0.36
Li, 2015	27	8	86759	0.97	0.92 - 1.02	0.264
Bao, 2016	33 <sup>a</sup>	11	123347	0.94	0.90 - 0.98	0.003
Jiang, 2018	38 <sup>a</sup>	16	127435	0.93	0.90 - 0.97	< 0.01 <sup>b</sup>
Moreno-Grau, 2018	29	14	90913	0.98	0.93 - 1.04	0.472
Siokas, 2019 <sup>c</sup>	13	8	11996	0.99	0.84 - 1.18	0.95
<b>rs3865444 A vs. C - East Asian</b>						
Current Study	8	8	6768	1.01	0.81 - 1.27	0.93
Li, 2015	4	4	4186	1.22	0.86 - 1.73	0.276
Mao, 2015	5	5	4441	1.10	0.79 - 1.52	0.576
Bao, 2016	4	4	4186	0.86	0.58 - 1.27	n.a
Jiang, 2018	6	6	4988	0.87	0.65 - 1.17	< 0.01 <sup>b</sup>
<b>rs3865444 A vs. C - European</b>						
Current Study <sup>d</sup>	10	9	467369	0.99	0.98 - 0.99	3.41e-09
Li, 2015	17	2	74998	0.93	0.91 - 0.95	1.73e-09
Bao, 2016	29 <sup>a</sup>	7	119161	0.93	0.90 - 0.95	n.a
Jiang, 2018 <sup>d</sup>	31 <sup>a</sup>	9	122447	0.91	0.89 - 0.93	0.04 <sup>b</sup>
<b>rs3826656 G vs. A - All</b>						
Current Study	6	6	457993	1.11	0.95 - 1.29	0.20
Jiang, 2018	4	4	1810	0.94	0.62 - 1.41	< 0.01 <sup>b</sup>
<b>rs3826656 G vs. A - East Asian</b>						
Current Study <sup>d</sup>	4	4	2098	1.25	1.10 - 1.42	8.20e-04
Mao, 2015	2	2	626	1.39	1.09 - 1.76	0.008

<sup>a</sup> Überlappende Datensätze

<sup>b</sup> am ehesten P-Wert der Heterogenitätsanalysen

<sup>c</sup> Dominantes Modell

<sup>d</sup> Fixed effect Modell

Tabelle 5: Vergleich der Ergebnisse dieser Arbeit mit publizierten Metaanalysen

## 4.2 Genetische Determinanten der *CD33*-Expression im menschlichen Gehirn

Der zweite Analyseschwerpunkt dieser Arbeit befasste sich mit der Frage, ob und welche DNA-Sequenzvarianten eine signifikante Assoziation mit der Expression des *CD33*-Gens im menschlichen Gehirn zeigen. Diese Frage ist u. a. für eine initiale funktionelle Einordnung der in Abschnitt 4.1. beschriebenen Metaanalyseergebnisse der genetischen Assoziationsstudien interessant: Zeigen die bisher im Fokus der Literatur stehenden und oben metaanalysierten *CD33*-SNPs auch eine Assoziation mit der Expression des *CD33*-Gens? Lassen sich hierdurch evtl. die genetischen Assoziationsergebnisse besser einordnen und / oder priorisieren? Zur Beantwortung dieser Frage hatte ich Zugang zu in der Arbeitsgruppe im Rahmen separater Projekte generierten RNA-Sequenzierungsdaten sowie genomweiten SNP-Genotypisierungsdaten, die ich im Kontext einer eQTL-GWAS zusammengeführt habe. In der nachfolgenden Ergebnisdarstellung konzentriere ich mich zuerst auf die bekannten, in Abschnitt 4.1. beschriebenen *CD33*-SNPs, d.h. mögliche *cis* eQTL-Effekte. Da mir für diese Analysen aber genomweite SNP-Genotypisierungsdaten zur Verfügung standen, konnte ich die Suche nach möglichen eQTLs der *CD33*-Expression auf das gesamte Genom ausweiten. Die Darstellung der so generierten *trans* eQTL-GWAS-Ergebnisse erfolgt im zweiten Teil dieses Abschnitts.

### 4.2.1 Ergebnisse der *cis* eQTL-Analysen zur *CD33*-Genexpression

Die folgende Tabelle 6 gibt einen Überblick über die *cis* eQTL-Effekte von den in dem Abschnitt 4.1 beschriebenen sechs *CD33*-Polymorphismen. Keiner dieser SNPs zeigte eine signifikante Assoziation mit der Expression des *CD33*-Gens im Gehirn. Die SNPs rs1697553 und rs2455069 zeigten allerdings Ergebnisse, die das nominale statistische Signifikanzniveau nur leicht überschritten haben (P-Werte von 0,0515 und 0,0534, respektive). Insbesondere der in der Literatur am häufigsten mit AD in Verbindung gebrachte SNP rs3865444 zeigte aber keine Assoziation mit der *CD33*-Expression im Gehirn (P-Wert = 0,589).

SNP	Genort	Effekt vs. Ref. Allel	EAF	P-Wert
<b>rs3865444</b>	19:51727962	A vs. C	0.302017	5.89e-01
<b>rs3826656</b>	19:51726613	G vs. A	0.781886	8.07e-02
<b>rs1697553</b>	19:51727322	G vs. A	0.459989	5.15e-02
<b>rs2455069</b>	19:51728641	G vs. A	0.459981	5.34e-02
<b>rs12459419</b>	19:51728477	T vs. C	0.302161	5.92e-01
<b>rs35112940</b>	19:51738917	A vs. G	0.184041	4.47e-01

Tabelle 6: *Cis* eQTL-Statistiken der sechs behandelten AD-Risiko-SNPs der Metaanalysen aus Abschnitt 4.1

In meiner eQTL-GWAS-Analyse habe ich aber auch noch weitere SNPs in der *CD33*-Region auf Assoziation mit der Genexpression testen können. In Tabelle 7 sind die Top-10 (nach P-Wert) *cis* eQTL-Effekte aus diesen Analysen zusammengefasst. Diese zehn SNPs gehören der generellen *CD33*-Region auf Chromosom 19, +/-500kb (von 51,228,335 bis 52,243,274) an. Insgesamt wurden in diesem Abschnitt 3442 SNPs auf Assoziation getestet, von denen 382 unabhängig (d.h. nicht in LD; hierfür wurde ein Grenzwert von  $r^2=0.3$  genutzt) waren. Die Resultate dieser Analysen zeigten zwar zahlreiche nominal signifikante Befunde, allerdings keines mit genomweiter Signifikanz. Obwohl die Top-10 Polymorphismen das nominale Signifikanzniveau z. T. deutlich unterschritten (die P-Werte variieren von 0.0005 [rs75310023] bis 0.0037 [rs77742698]), ergab die Gen-weite Bonferroni-Korrektur (d.h. Adjustierung für alle 382 unabhängigen SNPs in dieser Region,  $\alpha = 0.05/382 = 1.31e-04$ ), dass keiner dieser SNPs den Schwellenwert der Gen-weiten Signifikanz erreichen. D.h. dass selbst die signifikanteste der hier beobachteten *cis* eQTL-Assoziation noch im Bereich des zufällig Erwartbaren (=Nullhypothese) liegt. Keiner dieser zehn Top *cis* eQTL-Polymorphismen zeigt eine Überlappung mit AD-Risiko-SNPs. Im Gegenteil: im Vergleich zu den sechs in Kapitel 4.1 beschriebenen SNPs, die sich allesamt im 5' Abschnitt des *CD33*-Gens befinden, liegen diese nominal assoziierten Varianten eher in der Nähe des 3'-Endes. Im Anhang 9 sind die Resultate der *cis* eQTL-Analysen in der „*cis*“ *CD33*-Region zusammengefasst.

SNP	Genort	Effekt vs. Ref. Allel	EAF	P-Wert
<b>rs75310023</b>	19:52195270	T vs. C	0.0628541	5.21e-04
<b>rs7251465</b>	19:51350791	G vs. A	0.0832032	1.00e-03
<b>rs78177998</b>	19:51345264	C vs. T	0.0834825	1.63e-03
<b>rs56397626</b>	19:51346522	G vs. A	0.0831101	1.78e-03
<b>rs146018038</b>	19:52203444	A vs. T	0.0359383	2.97e-03
<b>rs78652800</b>	19:52203223	C vs. A	0.0362542	3.07e-03
<b>rs74552391</b>	19:52197292	G vs. A	0.046586	3.63e-03
<b>rs77133866</b>	19:52197489	G vs. A	0.046586	3.65e-03
<b>rs35018336</b>	19:52197457	C vs. T	0.046681	3.66e-03
<b>rs77742698</b>	19:52197496	T vs. C	0.0466169	3.70e-03

Tabelle 7: Teststatistiken der Top-10 *cis* eQTL-SNPs der *CD33*-Genexpression im Gehirn

Interessanterweise zeigte eine *in silico* Modellierung der potentiellen funktionellen Effekte (mittels des „variant effect predictor“ Tools der Ensembl-Datenbank [34]), dass viele dieser Top-10 *CD33 cis* eQTL-Varianten in Bereichen lokalisiert sind, die möglicherweise einen genregulatorischen Effekt haben (z. B. Kolokalisierung mit long non-coding RNAs oder micro RNAs; s. Anhang 10). Trotz der fehlenden Gen-weiten Signifikanz der eQTL-

Assoziationen, erscheint ein funktioneller Zusammenhang zwischen diesen Varianten und der CD33-Genexpression zumindest nicht unplausibel. Ein kausaler funktioneller Zusammenhang kann allerdings nur in dafür geeigneten laborexperimentellen Ansätzen nachgewiesen werden.

#### 4.2.2 Ergebnisse der *trans* eQTL-GWAS-Analysen zur CD33-Genexpression

Wegen des Vorliegens von genomweiten SNP-Genotypisierungsdaten konnte ich in dieser Arbeit auch nach möglichen *trans* eQTL-Effekten suchen. Als „*trans*“ wurden hier alle Varianten gezählt, die nicht in der unmittelbaren CD33-Region lokalisiert waren („*cis*“, s. Abschnitt 4.2.1). Nach ausführlicher QC (s. Methoden) blieben für die genomweiten eQTL-Analysen 7.417.011 SNPs übrig. In Abbildung 19 ist der dazugehörige sog. Manhattan-Plot der SNP-basierten GWAS-Analysen dargestellt. Die rote Linie stellt die weithin im Genomanalyse-Feld verwendete Schwelle für genomweite Signifikanz dar ( $P$ -Wert  $<5\text{e-}08$ ). Diese Schwelle wurde von insgesamt drei SNPs auf den Chromosomen 5, 6, und 8 unterschritten (Tabelle 8). Die MAF für alle drei genomweit assoziierten SNPs war mit Werten zwischen 1 und 5% allerdings vergleichsweise niedrig, so dass ein falsch-positiver Befund aufgrund der kleinen Stichprobengröße nicht ausgeschlossen werden kann. Darüber hinaus wurden zusätzlich zu den drei SNPs weitere 90 Loci gefunden, die mindestens genomweit suggestive Assoziationsevidenz ( $P$ -Wert  $< 1\text{e-}05$ ; Anhang 8) zeigten. Der in Abbildung 20 dargestellte Quantil-Quantil-(QQ)-Plot zeigte keine nennenswerte Evidenz für eine Inflation der genomweiten Teststatistiken. Im Nachfolgenden konzentriere ich mich v. a. auf eine Darstellung der drei in diesen Analysen identifizierten genomweit-signifikanten *trans* eQTL-GWAS-Signale. Im letzten Abschnitt werden auch noch kurz zwei Beispiele der Gen basierten GWAS-Ergebnisse dargestellt, von denen allerdings keines die Schwelle der genomweiten Signifikanz unterschritt.

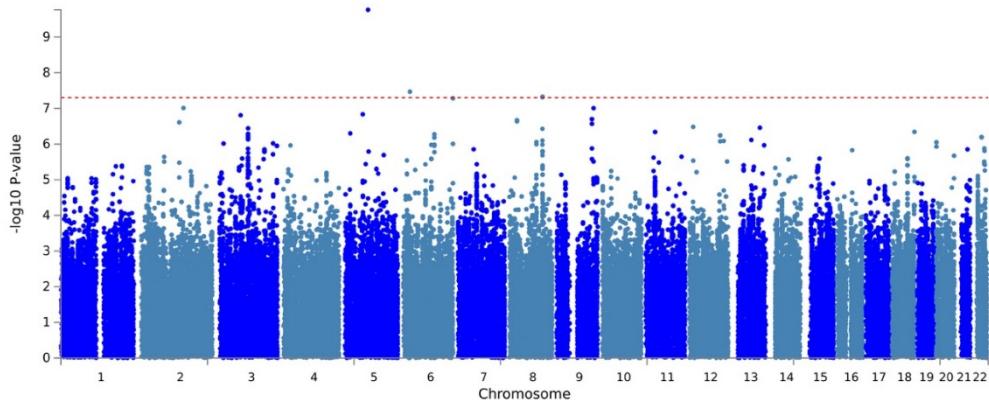


Abbildung 19: SNP-basierter Manhattan-Plot der eQTL-GWAS-Resultate zur CD33-Genexpression

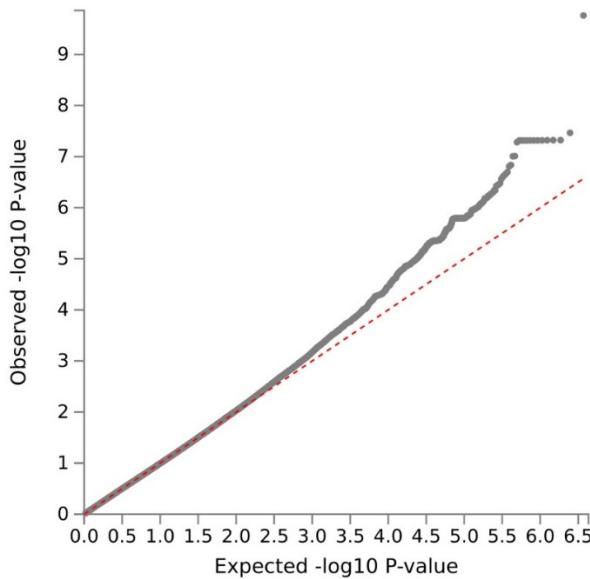


Abbildung 20: Quantile-quantile-(QQ-)Plot von GWAS Übersichtsstatistik

SNP	Genort	Nächstgelegenes Gen	Effekt vs. Ref. Allel	MAF	P-Wert
<b>rs112851200</b>	<b>5:79283742</b>	<i>MTX3</i>	A vs. G	0.05964	1.72742e-10
<b>rs55934141</b>	<b>6:23536354</b>	RP4-810F7.1	A vs. C	0.008946	3.42703e-08
<b>rs2353260</b>	<b>8:113119752</b>	RNU4-37P / <i>CSMD3</i>	A vs. G	0.01988	4.75634e-08

Tabelle 8: Top SNPs der trans eQTL-GWAS-Analysen zur CD33-Genexpression

#### SNP rs112851200 auf Chromosom 5q14.1

SNP rs112851200 ist in einem nicht-kodierenden, intronischen Genabschnitt des *MTX3*-Gens lokalisiert. Mit einem P-Wert von 1.72742e-10 zeigte dieser SNP die signifikanteste Assoziation mit der Expression vom *CD33*-Gen im Gehirn in den hier durchgeföhrten Analysen. Die GTEx-Datenbank (v8) zeigt weitere signifikante eQTL-Effekte in Bezug auf die Expression von *MTX3* in der Schilddrüse und in kultivierten Fibroblastenzelllinien [16], aber keine in den von GTEx analysierten Gehirnproben.

