

Molekulargenetische Untersuchungen bei schizophrenen Störungen

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Anna Carina Koller

aus

Augsburg

Bonn 2025

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer: Prof. Dr. Markus M. Nöthen

Gutachter: Prof. Dr. Walter Witke

Tag der Promotion: 09.02.2026

Erscheinungsjahr: 2026

Abkürzungsverzeichnis

1 Einleitung.....	1
1.1 Schizophrene Störungen	1
1.1.1 Ätiologie und Epidemiologie.....	1
1.1.2 Diagnose und Therapie	2
1.2 Genetische Architektur der Schizophrenie.....	4
1.3 Strategien zur Identifizierung seltener genetischer Risikovarianten	7
2 Zielsetzung der Arbeit.....	11
3 Material & Methoden	12
3.1 Geräte	12
3.2 Chemikalien, Puffer und Lösungen	14
3.3 Ergänzende Materialien zur Analyse von Nukleinsäuren	14
3.4 Kommerzielle Systeme	14
3.5 Software und Datenbanken.....	15
3.6 Patienten- und Kontrollkollektive	16
3.7 Genomweite Genotypisierung.....	20
3.8 Ausschluss von Schizophrenie-assoziierten Kopienzahlvarianten	20
3.9 Exom-Sequenzierung	20
3.9.1 Probenvorbereitung	20
3.9.2 Sequenzierung	21
3.9.3 Bioinformatische Auswertung und Filtern der detektierten Varianten	21
3.10 Genomweite Analyse von Kopienzahlvarianten.....	24
3.10.1 Verwendete Algorithmen zur Detektion von Kopienzahlvarianten.....	24
3.10.2 Filtern der detektierten Kopienzahlvarianten.....	25
3.11 Experimentelle Verifikation	25
3.11.1 Verifikation der detektierten Einzelnukleotid-Varianten und Indels	25
3.11.2 Verifikation der detektierten Kopienzahlvarianten mittels qPCR.....	26
3.12 Priorisierung der detektierten Gene	27
3.12.1 Genetische Evidenz – Abgleich mit publizierten Studien	27
3.12.2 Genetische Evidenz – Gen-basierte Tests	28
3.12.3 Protein-Protein Interaktionsanalysen	28
3.12.4 Genexpressionsanalysen	30
4 Ergebnisse.....	31
4.1 Familien-Pilotstudie	32
4.2 CNV-Studie	46

5 Diskussion	66
5.1 Pilotstudie – Evidenz für die Relevanz der Chromatinmodifikation in der Pathogenese der Schizophrenie	66
5.2 CNV-Studie – Identifizierung exonischer Kopienzahlvarianten in einer großen Kohorte von SCZ multiplex Familien.....	71
5.3 Ausblick	73
6 Zusammenfassung.....	76
7 Literaturverzeichnis	79
8 Anhang.....	93

Abkürzungsverzeichnis

ASD	Autismus-Spektrum Störung
bp	Basenpaar
CADD	<i>Combined Annotation Dependent Depletion</i>
cDNA	<i>complementary DNA</i> , komplementäre DNA
chr	Chromosom
CN	<i>Copy Number</i> , Kopienzahl
CNV	<i>Copy Number Variant</i> , Kopienzahlvariante
CoNIFER	<i>Copy Number Inference From Exome Reads</i>
DNA	<i>deoxyribonucleic acid</i> , Desoxyribonukleinsäure
DSM	<i>Diagnostic and Statistical Manual of Mental Disorders</i> ; Diagnostisches und Statistisches Manual Psychischer Störungen
EVS	<i>Exome Variant Server</i>
ExAC	<i>Exome Aggregation Consortium</i>
GTE _x	<i>Genotype-Tissue Expression</i>
GWAS	Genomweite Assoziationsstudie
HIPPIE	<i>Human Integrated Protein-Protein Interaction rEference</i>
ICD	<i>International Statistical Classification of Diseases and Related Health Problems</i> ; Internationale Klassifikation der Krankheiten und verwandter Gesundheitsprobleme
ID	Intelligenzminderung
Indel	Insertion und Deletion
iPSC	<i>induced pluripotent stem cell</i> ; Induzierte pluripotente Stammzelle
kb	Kilobase
MAF	<i>minor allele frequency</i> , Frequenz des selteneren Allels
Mb	Megabase
mDA	mesencephale dopaminerge Neurone
min	Minute
μl	Mikroliter
μM	Mikromolar
ng	Nanogramm
OR	<i>Odds Ratio</i>
PPI	Protein-Protein Interaktion
NFE	<i>non-Finnish European</i> , nicht-finnischer Europäer
NGS	<i>Next Generation Sequencing</i>
PCR	<i>polymerase chain reaction</i> , Polymerasekettenreaktion