Das *MTX3*-Gen kodiert das Protein Metaxin-3 und ist laut GTEx-Datenbank (v8) im Menschen ubiquitär exprimiert. Die genaue Funktion von Metataxin-3 ist noch nicht abschließend geklärt, laut der Datenbank UniProt wird aber eine Involvierung bei dem Transport von Proteinen in die Mitochondrien vermutet [107]. Im „GWAS-Catalog“, einer Datenbank zu publizierten GWAS an menschlichen Phänotypen, sind bisher keine SNPs in *MTX3* als GWAS-Signale aufgeführt. Obwohl in den Literaturdatenbanken bisher keine Informationen über die Assoziation des SNP rs112851200 oder anderen SNPs im *MTX3*-Gens mit der menschlichen Immunantwort oder mit AD-relevanten Phänotypen aufgeführt werden, ist erwähnenswert, dass in unmittelbarer Nähe (~50kb) das Gen *THBS4*

(Thrombospondin-4) lokalisiert ist, das erst kürzlich als Teil der astrozytischen Gensignatur bei chronischer Neurodegeneration identifiziert wurde [27]. Ob und wie dieser Befund im Verhältnis zur *CD33*-vermittelten Immunantwort steht ist aber unklar.

#### SNP rs55934141 auf Chromosom 6p22.3

Dieser SNP ist laut UCSC-Genombrowser in keinem kodierenden Element lokalisiert. Das mit ca. 138 kbp nächstgelegene annotierte Gen ist RP4-810F7.1 (ENSG00000219453), das laut Ensembl-Datenbank ein Ferritin-Pseudogen ohne bekannte Funktion ist. Mit einem P-Wert von 3.42703e-08 zeigte SNP rs55934141 eine genomweit signifikante Assoziation mit der Expression des *CD33*-Gens im Gehirn. Die GTEx-Datenbank zeigte eine generell nur sehr schwache Expression von RP4-810F7.1 im Menschen und keine weiteren eQTL-Effekte für den SNP rs55934141 auf GTEx reportiert.

Auch im GWAS-Katalog wurde bisher keine Daten über die Assoziation dieses Gens mit der menschlichen Immunantwort oder der AD aufgeführt. Die erweiterte Suche nach Einträgen in der unmittelbaren Nähe des assoziierten SNPs (Region chr6:23286454-23786454) ergab allerdings mehrere GWAS-Befunde mit Bezug zum Immunsystem, z. B. der Lymphozyten- und Monozyten-Zahl im Blut sowie der Immunantwort im Zuge einer Infektion mit Pockenviren (Variola) [75]. Eine funktionelle Verbindung zur *CD33*-vermittelten Immunantwort erscheint somit zumindest möglich, wenngleich es derzeit keine Studien gibt, die dies direkt belegen.

#### SNP rs2353260 auf Chromosom 8q23.3

Auch der dritte identifizierte SNP rs2353260 liegt laut UCSC-Genombrowser nicht in einem bekannten kodierenden Abschnitt des menschlichen Genoms. Dieser SNP zeigte eine gerade noch genomweit signifikante Assoziation (P-Wert von 4.75634e-08) mit der Expression des *CD33*-Gens im Gehirn. Die nächstgelegenen Gene sind RNU4-37P (ENSG00000222146; ebenfalls als Pseudogen annotiert; Abstand ca. 42kb) und *CSMD3* (CUB and sushi domain-containing protein 3; Abstand ca. 120kb).

Letzteres kodiert für ein Protein aus der Familie der CSMD-Proteine, die Reaktionen zwischen extrazellulären Proteinen und Zellmembranproteinen vermittelt, sowie die Synapsenbildung im ZNS reguliert [69, 99]. Dazu passend ist die prädominante Expression dieses Gens im menschlichen ZNS-Gewebe in GTEx (v8) [17]. Mizukami et al. berichteten im Jahr 2016, dass das *CSMD3*-Gen an der Verzweigung der Dendriten der Hippocampus-Neuronen beteiligt ist [69]. Im Jahr 2003 wurde *CSMD3* von Shimizu et al. als ein Gen-Kandidat für familiäre myoklonische Epilepsie postuliert [98]. Außerdem wurden Mutationen des *CSMD3*-Gens auch bei Patienten mit Autismus und Schizophrenie beschrieben [69].

Basierend auf diesen Erkenntnissen kann konkludiert werden, dass die Dysfunktion des *CSMD3*-Gens zu Anomalien der dendritischen Entwicklung und dadurch möglicherweise zu neurologischen und psychiatrischen Störungen führen kann. Dazu passen die Einträge im GWAS-Katalog, der Varianten in diesem Gen mit mehreren ZNS-relevanten Funktionen aufführt, z. B. dem Tourette-Syndrom und Insomnie [76]. Allerdings werden dort derzeit keine Assoziationen mit Immunantwort-relevanten Phänotypen beschrieben, so dass der Zusammenhang mit der *CD33*-Funktion unklar bleibt.

#### 4.2.3 Gen-basierte eQTL-GWAS Ergebnisse zur *CD33*-Genexpression im menschlichen Gehirn

In den Gen-basierten GWAS-Analysen werden die SNP-basierten GWAS-Ergebnisse mit einem im FUMA-Tool integrierten Algorithmus (MAGMA; [57]) aggregiert. Dies erlaubt eine übergeordnete Einschätzung der Assoziationsevidenz auf Gen-Ebene. Durch die kleinere Anzahl an annotierten Genen reduziert sich damit die genomweite Signifikanzschwelle auf  $\alpha = 2.684\text{e-}6$ . Diese Schwelle ist im Manhattan-Plot in Abbildung 21 wieder durch die rote Linie markiert. Wie ersichtlich, wurde dieser Schwellenwert aber von keinem der eingeschlossenen Gene unterschritten, d.h. dass in der Gen-weiten GWAS i. Ggs. zu den SNP-basierten Analysen (s.o.) keine genomweiten Signale identifiziert wurden. Der dazugehörige QQ-Plot (Abbildung 22) zeigt keine Evidenz für eine Inflation der genomweiten Gen-basierten Teststatistiken.

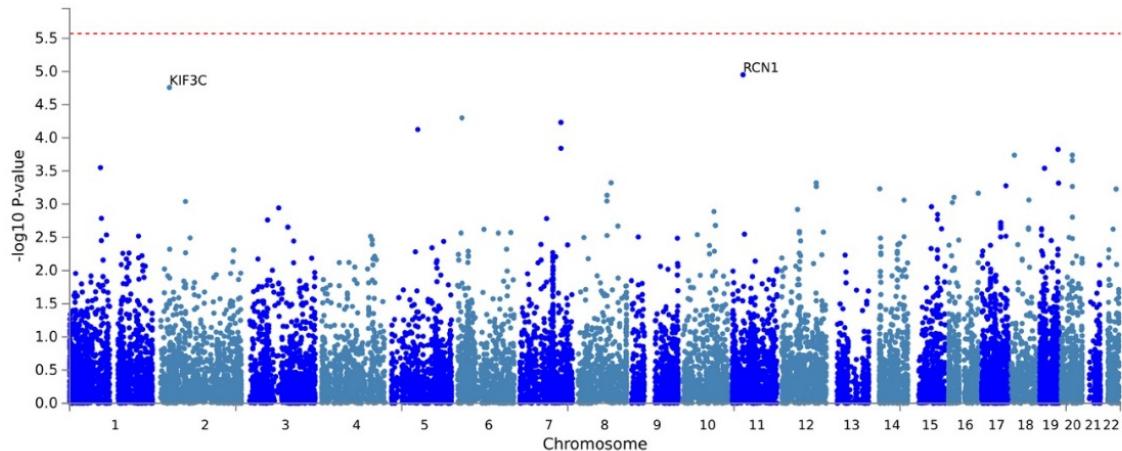


Abbildung 21: Gen-basierter Manhattan-Plot

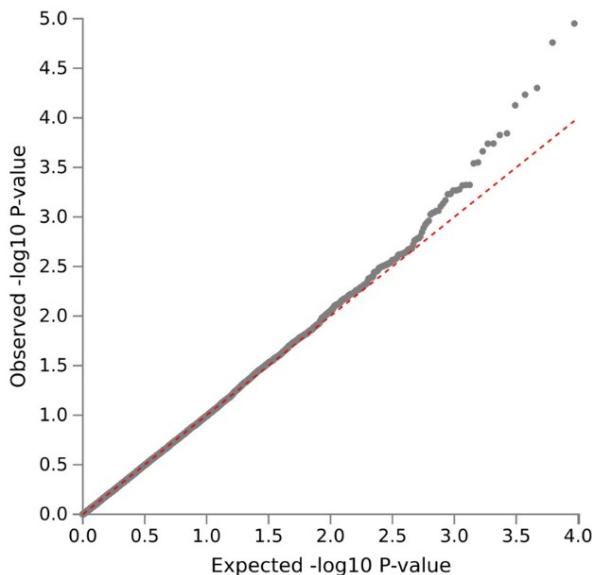


Abbildung 22: Quantile-quantile-(QQ)-Plot des genbasierten Tests

Trotz der fehlenden Evidenz für eine genomweit signifikante Assoziation mit der CD33-Genexpression sollen hier die zwei signifikantesten Ergebnisse der Gen-basierten GWAS-Analysen kurz zusammengefasst werden. Die stärkste Assoziation (P-Wert 1.1188e-05) wurde mit dem *RCN1*-Gen (Reticulocalbin-1) auf Chromosom 11p13 identifiziert. Reticulocalbin-1 ist ein kalzium-bindendes Protein, von dem vermutet wird, dass es vor allem die kalzium-abhängige Zelladhäsion beeinflusst [82]. Die Dysregulation des *RCN1*-Proteins wurde bisher bei einer Vielzahl von Krankheiten festgestellt, u.a. bei onkologischen, kardiovaskulären und neuromuskulären Erkrankungen [21]. Eine wachsende Anzahl von Studien hat die Überexpression von *RCN1* bei Mammakarzinomen, Leberzellkarzinomen, Nierenzellkarzinomen, Kolonkarzinomen und anderen Tumoren gefunden. Basierend auf diesen Ergebnissen wird *RCN1* als Onkogen bei der Tumorentstehung und Tumorprogression eingeordnet [21]. Einen Zusammenhang mit der menschlichen Immunantwort ist in den im GWAS-Katalog aufgeführten Studien zum Zeitpunkt der Schriftlegung dieser Arbeit allerdings nicht evident [77].

Das in den Gen-basierten GWAS-Analysen am zweitstärksten assoziierte Gen (P-Wert 1.7377e-05) ist *KIF3C* (engl. „kinesin superfamily protein 3C“), das auf Chromosom 2p23.3 lokalisiert ist. Das *KIF3C*-Protein gehört zur Kinesin-Superfamilie, die eine Klasse der Mikrotubulus-abhängigen Motorproteine darstellt [36]. Diese Proteine wandeln die ATP-Energie in eine mechanische Energie um. Somit steuern die den intrazellulären Transport von Makromolekülen und Organellen. Im Gegensatz zu *RCN1* wird *KIF3C* im ZNS besonders stark exprimiert [18]. In separaten Analysen wurde berichtet, dass *KIF3C* im ZNS nicht nur in Neuronen, sondern auch in Astrozyten exprimiert wird [36]. Des Weiteren gibt

es Hinweise, dass das KIF3C-Protein am Wachstum und der Regeneration von Axonen nach Verletzung beteiligt ist [42]. Gao et al. [36] fanden im Jahr 2020 heraus, dass die Überexpression des *KIF3C*-Gens in den Gliazellen deren Zellproliferation, Migration und Invasion fördert und die Zellapoptose unterdrückt. Interessanterweise gibt es im GWAS-Katalog derzeit keine klaren Evidenzen für eine Involvierung mit Gehirnfunktion-relevanten Phänotypen, wohl aber mit einigen möglicherweise Immunsystem-relevanten Ausprägungen, wie z. B. der Leukozytenanzahl im Blut [78]. Ob und wie dieser Befund in Verbindung mit der *CD33*-vermittelten Immunantwort steht bleibt aber zur Schriftlegung unklar.

#### 4.2.4. Zusammenfassung der eQTL-GWAS-Befunde in Bezug auf die *CD33*-Genexpression im Gehirn

Zusammenfassend lässt sich feststellen, dass in den von mir durchgeführten eQTL-GWAS-Analysen zur *CD33*-Expression im Gehirn einige interessante Assoziationssignale beobachtet wurden, die möglicherweise mit der *CD33*-Funktion in Verbindung stehen. Diese wurden v. a. als *trans*-Effekte evident, d.h. mit Genloci, die allesamt in größerer Entfernung (>500kb) zum *CD33*-Gen selbst lokalisiert waren. Genloci in *cis*, d.h. die in unmittelbarer Umgebung des *CD33*-Gens lokalisiert sind, zeigten dagegen keine (statistisch) besonders herausragenden Effekte basierend auf den generierten GWAS P-Werten. Der kleinste *cis* eQTL-Effekt wurde mit SNP rs75310023 beobachtet (P-Wert=5.21e-04), war aber nach Bonferroni-Korrektur nicht-signifikant auf Gen-Ebene. Insbesondere von den in Abschnitt 4.1. ausführlich analysierten AD-Risiko-SNPs zeigte keiner eine nennenswerte Assoziation mit der Expression des *CD33*-Gens in diesem Datensatz. Dies deutet entweder darauf hin, dass solche Effekte mit den AD-Risiko-SNPs nicht existieren (und damit der Mechanismus der beobachteten genetischen Assoziation nicht primär über die Genexpression vermittelt wird) oder dass die statistische Trennschärfe (Power) solche Effekte zu identifizieren in dem von mir verwendeten Gehirn-Datensatz nicht ausreichend war (s. auch Diskussion).

## 5 Diskussion

### 5.1 Diskussion der wissenschaftlichen Relevanz der vorliegenden Dissertation

*Allgemeine Relevanz.* Die Aktualität und die wissenschaftliche Relevanz des Themas der vorliegenden Dissertation ergibt sich aus der hohen Prävalenz der AD. Laut WHO leben derzeit mehr als 55 Millionen Menschen mit Demenz weltweit [114]. Da die Bevölkerung in fast allen Ländern älter werden wird, wird erwartet, dass die Prävalenz der AD (die ca. 70% aller Demenz-Fälle ausmacht) im Jahr 2030 auf 78 Millionen und im Jahr 2050 auf 139 Millionen ansteigen wird [114]. Basierend auf diesen hohen und steigenden Erkrankungszahlen sind auch die damit assoziierten Krankheitskosten enorm. Allein im Jahr 2015 wurden sie global auf 818 Milliarden US-Dollar geschätzt [63]. Im Zuge der vorausgesagten Prävalenz-Erhöhung werden auch diese Krankheitskosten (85% hiervon kommen durch familiäre und soziale Belastungen zustande) weiter dramatisch steigen [63]. Darüber hinaus hat sich seit dem Anfang der COVID-19 Pandemie die AD als eine der wichtigsten Komorbiditäten dieser respiratorischen Krankheit herausgestellt, was wirtschaftliche Kosten und Zeitaufwand für pflegende Angehörige nochmals erhöht hat [116].

*Spezifische Relevanz.* In einer kürzlich in *Nature Genetics* veröffentlichten Analyse zum Erfolg von Arzneimittelstudien kamen die Autoren zu dem Schluss, dass die Auswahl von „Drug-Targets“, die durch genetische Studien supportiert sind, die Erfolgsrate in klinischen Studien verdoppeln könnte [81]. Das Ziel dieser Promotion lag darin, ein besseres Verständnis über die komplexen Zusammenhänge zwischen genetischer Veranlagung und dem AD-Risiko in Bezug auf das CD33-Gen zu gewinnen, das von einigen Autoren bereits als möglicher pharmakologischer Therapieansatz ausgemacht wurde [13, 39].

Um unser Verständnis der Rolle des CD33-Gens in der Pathogenese der AD zu vertiefen habe ich in dieser Arbeit zwei Ansätze verfolgt. Erstens habe ich unter Zusammenführung sämtlicher derzeit verfügbarer publizierter Evidenz untersucht, welche genetischen Polymorphismen des CD33-Gens das AD-Risiko beeinflussen. Hier habe ich sechs Polymorphismen mit mindestens nominal signifikanter Assoziation identifiziert, drei davon mit einem protektiven Effekt, die anderen drei mit Risiko-erhöhenden Effekten. Zweitens habe ich mittels genomweiter Daten analysiert, welche genetischen Polymorphismen eine Assoziation mit der Expression des CD33-Gens im Gehirn von AD Patienten und Kontrollpersonen zeigen. In diesen Analysen habe ich herausgefunden, dass drei SNPs die Schwelle der genomweiten Signifikanz unterschritten. Alle drei assoziierten SNPs waren *trans* Varianten, d.h. sie lagen nicht in unmittelbarer Nähe des CD33-Gens. Im Gegensatz dazu zeigte keiner der sechs in dem ersten Teil der Arbeit metaanalysierten AD-Risiko-

SNPs (alle in *cis*) einen nennenswerten Einfluss auf die *CD33*-Expression in menschlichen Gehirnproben.

Im Nachfolgenden diskutiere ich die Ergebnisse meiner Arbeit im Kontext der aktuellen Literatur und hebe die Stärken und insbesondere die Schwächen der von mir durchgeführten Analysen hervor.

## 5.2 Diskussion der Ergebnisse der Metaanalysen

### 5.2.1 Zusammenfassung der wichtigsten Ergebnisse

In den durchgeführten Metaanalysen wurden die SNPs rs3865444, rs3826656, rs1697553, rs2455068, rs12459419 und rs35112940 des *CD33*-Gens auf eine Assoziation mit dem LOAD-Risiko anhand der Zusammenführung von 24 relevanten Primärstudien, die sich in ihren Ergebnissen zum Teil unterschieden haben, untersucht. Zusammenfassend lässt sich hierzu festhalten, dass zwischen allen untersuchten Polymorphismen in mindestens einer der analysierten Populationen eine statistisch signifikante Assoziation mit dem Auftreten der AD besteht.

Der für den Forschungsdiskurs bisher relevanste SNP rs3865444 und im kleineren Ausmaß der SNP rs3826656 haben, basierend an den Erkenntnissen der vorliegenden Arbeit, den AD-assoziierten Status bekräftigt. Für den SNP rs3865444 wurde in die Metaanalyse die größte Anzahl an Primärstudien eingeschlossen, inklusive der zur Schriftlegung aktuellsten und umfangreichsten AD-GWAS aus dem Jahr 2019 [48]. Das Minor-(A)-Allel des SNPs rs3865444 zeigte sich hierbei in der Gruppe der Probenden mit nordeuropäischer Abstammung als signifikanter AD-protektiver Polymorphismus. In den Analysen für andere Populationen wurden dahingegen keine statistisch signifikanten Assoziationen detektiert. Von besonderem Interesse ist die Beobachtung, dass die Effektrichtung der Metaanalyse des SNPs rs3865444 in der ostasiatischen Population konträr zu der Effektrichtung der Nordeuropäer war. Die LD-Berechnung kann hierfür als Erklärung dienen: die SNPs rs3865444 und rs3826656 befinden sich in der Population der Ostasiaten, i. Ggs. zu den Nordeuropäern, in moderatem LD (Tabelle 4). Genau in dieser Population zeigt sich das Major-(G)-Allel des SNPs rs3826656 in meiner Analyse als signifikanter AD-Risikofaktor.

Weiterhin lässt sich unter Berücksichtigung der LD-Struktur herleiten, dass die SNPs rs12459419 und rs35112940 eine signifikante Assoziation mit einem reduziertem AD-Risiko aufgrund ihres starken LD mit rs3865444 zeigen.

Um möglichst umfangreiche Ergebnisse zu der Rolle des *CD33*-Gens bei der Entstehung der AD zu erhalten, habe ich in dieser Arbeit erstmals Metaanalysen zu zwei weiteren

SNPs, d.h. rs1697553 und rs2455069, durchgeführt. Diese Polymorphismen befinden sich sowohl in Nordeuropäern als auch in Ostasiaten in komplettem LD miteinander. Folglich zeigten die Minor-(G)-Allele beider SNPs als Ergebnis eine signifikante Assoziation mit dem Auftreten der AD.

Als Fazit lässt sich festhalten, dass ich durch die Integration der Ergebnisse mehrerer Primärstudien in die Metaanalysen ein besseres Verständnis über die komplexen Zusammenhänge zwischen *CD33*-Polymorphismen und der AD gewinnen und die Rolle des *CD33*-Gens als wichtigem AD-Risikogen bestätigen konnte.

### 5.2.2 Vergleich der erhobenen Metaanalyse-Ergebnisse mit der Literatur

Bis zum Zeitpunkt der Literatursuche (31.12.2020) wurden sechs Metaanalysen, die eine quantitative Kombination der Ergebnisse mehrerer genetischen Assoziationsstudien darstellen, zu diesem Thema publiziert. Im Kapitel 4.1.10 habe ich einen Überblick über die (z. T. unterschiedlichen) Ergebnisse dieser Studien zusammengestellt. Auch wenn Metaanalysen in der Hierarchie der Evidenz die höchstwertigste Stellung einnehmen [1] ist damit keine pauschale Aussage über den wissenschaftlichen, qualitativen Wert einer individuellen Metaanalyse möglich. So kann eine falsch ausgewertete Metastudie sogar einen geringeren bzw. falschen wissenschaftlichen Wert im Vergleich zu einer Einzelnen, sauber durchgeführten Assoziationsstudie haben. Nachfolgend möchte ich daher eine Einschätzung der Berichtsqualität der bisher veröffentlichten Metaanalysen ausführen.

#### Qualitative Beurteilung der Metaanalyse von Jiang et al. [50]

Die 2018 in *Annals of Translational Medicine* publizierte Metaanalyse von Jiang et al. gehört zu den größten und aktuellsten Übersichtsarbeiten zum Thema. Dennoch weist diese Arbeit in der methodischen Vorgehensweise, sowie auch in der Darstellung der Ergebnisse mehrere Limitationen auf, welche die Aussagekraft m. E. reduzieren. Zuerst erfolgte in der Metaanalyse laut Methodenbeschreibung keine Kontrolle auf und keine Eliminierung von duplizierten Datensätzen. So überlappen die „ADGC“-Datensätze („ADGC-GWAS“ und „ADGC-REP“) von Naj et al. [73] mit dem „ADGC“-Datensatz von Lambert et al. [55]. Ferner sind die Datensätze „GERAD1“ und „EADI1“ von Hollingworth et al. [45] sind dieselben wie die Datensätze „GERAD“ und „EADI“ von Lambert et al. [55]. Der Datensatz „Mayo2“ („Jacksonville“, „Rochester“ und „Autopsy“) von Carrasquillo et al. [20] wurde ebenfalls in den „ADGC“-Datensatz von Lambert et al. [55] zusammengeführt. Keine dieser Überlappungen der verwendeten Studienpopulationen wurde augenscheinlich von Jiang et al. berücksichtigt. D.h., dass von den 127.435 Probanden, die von Jian et al. in die Metaanalyse von SNP rs3865444 eingeschlossenen wurden, 39.750 Teilnehmer (also fast

1/3 der gesamten Stichprobe) doppelt analysiert wurden. Des Weiteren fiel auf, dass die Autoren dieser Arbeit bei manchen eingeschlossenen Studien nicht korrekt ermittelt haben, ob es sich bei dem Effektallel (und den entsprechenden OR-Schätzern) um das Minor- oder das Major-Allel handelt. Somit wurde die Effekt-Richtung in den Fall-Kontroll-Studien von Deng et al. [28] und Jiao et al. [51] für den SNP rs3865444 nicht entsprechend angepasst und daher falsch von Jiang et al. berücksichtigt. Auch in der Metaanalyse zu dem SNP rs3826656 wurde offensichtlich kein Abgleich der Effektallele vorgenommen: die Effektrichtung der Studie von Mao et al. [66] ist mit Effektrichtungen der Studien von Yuan et al. [118] und Jiao et al. [51] in der Metaanalyse von Jiang et al. entgegengesetzt dargestellt, obwohl sie in der Darstellung der Primärstudien in dieselbe Richtung zeigen. Der OR-Wert von einer weiteren Fall-Kontroll-Studien, die in die Metaanalyse von Jiang et al. zu dem Polymorphismus rs3826656 eingeschlossen wurden [71], ist nicht nachvollzierbar, weil er sich aus keinen in der Primärstudie gegebenen Daten errechnen lässt. Zuletzt möchte ich erwähnen, dass die in der Publikation als P-Werte der Metaanalysen dargestellten Werte am wahrscheinlichsten die P-Werte der Heterogenitätsanalysen darstellen. Dies lässt sich dadurch belegen, dass die in den Forrest-Plots dargestellten P-Werte sich auf die Heterogenitätsanalysen beziehen und *dieselben* Werte sich aber in der Beschreibung der Signifikanz wiederholen. Somit sind die für diese SNPs angegebenen Signifikanz-Werte dieser Metaanalysen höchstwahrscheinlich schlicht falsch dargestellt. Die vergleichsweise großen Unterschiede in den Ergebnissen der hier vorgelegten Dissertation und der Publikation von Jiang et al. sind m. E. durch diese Unterschiede (und wahrscheinlich: Fehler in der Jiang et al. Publikation) erklärlich.

#### Qualitative Beurteilung der Metaanalyse von Moreno-Grau et al. [72]

Die Metaanalyse von Moreno-Grau et al. wurde ebenfalls im Jahr 2018 publiziert, im Journal *Oncotarget*. Die Autoren dieser Studie haben sich ausschließlich auf den SNP rs3865444 des CD33-Gens konzentriert und eine eigene Fall-Kontroll-Studie durchgeführt, die später mit den Ergebnissen aus 13 anderen Datensätzen per Metaanalyse kombiniert wurde. Die duplizierten Datensätze aus dem IGAP-Dataset (s.o.) wurden hier richtigerweise aus der Analyse ausgeschlossen. Trotzdem fiel auch diese Arbeit dem Fehler anheim, die Effektrichtung aus der Studie von Jiao et al. [51] nicht anzupassen, obwohl die Allele in der Arbeit offensichtlich auf dem komplementären Strang dargestellt sind. Im Vergleich zu der vorliegenden Dissertation erfolgte diese Metaanalyse nur zu einem der CD33-Polymorphismen (rs3865444) und ausschließlich in der Gesamtpopulation, ohne dass die ethnischen Subgruppenanalysen durchgeführt wurden. Des Weiteren war die

Stichprobengröße der Metaanalyse von Moreno-Grau deutlich kleiner als die der vorliegenden Arbeit (90 913 [72] vs. 476.187 [diese Arbeit]), was durch den fehlenden Einschluss der Daten der Jansen-GWAS [48] bedingt war. Trotz der Unterschiede gab es keine relevanten Unterschiede in den Ergebnissen (Effektrichtung bzw. Signifikanz).

#### [Qualitative Beurteilung der Metaanalyse von Li et al. \[60\]](#)

Li et al. untersuchten in ihrer systematischen Übersichtsarbeit und Metaanalyse, die im Jahr 2015 in der Zeitschrift *Molecular Neurobiology* publiziert wurde, lediglich den Zusammenhang zwischen dem SNP rs3865444 und dem AD-Risiko. Hierbei wurden die Daten für ostasiatische, europäische und nordamerikanische Populationen getrennt analysiert. Keiner der anderen *CD33*-Polymorphismen wurden für dieses Papier berücksichtigt. Nach meiner Durchsicht dieser Arbeit gab es i. Ggs. zu den oben dargestellten keinen Anhalt für methodische Fehler. Im direkten Vergleich mit der vorliegenden Arbeit war die Anzahl der in die Metaanalysen für den SNP rs3865444 eingeschlossen Studien allerdings sehr viel kleiner (8 [60] vs. 21 [hier] in der Gesamtanalyse aller Populationen, 2 [60] vs. 9 [hier] in der europäischen Population und 4 [60] vs. 8 [hier] in der ostasiatischen Population), sowie die der Stichprobengröße (86.756 [60] vs. 476.187 [hier] in der Gesamtanalyse aller Populationen, 74.998 [60] vs. 467.369 [hier] in der europäischen Population und 4.186 [60] vs. 6.768 [hier] in der ostasiatischen Population), was der großen zeitlichen Differenz in der Durchführung beider Studien, 2015 [60] vs. 2020 [diese Arbeit] geschuldet ist. Die Ergebnisse haben sich allerdings nicht in der Effektrichtung oder in der Signifikanz unterschieden.

#### [Qualitative Beurteilung der Metaanalyse von Mao et al. \[66\]](#)

Mao et al. fokussierten sich in deren Arbeit, die im Jahr 2016 in der Zeitschrift *Neuroscience Letters* erschien, ausschließlich auf Datensätze ostasiatischer Herkunft. Sie haben hierfür eine eigene genetische Assoziationsstudie für AD durchgeführt und anschließend die Ergebnisse mit anderen Fall-Kontroll-Studien zu diesem Thema metaanalysiert. Für das *CD33*-Gen wurden im Rahmen dieser Metaanalysen zwei Polymorphismen untersucht (rs3865444 und rs3826656). Die wichtigsten Unterschiede der Analysen der hier vorgelegten Dissertation im Vergleich mit der Metaanalyse von Mao et al. liegen in der Anzahl der eingeschlossenen Studien (5 [66] vs. 8 [hier] für den SNP rs3865444 und 2 [66] vs. 4 [hier] für den SNP rs3826656) und in den Stichprobengrößen (4.441 [66] vs. 6.768 [hier] für den SNP rs3865444 und 626 [66] vs. 2098 [hier] für den SNP rs3826656). Was die Ergebnisse betrifft, gab es keine gravierenden Unterschiede, d.h. sowohl Mao et al. als auch diese Arbeit kamen zu dem Schluss, dass die SNPs rs3865444 und rs3826656 in der

ostasiatischen Population zu den AD-Risikoloci gehören, allerdings zeigte lediglich der SNP rs3826656 eine signifikante Assoziation mit der AD.

#### Qualitative Beurteilung der Metaanalyse von Bao et al. [7]

Die Metaanalyse von Bao et al., die im Jahr 2016 in der Zeitschrift *Medical Science Monitor* publiziert wurde, konzentrierte sich auf den CD33-Polymorphismus rs3865444 und auf zwei andere AD-Risikogene (*ABCA7* und *TOMM40*). Diese Metaanalyse wies in der Art der Durchführung und Darstellung der Ergebnisse beim SNP rs3865444 jedoch deutliche Einschränkungen auf, welche die Aussagekraft der erhobenen Ergebnisse meiner Ansicht nach reduzieren. Ähnlich wie in der oben beschriebenen Metaanalyse von Jiang et al. erfolgte im Rahmen dieser Arbeit keine Kontrolle bzw. Eliminierung von duplizierten Datensätzen (zusammengeführte Datensätze der IGAP-Studie, s. o.). Dadurch wurden 39.750 von 123.347 Probanden doppelt in die Analyse eingeschlossen. Auch die Richtung des Effekts der Studie von Deng et al. [28] wurde nicht angepasst, was zu einer falschen Berücksichtigung dieser Studie in den Metaanalysen führte. Ferner wurden zu den Subanalysen in Populationen ostasiatischer und europäischer Abstammung keine P-Werte angegeben, was es erschwert, eine Aussage zur statistischen Signifikanz der Ergebnisse zu machen. Anhand der publizierten OR-Werten und CI lässt sich aber annehmen, dass der SNP rs3865444 in europäischen Stichproben signifikant (OR = 0.93; CI: 0.90-0.95) und in ostasiatischen Stichproben nicht signifikant (OR = 0.86; CI: 0.58-1.27) mit der AD assoziiert war.