PGC	<i>Psychiatric Genomics Consortium</i>
qPCR	quantitative Polymerase-Kettenreaktion
RNA	<i>ribonucleic acid</i> , Ribonukleinsäure
s	Sekunde
SCZ	Schizophrenie
smMIP	<i>single molecule molecular inversion probe</i>
SNP	<i>single nucleotide polymorphism</i> , Einzelnukleotid-Polymorphismus
SNV	<i>single nucleotide variant</i> , Einzelnukleotid-Variante
UTR	untranslatierte Region
V	Volt
WGA	<i>whole genome amplification</i> , Genomamplifikation
XHMM	<i>eXome-Hidden Markov Model</i>

1 Einleitung

1.1 Schizophrene Störungen

1.1.1 Ätiologie und Epidemiologie

Schizophrenie (SCZ) ist eine schwere neuropsychiatrische Erkrankung, die unter anderem zu einer Störung der Wahrnehmung, des Denkens und des Bewusstseins bzw. des Ich-Erlebens führen kann (Rey, 2011). Die Wahrscheinlichkeit, im Laufe des Lebens an SCZ zu erkranken (Lebenszeitrisiko), liegt bei ca. 1% (Perälä et al., 2007) und ist in allen Ländern und Kulturen etwa gleich (Jablensky et al., 1992). Jedes Jahr erkranken in Deutschland durchschnittlich pro 100.000 Menschen etwa 15 neu an SCZ, dies entspricht ca. 12.000 Neuerkrankungen pro Jahr (Rössler, 2011). Die Inzidenz ist bei Männern etwas höher (~1,4-fach) als bei Frauen (Aleman et al., 2003; van der Werf et al., 2014).

SCZ ist eine multifaktorielle Erkrankung. Das bedeutet, dass bei der Entstehung der SCZ sowohl genetische als auch Umgebungsfaktoren eine Rolle spielen. Die Heritabilität (der Anteil der phänotypischen Varianz, der durch genetische Faktoren erklärt wird) der SCZ wird auf etwa 60-80% geschätzt (Hilker et al., 2018; Wray & Gottesman, 2012).

Zu den Umgebungsfaktoren zählen unter anderem Cannabiskonsum, das Aufwachsen in Großstädten, Kindheitstraumata (van Os et al., 2010) oder Infektionen der Mutter während der Schwangerschaft (z.B. mit *Toxoplasma gondii*) (Löhrs & Hasan, 2019). Das Vorhandensein dieser Faktoren führt jedoch jeweils nur zu einer relativ geringen Erhöhung des Erkrankungsrisikos (*Odds Ratio*; OR). So beträgt das OR für Cannabiskonsum ~1,3 (Vaucher et al., 2018), für das Aufwachsen in Großstädten etwa 2,5 (G. Lewis et al., 1992; McGrath et al., 2001; Os et al., 2001), für Kindheitstraumata ~4,5 (Heins et al., 2011) und für eine Infektion der Mutter mit *Toxoplasma gondii* während der Schwangerschaft ~2,6 (Brown et al., 2005). Ein weiterer Risikofaktor scheint außerdem ein erhöhtes Alter des Vaters zum Zeitpunkt der Zeugung zu sein (Malaspina D et al., 2001; Miller et al., 2011).