#### Qualitative Beurteilung der Metaanalyse von Siokas et al. [102]

Die neuste Metaanalyse zur Assoziationsevidenz zwischen SNPs des CD33-Gens und dem AD-Risiko wurde von den Autoren Siokas et al. im Jahr 2019 in der Zeitschrift *Journal of Molecular Neuroscience* veröffentlicht. Diese Arbeit hat eine eigene Fall-Kontroll-Studie in einem griechischen Datensatz durchgeführt, die anschließend mit sieben weiteren Primärstudien metaanalysiert wurde. Die Ergebnisse haben keine signifikante Assoziation zwischen LOAD und dem CD33-Polymorphismus rs3865444 in der gemischten Analyse in fünf verschiedenen genetischen Modellen (co-dominant, dominant, recessive, over-dominant, log-additive) nachgewiesen. Erwähnenswert ist, dass die Primärstudienanzahl und Stichprobengröße in Siokas et al. deutlich kleiner als die in der hier vorgelegten Metaanalyse waren, d.h. 8 [102] vs. 21 [hier], sowie 11.996 [102] vs. 476. 187 [hier]. Dies ist v.a. der Tatsache geschuldet, dass Siokas et al. keine der publizierten GWAS eingeschlossen hatten. Der SNP rs3865444 wurde in der Metaanalyse von Siokas et al. in keinem der angewandten Modelle statistisch signifikant mit der AD assoziiert berichtet,

was allerdings den Ergebnissen meiner Arbeit in der gemischten Analyse nicht widerspricht. Weitere Subgruppenanalysen in verschiedenen Abstammungsgruppen wurden in der Metaanalyse von Siokas et al. allerdings nicht verfolgt.

### 5.2.3 Stärken und Schwächen dieses Teils der Arbeit

Aufgrund ihrer Aktualität ist eine der größten Stärken dieses Teils der Arbeit, dass die durchgeführten Metaanalysen jeweils den aktuellsten Forschungstand zur Beurteilung der Assoziation der CD33-Polymorphismen und des LOAD-Risikos darstellt, zumindest bis zum letzten Datum der Literatursuche am 31.12.2020. Im Vergleich zu bisherigen Metaanalysen wurde die größte Anzahl an Primärstudien und somit auch die größte Stichprobenanzahl in den Metaanalysen berücksichtigt, was die Aussagekraft der Ergebnisse erhöht. Des Weiteren wurde in dieser Arbeit bis dato die höchste Anzahl an Polymorphismen des CD33-Gens metaanalyisiert. Schließlich habe ich versucht durch die Anwendung der PRISMA-Checklist die Qualität der Übersichtsarbeit und der darin durchgeführten Metaanalysen in höchsten Maße zu gewährleisten. Z. B. habe ich durch die Darstellung eines an die PRISMA-Flow-Chart angelehnten Flussdiagrammes und die tabellarische Darstellung der Ergebnisse die Arbeit möglichst transparent und nachvollziehbar gemacht. Basierend auf den qualitativen Beurteilungen anderer zu diesem Thema publizierter Metaanalysen stelle ich außerdem fest, dass ich in meiner Arbeit einige offensichtliche Fehler der Vorgängerstudien vermeiden konnte, was die Aussagekraft der Ergebnisse dieser Dissertation weiter erhöht. Selbstverständlich kann ich dennoch nicht ausschließen, dass sich trotz sorgfältigster Vorgehensweise auch in meine Arbeit Fehler eingeschlichen haben (s. nächster Abschnitt).

Trotz der Befolgung gängiger Richtlinien zur Durchführung systematischer Metaanalysen von genetischen Assoziationsstudien gibt es auch in dieser Arbeit einige Schwächen, auf die ich im Folgenden hinweisen möchte. Die erste potenziellen Einschränkung betrifft die verwendete Suchstrategie, die möglicherweise einige in Frage kommende Studien nicht identifiziert hat, z. B. solche, die nicht in der Literaturdatenbank PubMed aufgeführt wurden. Ich habe versucht, diese mögliche Limitation durch Abgleich der Referenzlisten der identifizierten Primärpublikationen zu minimieren. Zweitens, kann ein zumindest teilweise fehlerhaftes Verfahren zur Datenextraktion und -darstellung trotz mehrfacher unabhängiger Kontrolle und Prüfung durch eine zweite und teilweise dritte fachkundige Person nicht komplett ausgeschlossen werden. Drittens erfolgte aufgrund fehlender Daten in einigen eingeschlossenen Studien die Berechnung der Effektstärken anhand der angegebenen MAF-Werte/Allele bzw. die eine Umrechnung der publizierten Daten, um möglichst „reine“ (d.h. nicht APOE-adjustierte) OR-Werte zu erhalten. Dieses Prozedere wurde analog zu

den Vorläuferstudien der Arbeitsgruppe gewählt [9]. Auch wenn sich in den vorgelegten Metaanalysen offensichtlich keine wesentlichen Veränderungen der Effektstärken ergeben haben, könnten unterschiedliche Berechnungsformen zu leicht unterschiedlichen Ergebnissen führen. Viertens habe ich für die Metaanalysen nur das additive Modell berechnet, andere genetische Transmissions-Modelle habe ich im Rahmen dieser Dissertation dagegen nicht berücksichtigt. Obwohl dies die gängige Praxis in genetischen Assoziationsstudien komplex-genetischer Erkrankungen darstellt [9, 62], kann nicht ausgeschlossen werden, dass die Berücksichtigung anderer Transmissionmodelle weitere Assoziationsevidenzen zu Tage gefördert hätte. Fünftens konnten hinsichtlich der ethnischen Abstammung lediglich Aussagen zu den nordeuropäischen und ostasiatischen Populationen getroffen werden. Dies war der Tatsache geschuldet, dass Daten zu anderen Abstammungsgruppen entweder ganz fehlten oder in nicht aussagekräftiger Anzahl vorlagen, so dass Metaanalysen zu diesen Populationen im Kontext dieser Arbeit nicht möglich waren. Um diese Situation in der Zukunft zu verbessern sind neue Assoziationsstudien zur Rolle der *CD33*-Polymorphismen und dem LOAD-Risiko in weiteren Abstammungsgruppen, z. B. denen aus Afrika oder dem südasiatischen Kontinent, notwendig.

### 5.3 Diskussion der Ergebnisse der eQTL GWAS Analysen

#### 5.3.1 Zusammenfassung der wichtigsten Ergebnisse

In diesem Teil der Arbeit habe ich selbständig eQTL-GWAS-Analysen zur Expression des *CD33*-Gens in humanen *postmortem* Gehirnproben durchgeführt. Dies wurde ermöglicht durch das parallele Vorliegen von genomweiten SNP-Genotypisierungs- und RNA-Sequenzierungsdaten in Proben des entorhinalen Kortex von n~200 AD-Patienten und Kontrollprobanden, die im Rahmen anderer Forschungsprojekte der Arbeitsgruppe generiert wurden. Die übergeordnete Frage dieses Teils der vorgelegten Dissertation war, ob und welche genetischen Polymorphismen (d.h. SNPs) einen signifikanten Effekt auf die Expression des *CD33*-Gens in dieser Stichprobe zeigen. Ein besonderes Augenmerk lag dabei auf den in Abschnitt 4.1. metaanalysierten sechs *CD33*-Polymorphismen. Die Klärung möglicher eQTL-Effekte ist nicht zuletzt daher von besonderem Interesse, weil von einigen Autoren bereits postuliert wurde [41, 65], dass der Mechanismus der beobachteten genetischen Assoziationen durch eine erhöhte *CD33*-Expression im Gehirn erklärt werden könnte. Zumindest in den im Rahmen dieser Arbeit analysierten Gehirnproben konnte in den durchgeföhrten *cis* eQTL-Analysen jedoch kein erwähnenswerter Zusammenhang in diese Richtung beobachtet werden: keiner der sechs in den Metaanalysen krankheitsassoziierten *CD33*-SNPs zeigte gleichzeitig eine signifikante Assoziation mit der Expression des *CD33*-Gens. Wenn man das Ergebnis mit den Ergebnissen der

durchgeführten Metaanalysen zusammen betrachtet, legen diese Befunde den Schluss nahe, dass der Mechanismus der beobachteten genetischen Assoziationen mit dem LOAD-Risiko nicht bzw. nicht hauptsächlich über die Genexpression vermittelt ist. Allerdings gibt es zu dieser Konklusion einige Limitationen, die ich weiter unten ausführe.

In einem zweiten Analyseschritt wurden zusätzlich weitere DNA-Sequenzvarianten in *cis* Lokalisation zum *CD33*-Gen, also in +/-500kb Entfernung zum Transkriptionsstart oder -Ende, auf mögliche Assoziationen mit der *CD33*-Expression im Gehirn untersucht. Von insgesamt 3442 analysierten SNPs, die in diesem Genabschnitt liegen (davon waren aufgrund der lokalen paarweisen LD-Struktur aber nur 382 als unabhängig zu betrachten) zeigte keine Variante eine Evidenz für genomweite Assoziation mit der *CD33*-Expression. Auch nach Bonferroni-Korrektur für alle unabhängigen im *CD33*-Bereich analysierten SNPs lag der signifikanteste eQTL P-Wert immer noch im Bereich der Nullhypothese. Die hier durchgeführten *cis* eQTL-Analysen legen somit nahe, dass genetische Polymorphismen im *CD33*-Gen nicht (bzw. nicht stark) in die eigene Expression im menschlichen Gehirn involviert sind, zumindest nicht in dem hier untersuchten Bereich des entorhinalen Kortex. Diese Ergebnisse decken sich weitestgehend mit denen der GTEx-Datenbank (V8), wo zwar einige moderate *cis* eQTL-Effekte beschrieben sind, diese sich aber alle auf subkortikale Regionen des Gehirns beschränken [19].

Zuletzt habe ich die Suche nach möglichen genetischen Determinanten der *CD33*-Expression auf das gesamte Genom ausgeweitet im Kontext von *trans* eQTL-Analysen. Die SNP-basierte GWAS-Analyse ergab dabei drei SNPs, die das genomweite Signifikanzniveau (P-Wert <5x10<sup>-8</sup>) unterschritten haben. Die signifikanteste Assoziation zeigte hierbei der SNP rs112851200, der sich in einem Intron des *MTX3*-Gens befindet. In diesem Zusammenhang ist interessant, dass ein anderes in kurzem Abstand zu *MTX3* lokalisiertes Gen (*THBS4*), bereits mit neurodegenerativen Prozessen in Verbindung gebracht wurde [27]. Der zweite genomweit signifikante *trans* eQTL-SNP (rs55934141) liegt im intergenischen, nicht-kodierenden Bereich des Genoms. Allerdings befinden sich laut GWAS-Katalog in unmittelbarer Nähe andere GWAS-Befunde, die einen Bezug zur Funktion des Immunsystems zeigen. Diese Koexistenz von unabhängigen Befunden legt nahe, dass ein Zusammenhang mit der *CD33*-vermittelten Immunantwort bestehen könnte, obwohl diese Konklusion der weiteren Überprüfung bedarf. Der dritte SNP rs2353260, der ebenfalls im intergenischen Bereich des Genoms liegt, hat das genomweite Signifikanzniveau nur marginal unterschritten. Interessanterweise wurde das dem SNP nächstliegende Gen, *CSMD3*, bereits funktionell mit dem Dendritenwachstum im Gehirn in Verbindung gebracht [98], was zumindest eine hypothetische Verbindung mit dem Entstehungsrisiko der AD darstellen könnte. Schließlich hat keines der Gen-basierten

Ergebnisse meiner *trans* eQTL-GWAS-Analysen die Schwelle der genomweiten Signifikanz unterschritten. Die zwei Gene mit der stärksten Assoziation mit der *CD33*-Expression im Gehirn (*RCN1* und *KIF3C*) wurden in Abschnitt 4.2.3 detaillierter geschildert. Erwähnenswert ist die Tatsache, dass das *KIF3C*-Gen nicht nur im zentralen Nervensystem stark exprimiert wird, sondern auch immunsystemrelevanten Funktionen ausweist, welche potenzielle Berührungs punkte mit der Immunantwort haben könnte, die durch das *CD33*-Gen vermittelt wird.

### 5.3.2 Stärken und Schwächen dieses Teils der Arbeit

Zu den Stärken der hier durchgeführten eQTL-Analysen gehört die Verwendung von genomweiten SNP- bzw. RNA-Expressionsdatensätzen, die in denselben Proben nach standardisierten Hochdurchsatz-Verfahren gewonnen wurden. Die zugrundeliegenden Proben entstammten ferner einer AD-relevanten Region (d.h. dem entorhinalen Kortex) des menschlichen Gehirns, für die es bisher nur sehr wenige Datensätze vergleichbarer Größe gibt. Durch das parallele Vorhandensein der genomweiten Genotypisierungsdaten, musste ich meine Suche nach potenziellen eQTLs der *CD33*-Expression nicht auf bestimmte Genomabschnitte beschränken, sondern konnte sie auf das gesamte Genom ausweiten.

Trotz dieser Stärken ist die Aussagekraft auch dieses Teils der Arbeit durch einige Limitation möglicherweise eingeschränkt. Hierzu gehört zuallererst die mit einer effektiven  $n=177$  vergleichsweise kleine Stichprobengröße. Obwohl dieser Stichprobenumfang für *postmortem* Analysen des menschlichen Gehirns noch als vergleichsweise groß anzunehmen ist (zum Vergleich: die GTEx-Datenbank hatte für viele AD-relevante kortikale Regionen in ihrer V8 eine kleinere effektive Stichprobengröße, z. B. Amygdala ( $n=129$ ), Hippocampus ( $n=165$ )), ist die Stichprobe im Kontext von GWAS-Analysen, die sich heutzutage häufig auf mehrere Hunderttausend Teilnehmer beziehen, als klein anzusehen. Dennoch gibt es nach meinem Kenntnisstand nur wenige Projekte, die eine vergleichbare oder gar größere Anzahl an *postmortem* Gehirnproben mit parallelen generierten genomweiten Genotypisierungs und RNA-Sequenzierungsdaten vorliegen haben, insofern sind die hier präsentierten Ergebnisse durchaus aus „state-of-the-art“ anzusehen. Sobald größere Stichprobengrößen verfügbar sind, sollten die hier ermittelten *cis* und *trans* eQTL-Ergebnisse aber unabhängig überprüft werden. Zweitens sind Genexpressionsstudien generell von einer Reihe von technischen (z. B. RNA-Extraktionsverfahren, Lagerungsbedingungen, RNA-Integrität, Post-Mortem-Interval) und biologischen (z. B. Alter, Zelltypzusammensetzung, Grad der Erkrankung, Todesumstände) Variablen beeinflusst, deren Messung und Berücksichtigung aufgrund der Vielzahl möglicher Faktoren schwierig bis unmöglich ist. Davon sind u. U. auch die Ergebnisse meiner

Analysen beeinträchtigt. Allerdings möchte ich hierbei hervorheben, dass inter-individuelle Unterschiede in diesen möglichen Störvariablen (engl. „confounder“) für die hier durchgeführten genetischen Assoziationsanalysen nur dann relevant sind, wenn sie sich systematisch über die verschiedenen Genotypgruppen hinweg unterscheiden. Diese Möglichkeit ist sehr unwahrscheinlich, da die Genotypen zum Zeitpunkt der Probenakquirierung nicht bekannt waren und die Gehirne in der Reihenfolge des Todeszeitpunktes gesammelt und gelagert wurden, nicht aber in Abhängigkeit von genetischen Faktoren, z. B. bestimmten *APOE*-Genotypen, o. ä. Drittens und teilweise zusammenhängend mit dem vorherigen Punkt, könnte der Diagnosestatus, d.h. ob es sich um einen AD-Patienten/Patientin oder eine Kontrollperson handelt, dann zu einer Verzerrung der eQTL-Analysen führen, wenn ein analysierter SNP gleichzeitig mit dem Krankheitsrisiko assoziiert ist, was ja z. B. für alle sechs metaanalysierten CD33-SNPs zutrifft. Diese mögliche Verzerrung wurde durch Einschluss des Diagnosestatus als Störvariable in die eQTL-Analysen ausgeschlossen (s. Kapitel 3). Viertens können in den hier durchgeführten eQTL-Analysen nur Aussagen zu möglichen Expressionseffekten im entorhinalen Kortex, nicht aber zu anderen Gehirnregionen getroffen werden. Fünftens lagen für die RNA-Sequenzierungsexperimente nur Gesamtgewebeproben vor (engl. „bulk tissue“), die nicht nach den vorliegenden unterschiedlichen Zelltypen aufgetrennt wurden. Das Vorliegen derartiger Zelltyp-spezifischer Datensätze aus menschlichen Gewebeproben würde auch eine Zelltyp-spezifische RNA-Sequenzierung erlauben (engl. „single cell sequencing“), die möglichen Unterschiede in den RNA-Sequenzprofilen und eQTL-Effekte beleuchten könnte. Die Durchführung von Zelltyp-spezifischen eQTL-Analysen wäre ein logischer und interessanter nächster Schritt bei der Charakterisierung von genetischen Effekten in der Genexpressionsregulierung.

#### 5.4 Schlussfolgerung und Ausblick

In der vorliegenden Dissertation wurde die Rolle des *CD33*-Gens in der Entstehung der AD anhand von genetischen Metaanalysen und genomweiten eQTL-Analysen evaluiert. Diese Arbeit stellt den aktuellsten Forschungsstand in beiden Bereichen dar und bestätigt, dass SNPs rs3865444, rs3826656, rs1697553, rs2455068, rs12459419 und rs35112940 zu den AD-assoziierten Polymorphismen gehören und somit das *CD33*-Gen ein AD-Risikogen darstellt.

Die gewonnenen genetischen Assoziationsergebnisse lassen sich durch die hier durchgeführten eQTL-Analysen besser einordnen, da die Ergebnisse dieses Teils der Arbeit nahelegen, dass die beobachteten genetischen Assoziation nicht bzw. nicht ausschließlich auf Effekte die die Genexpression betreffen zurückgeführt werden können.

Es wurden im Rahmen dieser Arbeit aber weitere genetische eQTL-assoziationssignale beobachtet, die mutmaßlich mit der immunologischen CD33-Funktion in Beziehung stehen können.

Obwohl aus der vorliegenden Arbeit neue Erkenntnisse über die Zusammenhänge zwischen dem *CD33*-Gen und AD-Risiko resultieren, bietet dieses Gebiet weitergehend viele Herausforderungen für die zukünftige Forschung. Diese Projekte könnten sich z. B. auf die Betrachtung weiterer AD-assoziierter Polymorphismen im *CD33*-Gen konzentrieren. Im Hinblick auf die untersuchten Abstammungsgruppen ist zu erkennen, dass der Hauptteil der bisherigen Untersuchungen in nordeuropäischen Datensätzen durchgeführt worden ist, gefolgt von Datensätzen ostasiatischer Abstammung. Genetische Assoziationsstudien mit weiteren Abstammungsgruppen könnten hier weitere Subgruppenanalysen ermöglichen und sind daher wichtig. Des Weiteren könnte und sollte ein kontinuierliches Verfolgen der Literaturdatenbanken erfolgen, um neue Publikationen zu diesem Thema in aktualisierten Metaanalysen einfließend zu bewerten.

Das Forschungsgebiet bietet jedoch noch weitere Potentiale. Zukünftige wissenschaftliche Arbeiten könnten an den bisherigen Erkenntnissen anknüpfen, indem weitere Analysen zu epigenetischen Modifikationen durchgeführt werden.

## 6 Zusammenfassung (Abstract)

Genetik spielt bei der Entstehung der Alzheimer-Krankheit (AD) eine wichtige Rolle, die geschätzte Heritabilität für die spätmanifeste AD beträgt 56-79% [37]. Dank einer Vielzahl von genetischen Assoziationsstudien (v.a. GWAS) sind seit dem Jahr 2008 fast 50 AD-Riskoloci reportiert worden [100]. Zahlreichen, v.a. in neueren GWAS identifizierten, AD-Risikogenen wird ein Bezug zur Immunabwehr zugeschrieben [10]. Die vorliegende Arbeit beschäftigte sich mit einem dieser AD-Riskoloci, d.h. mit dem Gen *CD33*. Ziel dieser Dissertation war es, zu untersuchen, ob und wie *CD33*-Polymorphismen die Entwicklung der AD und die *CD33*-Expression in menschlichen Gehirnen beeinflussen.

Für den ersten Teil meiner Arbeit habe ich einen systematischen Review der aktuellen Literatur der genetischen Assoziationsstudien von *CD33* und dem AD-Risiko durchgeführt und alle relevanten Daten in Metaanalysen kombiniert. Obwohl es schon andere publizierte Metaanalysen zu dieser Fragestellung gibt, stellt meine Arbeit die derzeit umfangreichste und aktuellste Zusammenführung von genetischen Assoziationsdaten zum Thema dar. Alle sechs identifizierten und metaanalysierten *CD33*-Polymorphismen zeigten dabei in mindestens einer der untersuchten Abstammungsgruppen einen statistisch signifikanten Zusammenhang mit dem AD-Risiko. Hierbei war die Datenlage für Probanden mit nordeuropäischer Abstammung am umfangreichsten. Insgesamt betrachtet liefern die hier generierten Ergebnisse überzeugende Beweise dafür, dass drei *CD33*-Polymorphismen das Risiko der AD signifikant erhöhen, wogegen drei weitere Varianten eine signifikante Assoziation mit reduzierten AD-Risiko zeigten.

Im zweiten Teil meiner Dissertation habe ich mich mit der Frage beschäftigt, ob und wenn ja welche genetischen Polymorphismen einen Einfluss auf die *CD33*-Genexpression im menschlichen Gehirn zeigen. Hierfür konnte ich *CD33*-Genexpressionsdaten (des entorhinalen Kortex) und genomweite SNP-Genotypisierungsprofile von insgesamt 177 Probanden nordeuropäischer Abstammung im Kontext von eQTL-GWAS-Analysen zusammenführen. Diese Analysen zeigten, dass nur die *trans* Varianten (also solche, die nicht in unmittelbarer Nähe zum *CD33*-Gen lokalisiert sind) genomweit signifikante Assoziation mit der Expression des *CD33*-Gens zeigten. Interessanterweise wiesen keiner der sechs in dem ersten Teil der Arbeit untersuchten Polymorphismen (und auch keine weiteren *cis* Varianten) nennenswerte Assoziationssignale auf.

Insgesamt lässt sich aus dieser Arbeit schließen, dass das *CD33*-Gen zwar einen wichtigen AD-Riskolocus darstellt, diese Assoziation aber eher nicht über die Genexpression vermittelt wird. Zukünftige Studien sollten die genetischen Analysen auf andere Ethnien ausweiten und die Pathomechanismen z. B. mittels funktioneller Experimente aufklären.

## 7 Literaturverzeichnis

1. Ahn E, Kang H (2018) Introduction to systematic review and meta-analysis. *Korean Journal of Anesthesiology* 71:103–112
2. AlzGene FIELD SYNOPSIS OF GENETIC ASSOCIATION STUDIES IN AD. <http://www.alzgene.org/>. Letzter Zugriff am: 20 Dec 2021
3. Alzheimer A (1907) Über eine eigenartige Erkrankung der Hirnrinde. *Allgemeine Zeitschrift für Psychiatrie und Psychisch-gerichtliche Medizin* 64:146–148
4. Ashford JW, Mortimer JA (2002) Non-familial Alzheimer's disease is mainly due to genetic factors. *Journal of Alzheimer's disease* 4:169–177
5. Bagyinszky E, Youn YC, An SSA, Kim S (2014) The genetics of Alzheimer's disease. *Clinical Interventions in Aging* 9:535–551
6. Bajaj T, Ramirez A, Wagner-Thelen H (2018) Genetik der Alzheimer-Krankheit. *medizinische genetik* 30:259–266
7. Bao J, Wang X, Mao Z (2016) Associations Between Genetic Variants in 19p13 and 19q13 Regions and Susceptibility to Alzheimer Disease: A Meta-Analysis. *Medical Science Monitor* 22:234–243
8. Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, Hogan MF, Schjeide BMM, Hooli B, Divito J, Ionita I, Jiang H, Laird N, Moscarillo T, Ohlsen KL, Elliott K, Wang X, Hu-Lince D, Ryder M, Murphy A, Wagner SL, Blacker D, Becker KD, Tanzi RE (2008) Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *American Journal of Human Genetics* 83:623–632
9. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature Genetics* 39:17–23
10. Bertram L, Tanzi RE (2020) Genomic mechanisms in Alzheimer's disease. *Brain Pathology* 30:966–977
11. BGI - Cognitive Genomics Lab Plink 2.0. <https://www.cog-genomics.org/plink/2.0/>. Letzter Zugriff am: 12 Dec 2021
12. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to Meta-Analysis, 1. Edition, 3–7, 61–62, 277–292. Wiley (Hrsg.), West Sussex, UK
13. Bradshaw EM, Chibnik LB, Keenan BT, Ottoboni L, Raj T, Tang A, Rosenkrantz LL, Imboywa S, Lee M, Von Korff A, Alzheimer Disease Neuroimaging Initiative, Morris MC, Evans DA, Johnson K, Sperling RA, Schneider JA, Bennett DA, De Jager PL (2013) CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. *Nature Neuroscience* 16:848–850
14. Breijyeh Z, Karaman R (2020) Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules* 25:5789 1–28

15. Brigham and Women's Hospital Profil Gregory A. Petsko. <https://researchfaculty.brighamandwomens.org/BRIProfile.aspx?id=6606>. Letzter Zugriff am: 7 Feb 2021
16. Broad Institute of MIT and Harvard GTEx Portal - rs112851200. <https://www.gtexportal.org/home/snp/rs112851200>. Letzter Zugriff am: 20 Dec 2021
17. Broad Institute of MIT and Harvard GTEx Portal - CSMD3. <https://www.gtexportal.org/home/gene/CSMD3>. Letzter Zugriff am: 20 Dec 2021
18. Broad Institute of MIT and Harvard GTEx Portal - KIF3C. <https://www.gtexportal.org/home/gene/KIF3C>. Letzter Zugriff am: 20 Dec 2021
19. Broad Institute of MIT and Harvard GTEx Portal - CD33. <https://www.gtexportal.org/home/gene/CD33>. Letzter Zugriff am: 20 Dec 2021
20. Carrasquillo MM, Belbin O, Hunter TA, Ma L, Bisceglie GD, Zou F, Crook JE, Pankratz V, Sando SB, Aasly JO, Barcikowska M, Wszolek ZK, Dickson DW, Graff-Radford NR, Petersen RC, Passmore P, Morgan K, Younkin SG (2011) Replication of EPHA1 and CD33 associations with late-onset Alzheimer's disease: a multi-centre case-control study. *Molecular Neurodegeneration* 6:54 1–9
21. Chen X, Shao W, Huang H, Feng X, Yao S, Ke H (2019) Overexpression of RCN1 correlates with poor prognosis and progression in non-small cell lung cancer. *Human Pathology* 83:140–148
22. Chung SJ, Lee J-H, Kim SY, You S, Kim MJ, Lee J-Y, Koh J (2013) Association of GWAS Top Hits With Late-onset Alzheimer Disease in Korean Population. *Alzheimer Disease & Associated Disorders* 27:250–257
23. Cochrane Training RevMan. <https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman>. Letzter Zugriff am: 8 Dec 2021
24. Çoğaltay N, Karadağ E (2015) Introduction to Meta-Analysis. In: Leadership and Organizational Outcomes. Springer International Publishing (Hrsg.), Eskisehir, Turkey, 19–28
25. Cook DJ (1997) Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions. *Annals of Internal Medicine* 126:376–380
26. Corder EH, Saunders AM, Strittmatter WJ, Schmeichel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261:921–923
27. Das S, Li Z, Noori A, Hyman BT, Serrano-Pozo A (2020) Meta-analysis of mouse transcriptomic studies supports a context-dependent astrocyte reaction in acute CNS injury versus neurodegeneration. *Journal of Neuroinflammation* 17:227 1–17
28. Deng Y-L, Liu L-H, Wang Y, Tang H-D, Ren R-J, Xu W, Ma J-F, Wang L-L, Zhuang J-P, Wang G, Chen S-D (2012) The prevalence of CD33 and MS4A6A variant in Chinese Han population with Alzheimer's disease. *Human Genetics* 131:1245–1249

29. Dobricic V, Schilling M, Schulz J, Zhu L-S, Zhou C-W, Fuß J, Franzenburg S, Zhu L-Q, Parkkinen L, Lill CM, Bertram L (2021) Differential microRNA expression analyses across two brain regions in Alzheimer's disease. *bioRxiv* 446406 (Preprint)
30. Dubois B, Hampel H, Feldman HH, Scheltens P, Aisen P, Andrieu S, Bakardjian H, Benali H, Bertram L, Blennow K, Broich K, Cavedo E, Crutch S, Dartigues J-F, Duyckaerts C, Epelbaum S, Frisoni GB, Gauthier S, Genton R, Gouw AA, Habert M-O, Holtzman DM, Kivipelto M, Lista S, Molinuevo J-L, O'Bryant SE, Rabinovici GD, Rowe C, Salloway S, Schneider LS, Sperling R, Teichmann M, Carrillo MC, Cummings J, Jack CR (2016) Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 12:292–323
31. Ebbert MTW, Ridge PG, Wilson AR, Sharp AR, Bailey M, Norton MC, Tschanz JT, Munger RG, Corcoran CD, Kauwe JSK (2014) Population-based Analysis of Alzheimer's Disease Risk Alleles Implicates Genetic Interactions. *Biological Psychiatry* 75:732–737
32. Egger M, Smith GD, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315:629–634
33. Ensembl Gene: CD33. [http://www.ensembl.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000105383;r=19:51225064-51243860](http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000105383;r=19:51225064-51243860). Letzter Zugriff am: 8 Dec 2021
34. Ensembl Variant Effect Predictor Tool. <https://www.ensembl.org/Tools/VEP>. Letzter Zugriff am: 20 Dec 2021
35. Estus S, Shaw BC, Devaney N, Katsumata Y, Press EE, Fardo DW (2019) Evaluation of CD33 as a Genetic Risk Factor for Alzheimer's Disease. *Acta Neuropathologica* 138:187–199
36. Gao Y, Zheng H, Li L, Zhou C, Chen X, Zhou X, Cao Y (2020) KIF3C Promotes Proliferation, Migration, and Invasion of Glioma Cells by Activating the PI3K/AKT Pathway and Inducing EMT. *BioMed Research International* 2020:6349312 1–10
37. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL (2006) Role of genes and environments for explaining Alzheimer disease. *Archives of General Psychiatry* 63:168–174
38. Gogtay N, Thatte U (2017) An Introduction to Meta-Analysis. *Journal of The Association of Physicians of India* 65:78–85
39. Griciuc A, Federico AN, Natasan J, Forte AM, McGinty D, Nguyen H, Volak A, LeRoy S, Gandhi S, Lerner EP, Hudry E, Tanzi RE, Maguire CA (2020) Gene therapy for Alzheimer's disease targeting CD33 reduces amyloid beta accumulation and neuroinflammation. *Human Molecular Genetics* 29:2920–2935
40. Griciuc A, Serrano-Pozo A, Parrado AR, Lesinski AN, Asselin CN, Mullin K, Hooli B, Choi SH, Hyman BT, Tanzi RE (2013) Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron* 78:631–643
41. Griciuc A, Tanzi RE (2021) The role of innate immune genes in Alzheimer's disease. *Current Opinion in Neurology* 34:228–236

42. Gumy LF, Chew DJ, Tortosa E, Katrukha EA, Kapitein LC, Tolkovsky AM, Hoogenraad CC, Fawcett JW (2013) The kinesin-2 family member KIF3C regulates microtubule dynamics and is required for axon growth and regeneration. *The Journal of Neuroscience* 33:11329–11345
43. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, Welch V (2019) Cochrane Handbook for Systematic Reviews of Interventions, 2. Edition, xxiii–xxvi, 3–280, 321–370., John Wiley & Sons (Hrsg.), Chichester, UK
44. Higgins JPT, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ* 327:557–560
45. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Jones N, Stretton A, Thomas C, Jöckel K-H, [...], Scarpini E, Galimberti D, Brice A, Hannequin D, Licastro F, Jones L, Holmans PA, Jonsson T, Riemenschneider M, Morgan K, Younkin SG, Owen MJ, O'Donovan M, Amouyel P, Williams J (2011) Common variants in ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genetics* 43:429–435
46. Hong S, Dobricic V, Ohlei O, Bos I, Vos SJB, Prokopenko D, Tijms BM, Andreasson U, Blennow K, Vandenberghe R, Gabel S, Scheltens P, Teunissen CE, Engelborghs S, Frisoni G, Blin O, Richardson JC, Bordet R, Alzheimer's Disease Neuroimaging Initiative (ADNI), Lleó A, Alcolea D, Popp J, Clark C, Peyratout G, Martinez-Lage P, Tainta M, Dobson RJB, Legido-Quigley C, Sleegers K, Van Broeckhoven C, Tanzi RE, Ten Kate M, Wittig M, Franke A, Lill CM, Barkhof F, Lovestone S, Streffer J, Zetterberg H, Visser PJ, Bertram L (2021) TMEM106B and CPOX are genetic determinants of cerebrospinal fluid Alzheimer's disease biomarker levels. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 17:1628–1640
47. Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, Riley J, Purvis I, Xu C-F (2004) Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *European Journal of Human Genetics* 12:395–399
48. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, Sealock J, Karlsson IK, Hägg S, Athanasiu L, Voyle N, Proitsi P, Witoelar A, Stringer S, Aarsland D, [...], Ulstein ID, Wang Y, White LR, Hardy J, Hjerling-Leffler J, Sullivan PF, van der Flier WM, Dobson R, Davis LK, Stefansson H, Stefansson K, Pedersen NL, Ripke S, Andreassen OA, Postuma D (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics* 51:404–413
49. Javor J, Ďurmanová V, Párnická Z, Minárik G, Králová M, Pečeňák J, Vašečková B, Režnáková V, Šutovský S, Gmitterová K, Hromádka T, Peterajová L, Shawkatová I (2020) Association of CD33 rs3865444:C>A polymorphism with a reduced risk of late-onset Alzheimer's disease in Slovaks is limited to subjects carrying the APOE ε4 allele. *International Journal of Immunogenetics* 47:397–405
50. Jiang Y-T, Li H-Y, Cao X-P, Tan L (2018) Meta-analysis of the association between CD33 and Alzheimer's disease. *Annals of Translational Medicine* 6:169–178
51. Jiao B, Liu X, Zhou L, Wang MH, Zhou Y, Xiao T, Zhang W, Sun R, Waye MMY, Tang B, Shen L (2015) Polygenic Analysis of Late-Onset Alzheimer's Disease from Mainland China. *PLOS ONE* 10:e0144898: 1-10

52. Krnic Martinic M, Pieper D, Glatt A, Puljak L (2019) Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks. *BMC Medical Research Methodology* 19:203–214
53. Kumar A, Sidhu J, Goyal A, Tsao JW (2021) Alzheimer Disease. In: StatPearls. StatPearls Publishing (Hrsg.), Treasure Island (Florida), USA, 1–9
54. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, van der Lee SJ, Amlie-Wolf A, Bellenguez C, Frizatti A, Chouraki V, Martin ER, Sleegers K, [...], Haines JL, Deleuze J-F, Gudnason V, Mayeux R, Escott-Price V, Psaty BM, Ramirez A, Ruiz A, van Duijn CM, Holmans PA, Seshadri S, Amouyel P, Schellenberg GD, Lambert J-C, Pericak-Vance MA (2019) Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nature Genetics* 51:414–430
55. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, DeStefano AL, Bis JC, Beecham GW, Grenier-Boley B, Russo G, Thornton-Wells TA, Jones N, Smith AV, [...], Graff C, Psaty BM, Jones L, Haines JL, Holmans PA, Lathrop M, Pericak-Vance MA, Farrer LA, van Duijn CM, Van Broeckhoven C, Moskvina V, Seshadri S, Williams J, Schellenberg GD, Amouyel P (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* 45:1452–1458
56. Lane CA, Hardy J, Schott JM (2018) Alzheimer's disease. *European Journal of Neurology* 25:59–70
57. de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015) MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Computational Biology* 11:e1004219: 1–19
58. Lewis CM (2002) Genetic association studies : Design, analysis and interpretation. 3:146–153
59. Li W-W, Wang Z, Fan D-Y, Shen Y-Y, Chen D-W, Li H-Y, Li L, Yang H, Liu Y-H, Bu X-L, Jin W-S, Zeng F, Xu Z-Q, Yu J-T, Chen L-Y, Wang Y-J (2020) Association of Polygenic Risk Score with Age at Onset and Cerebrospinal Fluid Biomarkers of Alzheimer's Disease in a Chinese Cohort. *Neuroscience Bulletin* 36:696–704
60. Li X, Shen N, Zhang S, Liu J, Jiang Q, Liao M, Feng R, Zhang L, Wang G, Ma G, Zhou H, Chen Z, Jiang Y, Zhao B, Li K, Liu G (2015) CD33 rs3865444 Polymorphism Contributes to Alzheimer's Disease Susceptibility in Chinese, European, and North American Populations. *Molecular Neurobiology* 52:414–421
61. Lill CM, Rengmark A, Pihlstrøm L, Fogh I, Shatunov A, Sleiman PM, Wang L-S, Liu T, Lassen CF, Meissner E, Alexopoulos P, Calvo A, Chio A, Dizdar N, Faltraco F, [...], Veldink JH, Lindenberger U, Steinhagen-Thiessen E, SLAGEN Consortium, Teipel S, Perneczky R, Hakonarson H, Hampel H, von Arnim CAF, Olsen JH, Van Deerlin VM, Al-Chalabi A, Toft M, Ritz B, Bertram L (2015) The role of TREM2 R47H as a risk factor for Alzheimer's disease, frontotemporal lobar degeneration, amyotrophic lateral sclerosis, and Parkinson's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 11:1407–1416
62. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide B-MM, Schjeide LM, Meissner E, Zauft U, Allen NC, Liu T, Schilling M, Anderson KJ, Beecham G, Berg D, [...], Tung JY, Vance J, Wood NW, Zabetian CP, 23andMe TGE of PD (GEO-PC, Consortium (IPDGC) TIPDG, Consortium TPDG, Consortium 2 (WTCCC2)

- TWTCC, Young P, Tanzi RE, Khouri MJ, Zipp F, Lehrach H, Ioannidis JPA, Bertram L (2012) Comprehensive Research Synopsis and Systematic Meta-Analyses in Parkinson's Disease Genetics: The PDGene Database. *PLOS Genetics* 8:e1002548 1-10
63. Livingston G, Sommerlad A, Ortega V, Costafreda SG, Huntley J, Ames D, Ballard C, Banerjee S, Burns A, Cohen-Mansfield J, Cooper C, Fox N, Gitlin LN, Howard R, Kales HC, Larson EB, Ritchie K, Rockwood K, Sampson EL, Samus Q, Schneider LS, Selbæk G, Teri L, Mukadam N (2017) Dementia prevention, intervention, and care. *The Lancet* 390:2673–2734
64. Logue MW (2011) A Comprehensive Genetic Association Study of Alzheimer Disease in African Americans. *Archives of Neurology* 68:1569–1579
65. Malik M, Simpson JF, Parikh I, Wilfred BR, Fardo DW, Nelson PT, Estus S (2013) CD33 Alzheimer's Risk-Altering Polymorphism, CD33 Expression, and Exon 2 Splicing. *The Journal of Neuroscience* 33:13320–13325
66. Mao Y-F, Guo Z-Y, Pu J-L, Chen Y-X, Zhang B-R (2015) Association of CD33 and MS4A cluster variants with Alzheimer's disease in East Asian populations. *Neuroscience Letters* 609:235–239
67. Mehdizadeh E, Khalaj-Kondori M, Shaghaghi-Tarakdari Z, Sadigh-Eteghad S, Talebi M, Andalib S (2019) Association of MS4A6A, CD33, and TREM2 gene polymorphisms with the late-onset Alzheimer's disease. *Biolimpacts* 9:219–225
68. Miyashita A, Koike A, Jun G, Wang L-S, Takahashi S, Matsubara E, Kawarabayashi T, Shoji M, Tomita N, Arai H, Asada T, Harigaya Y, Ikeda M, Amari M, Hanyu H, [...], St. George-Hyslop P, Mayeux R, Haines JL, Pericak-Vance MA, Yoshida M, Nishida N, Tokunaga K, Yamamoto K, Tsuji S, Kanazawa I, Ihara Y, Schellenberg GD, Kuwano R (2013) SORL1 Is Genetically Associated with Late-Onset Alzheimer's Disease in Japanese, Koreans and Caucasians. *PLOS ONE* 8:e58618 1-13
69. Mizukami T, Kohno T, Hattori M (2016) CUB and Sushi multiple domains 3 regulates dendrite development. *Neuroscience Research* 110:11–17
70. Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 339:b2535 1-8
71. Moreno DJ, Ruiz S, Ríos Á, Lopera F, Ostos H, Via M, Bedoya G (2017) Association of GWAS Top Genes With Late-Onset Alzheimer's Disease in Colombian Population. *American Journal of Alzheimer's Disease & Other Dementias* 32:27–35
72. Moreno-Grau S, Hernández I, Heilmann-Heimbach S, Ruiz S, Rosende-Roca M, Mauleón A, Vargas L, Rodríguez-Gómez O, Alegret M, Espinosa A, Ortega G, Aguilera N, Abdelnour C, Neuroimaging Initiative AD, Gil S, Maier W, Sotolongo-Grau O, Tárraga L, Ramirez A, López-Arrieta J, Antúnez C, Serrano-Ríos M, Boada M, Ruiz A (2018) Genome-wide significant risk factors on chromosome 19 and the APOE locus. *Oncotarget* 9:24590–24600
73. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, [...], Saykin AJ, Reiman EM, Bennett DA, Morris JC, Montine TJ, Blacker D, Tsuang DW, Hakonarson H, Kukull WA, Foroud TM, Haines JL, Mayeux

- R, Pericak-Vance MA, Farrer LA, Schellenberg GD (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature Genetics* 43:436–441
74. National Center for Biotechnology Information PubMed. <https://pubmed.ncbi.nlm.nih.gov/>. Letzter Zugriff am: 20 Dec 2021
75. National Human Genome Research Institute EBI Catalog of human genome-wide association studies - chr6:23286454-23786454. <https://www.ebi.ac.uk/gwas/regions/chr6:23286454-23786454>. Letzter Zugriff am: 20 Dec 2021
76. National Human Genome Research Institute EBI Catalog of human genome-wide association studies - CSMD3. <https://www.ebi.ac.uk/gwas/genes/CSMD3>. Letzter Zugriff am: 20 Dec 2021
77. National Human Genome Research Institute EBI Catalog of human genome-wide association studies - RCN1. <https://www.ebi.ac.uk/gwas/genes/RCN1>. Letzter Zugriff am: 20 Dec 2021
78. National Human Genome Research Institute EBI Catalog of human genome-wide association studies - KIF3C. <https://www.ebi.ac.uk/gwas/genes/KIF3C>. Letzter Zugriff am: 20 Dec 2021
79. National Institute for Health and Care Excellence NICE. <https://www.nice.org.uk/>. Letzter Zugriff am: 8 Dec 2021
80. National Library of Medicine dbSNP. [https://www.ncbi.nlm.nih.gov/snp/rs3826656#frequency\\_tab](https://www.ncbi.nlm.nih.gov/snp/rs3826656#frequency_tab). Letzter Zugriff am: 9 Dec 2021
81. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J, Cardon LR, Whittaker JC, Sanseau P (2015) The support of human genetic evidence for approved drug indications. *Nature Genetics* 47:856–860
82. Nimmrich I, Erdmann S, Melchers U, Finke U, Hentsch S, Moyer MP, Hoffmann I, Müller O (2000) Seven genes that are differentially transcribed in colorectal tumor cell lines. *Cancer Letters* 160:37–43
83. Omoumi A, Fok A, Greenwood T, Sadovnick AD, Feldman HH, Hsiung G-YR (2014) Evaluation of late-onset Alzheimer disease genetic susceptibility risks in a Canadian population. *Neurobiology of Aging* 35:936.e5-936.e12
84. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* n71 1-10
85. Patel T, Brookes KJ, Turton J, Chaudhury S, Guetta-Baranes T, Guerreiro R, Bras J, Hernandez D, Singleton A, Francis PT, Hardy J, Morgan K (2018) Whole-exome sequencing of the BDR cohort: evidence to support the role of the PILRA gene in Alzheimer's disease. *Neuropathology and Applied Neurobiology* 44:506–521

86. Pati D, Lorusso LN (2018) How to Write a Systematic Review of the Literature. *Health Environments Research & Design Journal* 11:15–30
87. Petsko G (2008) The coming neurological epidemic. [https://www.ted.com/talks/gregory\\_petsko\\_the\\_coming\\_neurological\\_epidemic](https://www.ted.com/talks/gregory_petsko_the_coming_neurological_epidemic). Letzter Zugriff am: 7 Feb 2021
88. Portin P, Wilkins A (2017) The Evolving Definition of the Term “Gene.” *Genetics* 205:1353–1364
89. Raikwar SP, Thangavel R, Dubova I, Ahmed ME, Selvakumar PG, Kempuraj D, Zaheer S, Iyer S, Zaheer A Neuro-Immuno-Gene- and Genome-Editing-Therapy for Alzheimer’s Disease: Are We There Yet? *Journal of Alzheimer’s Disease* 65:321–344
90. Rücker G, Schwarzer G, Carpenter JR, Schumacher M (2008) Undue reliance on I<sup>2</sup> in assessing heterogeneity may mislead. *BMC Medical Research Methodology* 8:79–87
91. dos Santos LR, Pimassoni LHS, Sena GGS, Camporez D, Belcavello L, Trancozo M, Morelato RL, Errera FIV, Bueno MRP, de Paula F (2017) Validating GWAS Variants from Microglial Genes Implicated in Alzheimer’s Disease. *Journal of Molecular Neuroscience* 62:215–221
92. Sassi C, Nalls MA, Ridge PG, Gibbs JR, Ding J, Lupton MK, Troakes C, Lunnon K, Al-Sarraj S, Brown KS, Medway C, Clement N, Lord J, Turton J, Bras J, [...], Todd S, Heun R, Kölsch H, Kehoe PG, Vardy ERLC, Hooper NM, Mann DM, Pickering-Brown S, Brown K, Lowe J, Morgan K, Smith AD, Wilcock G, Warden D, Holmes C (2016) ABCA7 p.G215S as potential protective factor for Alzheimer’s disease. *Neurobiology of Aging* 46:235.e1–235.e9
93. Schmucker C, Nothacker M, Möhler R, Kopp I, Meerpolh JJ (2017) Bewertung des Verzerrungsrisikos von systematischen Übersichtsarbeiten: ein Manual für die Leitlinienerstellung, 1. Auflage, 1–35. Cochrane Deutschland, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement (Hrsg.), Marburg, Deutschland
94. Schwarzer G, Carpenter JR, Rücker G (2015) Meta-Analysis with R, 1. Edition, 3–16, 21–52, 117–119, 237–245. Springer International Publishing, Switzerland
95. Schwarzer G, Rücker G (2010) Statistische Methoden zur Detektion und Adjustierung von Publikationsbias. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 104:306–313
96. Serrano-Pozo A, Das S, Hyman BT (2021) APOE and Alzheimer’s disease: advances in genetics, pathophysiology, and therapeutic approaches. *The Lancet Neurology* 20:68–80
97. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 358:j4008 1-9
98. Shimizu A, Asakawa S, Sasaki T, Yamazaki S, Yamagata H, Kudoh J, Minoshima S, Kondo I, Shimizu N (2003) A novel giant gene CSMD3 encoding a protein with CUB

- and sushi multiple domains: a candidate gene for benign adult familial myoclonic epilepsy on human chromosome 8q23.3–q24.1. *Biochemical and Biophysical Research Communications* 309:143–154
99. Sia GM, Clem RL, Huganir RL (2013) The human language and epilepsy associated gene SRPX2 regulates synapse formation and vocalization in mice. *Science* 342:987–991
  100. Sims R, Hill M, Williams J (2020) The multiplex model of the genetics of Alzheimer's disease. *Nature Neuroscience* 23:311–322
  101. Singer JB (2009) Candidate Gene Association Analysis. In: DiPetrillo K (ed) *Cardiovascular Genomics: Methods and Protocols*. Methods in Molecular Biol. Humana Press (Hrsg.), Totowa, New Jersey, 223–230
  102. Siokas V, Aslanidou P, Aloizou A-M, Peristeri E, Stamatilou P, Liampas I, Arseniou S, Drakoulis N, Aschner M, Tsatsakis A, Mitsias PD, Bogdanos DP, Hadjigeorgiou GM, Dardiotis E (2020) Does the CD33 rs3865444 Polymorphism Confer Susceptibility to Alzheimer's Disease? *Journal of Molecular Neuroscience* 70:851–860
  103. Sommerer Y, Dobricic V, Schilling M, Ohlei O, Sabet SS, Wesse T, Fuß J, Franzenburg S, Franke A, Parkkinen L, Lill CM, Bertram L (2021) Entorhinal cortex EWAS meta-analysis highlights four novel loci showing differential methylation in Alzheimer's disease. *bioRxiv* 450878 (Preprint)
  104. Statista Statistiken zum Thema Demenz weltweit. <https://de.statista.com/themen/2032/demenzerkrankungen-weltweit/>. Letzter Zugriff am: 23 Feb 2021
  105. Strittmatter WJ, Saunders AM, Schmeichel D, Pericak-Vance M, Enghild J, Salvesen GS, Roses AD (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America* 90:1977–1981
  106. Tan L, Yu J-T, Zhang W, Wu Z-C, Zhang Q, Liu Q-Y, Wang W, Wang H-F, Ma X-Y, Cui W-Z (2013) Association of GWAS-linked loci with late-onset Alzheimer's disease in a northern Han Chinese population. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 9:546–553
  107. The UniProt Consortium UniProt: the universal protein knowledgebase in 2021. <https://www.uniprot.org/uniprot/Q5HYI7>. Letzter Zugriff am: 20 Dec 2021
  108. Veitch DP, Weiner MW, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Morris JC, Petersen RC, Saykin AJ, Shaw LM, Toga AW, Trojanowski JQ, Alzheimer's Disease Neuroimaging Initiative (2019) Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 15:106–152
  109. Victor A (2007) Bewertung genetischer Forschungsergebnisse - Methoden und vermeidbare Fehler. *Gesundheit und Gesellschaft : G+G : das AOK-Forum für Politik, Praxis und Wissenschaft* 4:14–22

110. Walker DG, Whetzel AM, Serrano G, Sue LI, Beach TG, Lue L-F (2015) Association of CD33 polymorphism rs3865444 with Alzheimer's disease pathology and CD33 expression in human cerebral cortex. *Neurobiology of Aging* 36:571–582
111. Watanabe K, Taskesen E, van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* 8:1826 1–11
112. Wattmo C, Minthon L, Wallin ÅK (2016) Mild versus moderate stages of Alzheimer's disease: three-year outcomes in a routine clinical setting of cholinesterase inhibitor therapy. *Alzheimer's Research & Therapy* 8:7 1-15
113. WHO reveals leading causes of death and disability worldwide: 2000-2019. <https://www.who.int/news-room/detail/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019>. Letzter Zugriff am: 29 Dec 2020
114. WHO Dementia - Key Facts. <https://www.who.int/news-room/fact-sheets/detail/dementia>. Letzter Zugriff am: 11 Sep 2021
115. Wolf SA, Boddeke HWGM, Kettenmann H (2017) Microglia in Physiology and Disease. *Annual Review of Physiology* 79:619–643
116. Xia X, Wang Y, Zheng J (2021) COVID-19 and Alzheimer's disease: how one crisis worsens the other. *Translational Neurodegeneration* 10:15 1-17
117. York U of Centre for Reviews and Dissemination - Centre for Reviews and Dissemination, University of York. <https://www.york.ac.uk/crd/>. Letzter Zugriff am: 8 Dec 2021
118. Yuan Q, Chu C, Jia J (2012) Association studies of 19 candidate SNPs with sporadic Alzheimer's disease in the North Chinese Han population. *Neurological Sciences* 33:1021–1028
119. Zhang D-F, Li J, Wu H, Cui Y, Bi R, Zhou H-J, Wang H-Z, Zhang C, Wang D, Alzheimer's Disease Neuroimaging Initiative (ADNI), Kong Q-P, Li T, Fang Y, Jiang T, Yao Y-G (2016) CFH Variants Affect Structural and Functional Brain Changes and Genetic Risk of Alzheimer's Disease. *Neuropsychopharmacology* 41:1034–1045
120. Zhao L (2019) CD33 in Alzheimer's Disease - Biology, Pathogenesis, and Therapeutics: A Mini-Review. *Gerontology* 65:323–331

## 8 Anhang

### **Anhang 1: Datensatz der für die Metaanalysen verwendeten Studien**

Der gesamte Datensatz der für die Metaanalysen verwendeten Studien befindet sich in Form einer MS-Excel-Datei auf der beigefügten CD-ROM

## Anhang 2: ausgefüllte PRISMA 2020 Checkliste für die vorliegende Dissertation

Section and Topic	Item	Checklist item	Location where item is reported
<b>TITLE</b>			
Title	1	Identify the report as a systematic review.	Kapitel 2.5 - Fragestellung und Ziel der Arbeit
<b>ABSTRACT</b>			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Kapitel 6 - Zusammenfassung (Abstract)
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Kapitel 2 – Einleitung und Fragestellung
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	
<b>METHODS</b>			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Kapitel 3.3 - Auswahlkriterien inklusive Ein- und Ausschlusskriterien
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Kapitel 3.4 - Informationsquellen und Datensuchstrategie
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Kapitel 3.5 - Auswahl der Studien und Datensammlung
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Kapitel 3.6 - Statische Auswertung und Reduzierung von Biases in der Metaanalyse

Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Kapitel 3.6 - Statische Auswertung und Reduzierung von Biases in der Metaanalyse
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	
<b>RESULTS</b>			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Kapitel 4.1.1 - PRISMA Flow Chart
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	
Study characteristics	17	Cite each included study and present its characteristics.	Kapitel 4.1.1 - PRISMA Flow Chart Kapitel 7 - Literaturverzeichnis
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	n/a
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Anhang 1
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Kapitel 4.1.2 - Überblick Ergebnisse der eigenen Metaanalysen

	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Kapitel 4.1 - Metaanalysen genetischer Assoziationsstudien zur Rolle des CD33-Gens in der AD
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Kapitel 4.1.2 - Überblick Ergebnisse der eigenen Metaanalysen
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	n/a
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Kapitel 4.1 - Metaanalysen genetischer Assoziationsstudien zur Rolle des CD33-Gens in der AD
<b>DISCUSSION</b>			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Kapitel 5.2 - Diskussion der Ergebnisse der Metaanalysen
	23b	Discuss any limitations of the evidence included in the review.	Kapitel 5.2.3 - Stärken und Schwächen der Arbeit
	23c	Discuss any limitations of the review processes used.	
	23d	Discuss implications of the results for practice, policy, and future research.	Kapitel 5.4 - Schlussfolgerung und Ausblick

### Anhang 3: R Code für die Metaanalysen

```
#Installiere Metastudien Auswertung-Libaries
install.packages("meta")
library(meta)

#Dataset aus Tabelle in R Studio laden
# Arbeitsordner angeben, in der die einzulesende csv liegt
setwd("//xy/Data/Doktorarbeit/ Datentabelle")

# Einlesen der Tabelle
library(readxl)
data = read_excel("rs35112940.xlsx", col_types = c("text", "numeric", "numeric",
                                                 "numeric", "numeric", "numeric"))

# Statische Auswertung mit Metagen
# comb.fixed / comb.random wurden je nach effect model anpassen
ergebnis = metagen(TE=log(OR),
                    seTE=SE,
                    data=data,
                    studlab = Autor,
                    n.e = cases,
                    n.c = control,
                    comb.fixed = FALSE,
                    comb.random = TRUE,
                    method.tau = "DL",
                    #hakn = TRUE,
                    #prediction = TRUE,
                    sm = "OR")
```

```

# Ergebnis plotten

# bei leftcols = "Null" wenn TE und seTE auch erscheinen sollen

# test.overall.random = TRUE oder test.overall.fixed = TRUE bzw. w.random oder

# w.fixed je nach effect model anpassen

forest(ergebnis, test.overall.random = TRUE, layout = "meta", scientific.pval = TRUE,
digits.pval.Q = 2, digits.pval = 2, leftcols = c("studlab", "Year", "n.e", "n.c"), leftlabs =
c("Involved Studies", "Year of Study", "Cases", "Controls"), rightlabs = NULL, rightcols =
c("effect", "ci", "w.random"), sortvar = Year, digits.weight = 2, type.fixed = "diamond",
col.diamond.fixed = "blue", col.diamond.lines.fixed = "black", type.random = "diamond",
col.diamond.random = "blue", col.diamond.lines.random = "black")

# Funnel-Plot erstellen

funnel(ergebnis)

# Egger's Test (Test auf Publication Bias)

metabias(ergebnis, method.bias = "Egger")

metabias(ergebnis, method.bias = "Egger", k.min=1)$pval

metabias(ergebnis, method.bias = "Egger", plotit = TRUE, k.min=1)

```

#### Anhang 4: ausgefüllte AMSTAR Checkliste für die vorliegende Dissertation

1. Wurde die Übersichtsarbeit a priori geplant/definiert?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

2. Wurde die Studienauswahl und Datenextraktion von 2 Personen unabhängig voneinander ausgeführt?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

3. Wurde eine umfassende und systematische Literatursuche durchgeführt?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

4. Wurden unpublizierte Studiendaten und graue Literatur in der Übersichtsarbeit berücksichtigt?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

5. Wurden die Referenzen der ein- und ausgeschlossenen Studien in der Übersichtsarbeit angegeben?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

6. Wurden die Studiencharakteristika (Charakteristika der Patienten, Intervention(en) und Endpunkte) der eingeschlossenen Studien in Tabellenform oder ausführlich in Textform angegeben?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

7. Wurde das Verzerrungsrisiko der eingeschlossenen Primärstudien nach etablierten Methoden bewertet?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

8. Wurde das Verzerrungsrisiko der eingeschlossenen Studien in der Ergebnisinterpretation der Übersichtsarbeit berücksichtigt?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

9. Wurden die Studienergebnisse statistisch adäquat ausgewertet?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

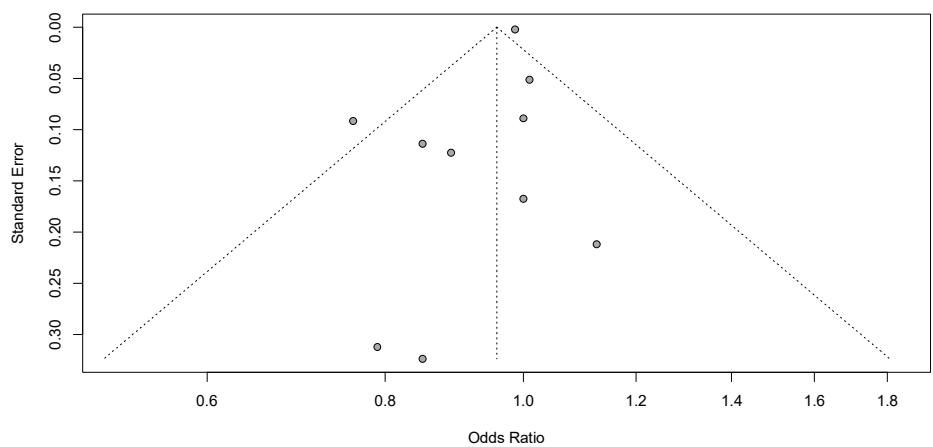
10. Wurde potentieller Publikationsbias adressiert?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

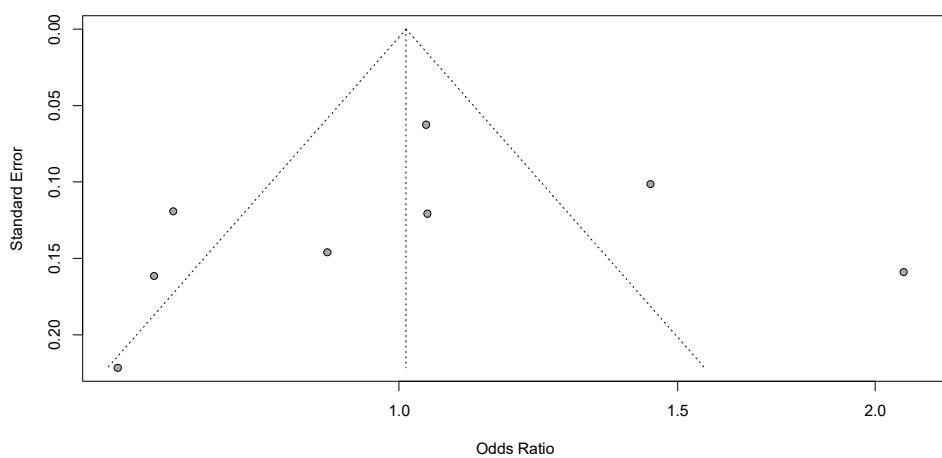
11. Wurden potentielle Interessenkonflikte adressiert?

JA  NEIN  UNKLAR  NICHT ANWENDBAR

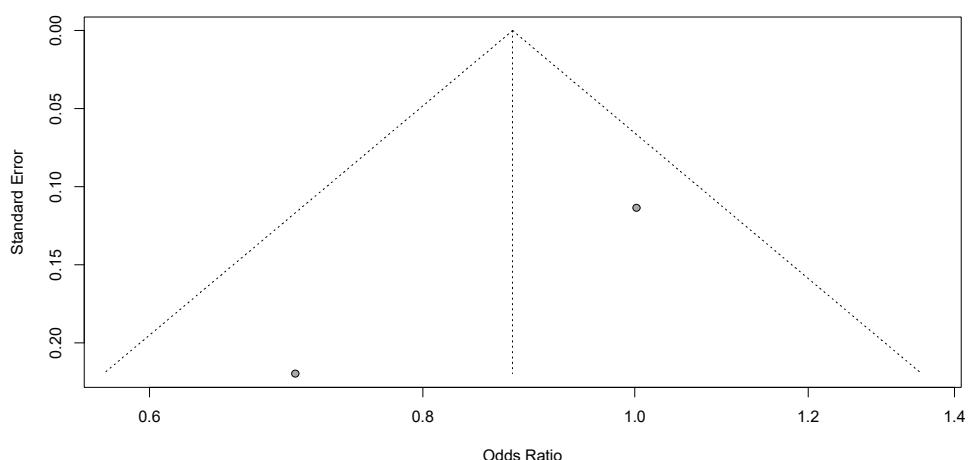
## Anhang 5: Funnel-Plots



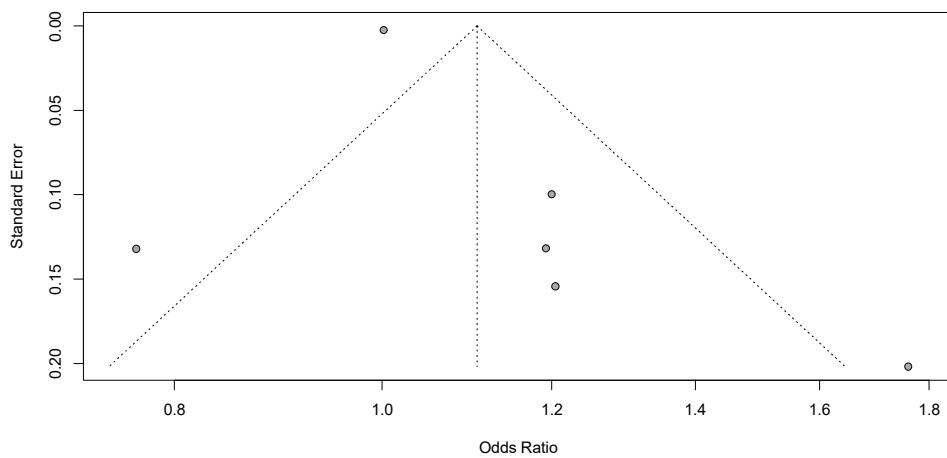
Anhang 5.1: Funnel-Plot für den SNP rs3865444; nur nordeuropäische Abstammung



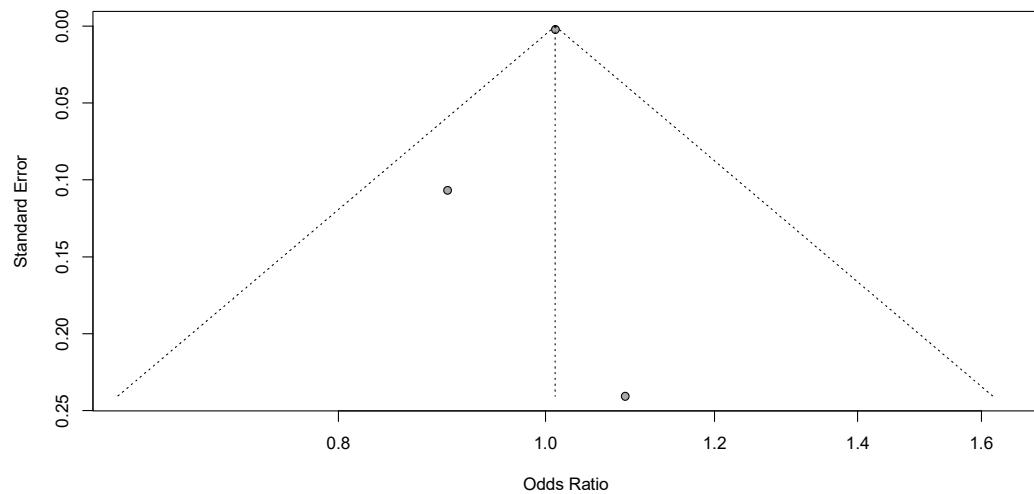
Anhang 5.2: Funnel-Plot für den SNP rs3865444; nur ostasiatische Abstammung



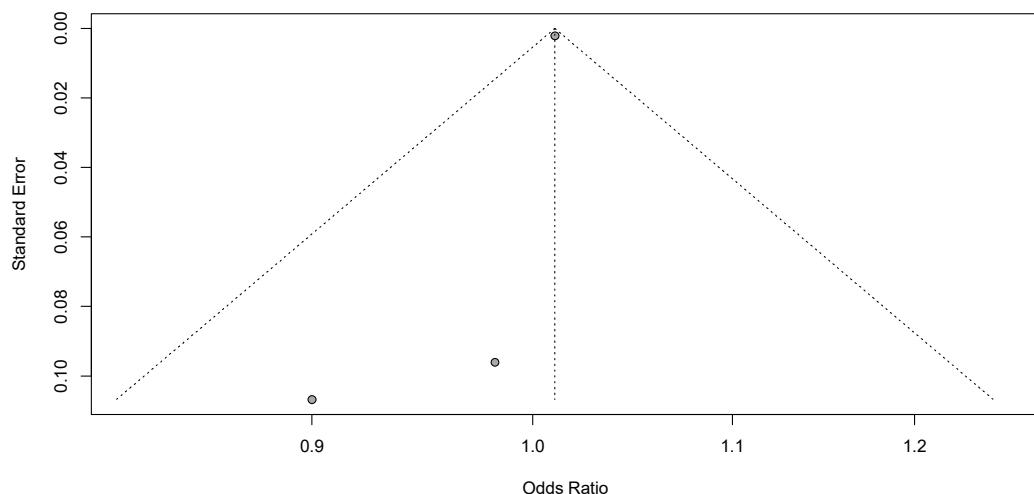
Anhang 5.3: Funnel-Plot für den SNP rs3865444; nur hispanische Abstammung



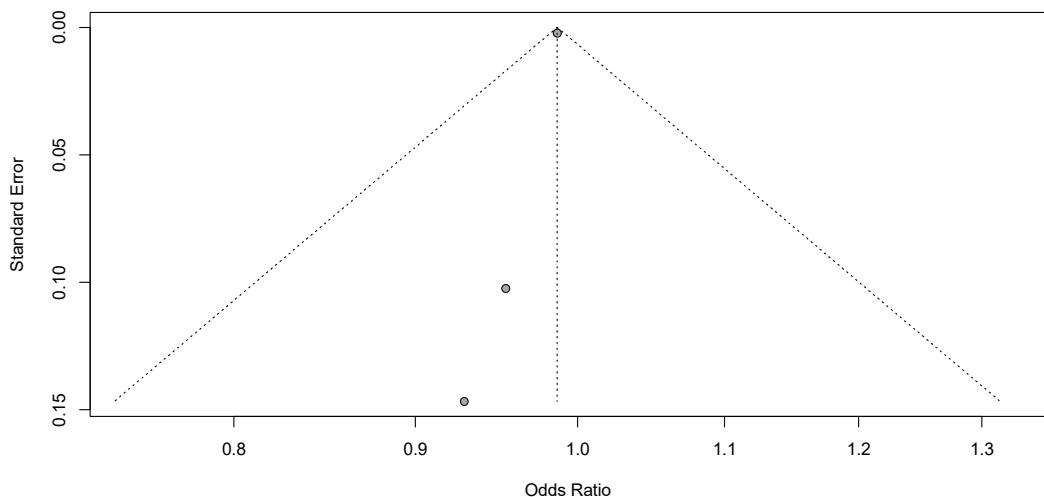
Anhang 5.4: Funnel-Plot für den SNP rs3826656; nach Einschluss aller ethnischen Abstammungsgruppen



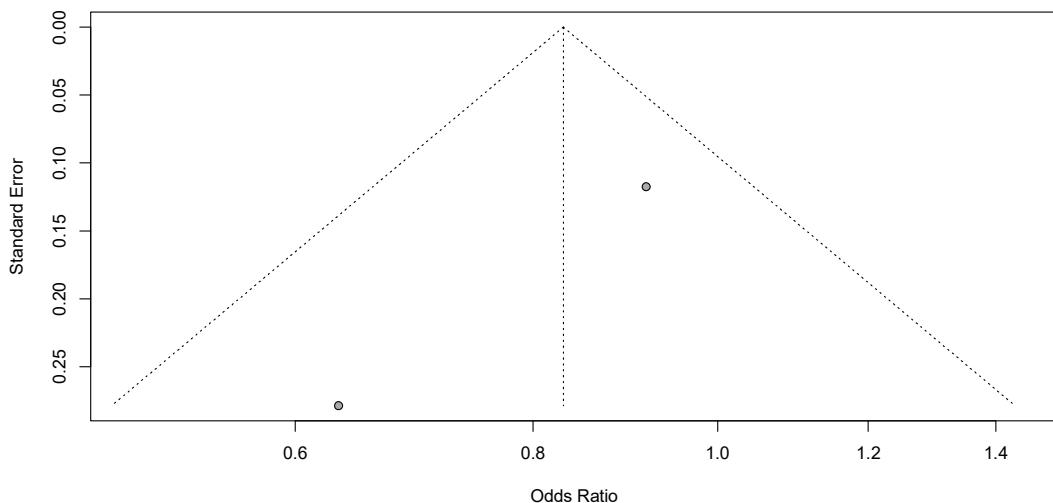
Anhang 5.5: Funnel-Plot für den SNP rs1697553; nach Einschluss aller ethnischen Abstammungsgruppen



Anhang 5.6: Funnel-Plot für den SNP rs2455069; nur nordeuropäische Abstammung



Anhang 5.7: Funnel-Plot für den SNP rs12459419; nach Einschluss aller ethnischen Abstammungsgruppen



Anhang 5.8: Funnel-Plot für den SNP rs35112940; nur nordeuropäische Abstammung

## Anhang 6: Egger's-Test Ergebnisse

SNP	Abstammungsgruppe	Egger's Test
rs3865444	Alle ethnischen Abstammungsgruppen	0.4156
rs3865444	Nordeuropäische Abstammung	0.1700
rs3865444	Ostasiatische Abstammung	0.7018
rs3865444	Hispanische Abstammung	0.0
rs3826656	Alle ethnischen Abstammungsgruppen	0.2744
rs3826656	Ostasiatische Abstammung	0.2273
rs1697553	Alle ethnischen Abstammungsgruppen	0.6825
rs2455069	Nordeuropäische Abstammung	0.3292
rs12459419	Alle ethnischen Abstammungsgruppen	0.0713
rs35112940	Nordeuropäische Abstammung	0.2551

## Anhang 7: Votum der Ethikkomission



UNIVERSITÄT ZU LÜBECK

Universität zu Lübeck · Ratzeburger Allee 160 · 23538 Lübeck

Herrn  
Prof. Dr. med. Lars Bertram  
Leiter Interdisziplinäre Plattform für Genomanalytik  
Institute für Neurogenetik und Kardiogenetik  
Maria-Goeppert-Straße 1 (MFC1) – 4.23  
23562 Lübeck

### Ethik-Kommission

Vorsitzender:  
Herr Prof. Dr. med. Alexander Katalinic  
Stellv. Vorsitzender:  
Herr Prof. Dr. med. Frank Gieseler

Ratzeburger Allee 160  
23538 Lübeck  
Sachbearbeitung: Frau Janine Erdmann  
Tel.: +49 451 3101 1008  
Fax: +49 451 3101 1024  
ethikkommission@uni-luebeck.de

Aktenzeichen: 19-392A  
Datum: 22. November 2019

The role of microRNA-RNA regulatory networks in neural circuits affected by Alzheimer's disease  
(MiRNetAD)

Sehr geehrter Herr Prof. Bertram,

mit Schreiben vom 16.11.2019 informierten Sie uns über das o.g. Studienvorhaben.  
Dabei analysieren Sie Biomaterial, das Ihnen von Kooperationspartnern aus Biobanken zur Verfügung gestellt wird. Sammlung, Lagerung und Weitergabe des Biomaterials ist dabei durch Ethikvoten (eingeholt von den Kooperationspartnern) abgedeckt.

Unter diesen Rahmenbedingungen nimmt die Ethik-Kommission das von Ihnen in Ihrem Anschreiben beschriebene Vorhaben im verkürzten Verfahren zustimmend zur Kenntnis.

Es lagen folgende Unterlagen vor:

- Ihr Anschreiben vom 16. November 2019
- Studienprotokoll in der Version 1.0 vom 15. November 2019.

Mit freundlichen Grüßen

  
Prof. Dr. med. Alexander Katalinic  
Vorsitzender

#### **Anhang 8: Top-SNP-Tabelle *trans* eQTL-Ergebnisse**

Die Top-SNP-Tabelle *trans* eQTL-Ergebnisse als Excel-Datei befindet sich in Form einer MS-Excel-Datei auf der beigelegten CD-ROM.

#### **Anhang 9: Resultate der *cis* eQTL-Analysen in der „*cis*“ CD33-Region**

Die gesamten Resultate der *cis* eQTL-Analysen in der „*cis*“ CD33-Region befinden sich in Form einer MS-Excel-Datei auf der beigelegten CD-ROM.

#### **Anhang 10: VEP-Outputs**

Die VEP-Outputs als Excel-Datei befinden sich in Form einer MS-Excel-Datei auf der beigelegten CD-ROM.

## 9 Danksagung

Mein herzlicher und großer Dank gilt Herrn Professor Dr. Lars Bertram, meinem Doktorvater, für das Überlassen des Themas und die exzellente Betreuung dieser Dissertation. Seine fachliche und menschliche Betreuung war die wichtigste Stütze für meine Doktorarbeit. Unermüdlich stand er mir mit viel Geduld zur Seite und steuerte aus seinem reichen Erfahrungsschatz viele Anregungen bei.

Weiterhin danke ich Olena Ohlei, die einen wichtigen Beitrag leistete, indem sie stets Verständnis für meine Fragestellungen zeigte und bei der Durchführung der Analysen mir jederzeit mit ihrem Rat zur Seite stand.

Mein ganz besonderer Dank gilt meiner Familie, die so viel für mich geopfert hat, um mir mein Studium & Lebensweg zu ermöglichen und der ich diese Arbeit widme. Denen, die noch leben und denen, die leider nicht mehr am Leben sind.

Tief verbunden und dankbar bin ich meinem Ehemann, Jan Paulsen, der dafür gesorgt hat, dass ich körperlich – aber vor allem seelisch diese intensive Zeit gut überstanden habe.

## 10 Lebenslauf