Eine positive Familienanamnese für SCZ ist bisher der größte bekannte Risikofaktor für eine Erkrankung mit SCZ. Dabei korreliert die Erkrankungswahrscheinlichkeit einer nicht betroffenen Person mit der Anzahl Erkrankter in der Familie und dem Verwandtschaftsgrad zum Indexpatienten. Im Vergleich zur Allgemeinbevölkerung ist das Risiko an SCZ zu erkranken leicht erhöht (ca. 3%), wenn ein zweitgradig Verwandter (z.B. Onkel oder Tante) erkrankt ist. Bei erstgradig Verwandten (Geschwister, Elternteil oder Kind) eines Betroffenen

beträgt das Risiko ca. 10%. Sind beide Elternteile betroffen, steigt das Risiko auf ca. 45% (Faraone et al., 1999; Harper, 2004).

1.1.2 Diagnose und Therapie

Die Diagnose des Krankheitsbildes SCZ wird klinisch, basierend auf standardisierten Kriterien und speziellen Fragebögen, gestellt. Am häufigsten werden dabei derzeit die Definitionskriterien des Diagnostischen und Statistischen Handbuchs Psychischer Störungen (*Diagnostic and Statistical Manual of Mental Disorders*; DSM)-IV der Amerikanischen Psychiatrischen Vereinigung (American Psychiatric Association, 1994) oder des Klassifikationssystems ICD-10 der Weltgesundheitsorganisation (*International Statistical Classification of Diseases and Related Health Problems*) (Dilling et al., 2015) verwendet.

Nach den DSM-IV Kriterien (American Psychiatric Association, 1994) müssen für die Diagnose SCZ mindestens zwei der folgenden Symptome für einen Monat oder länger auftreten:

1. Wahn
2. Halluzinationen
3. Desorganisiertes Sprechen
4. Desorganisiertes oder katatonisches Verhalten
5. Negative Symptome (wie z.B. Affektverflachung oder Sprachverarmung)

Weiterhin müssen vor Diagnosestellung andere Ursachen der Symptome wie Substanzeinfluss oder medizinische Krankheitsfaktoren (z.B. Schädel-Hirntrauma, Hirntumor oder Stoffwechselerkrankungen) ausgeschlossen werden.

Meist tritt die Erkrankung im frühen Erwachsenenalter auf, wobei Frauen im Durchschnitt ein paar Jahre später erkranken als Männer (Aleman et al., 2003; Jablensky et al., 1992; van der Werf et al., 2014). Das Ersterkrankungsalter liegt bei Männern bei etwa 20-24 Jahren, bei Frauen bei ca. 25-29 Jahren. Ein weiterer Unterschied besteht darin, dass bei Frauen ein zweiter Erkrankungspunkt im Alter von etwa 45-49 Jahren auftritt (Rey, 2011). Es wird vermutet, dass Östrogen eine protektive Wirkung besitzt, welche nach der Menopause verloren geht und so zu dem zweiten Erkrankungspunkt führt (Kirkbride et al., 2012).

Die Lebenserwartung von Patienten mit SCZ ist im Vergleich zur Allgemeinbevölkerung um bis zu 25 Jahre verkürzt (Hjorthøj et al., 2017; Nordentoft et al., 2013; Tiihonen et al., 2009). Dies liegt unter anderem an somatischen Komorbiditäten wie Herz-Kreislauf-Erkrankungen, Diabetes mellitus, Hyperlipidämie oder Adipositas (Pérez-Piñar et al., 2016). Weiterhin ist die

Suizidrate bei Patienten mit SCZ im Vergleich zur Allgemeinbevölkerung ~20-fach erhöht (Nordentoft et al., 2004).

Mindestens 60-70% der Patienten mit SCZ benötigen eine längerfristige, zum Teil lebenslange Behandlung. Diese setzt sich meist aus einer medikamentösen Behandlung (z.B. Neuroleptika), psychologisch-psychotherapeutischer Therapie, Familientherapie und Sozialtherapie (Rehabilitation) zusammen. Etwa 30% der Patienten erholen sich nach einer einmaligen akuten Phase wieder vollständig (Rey, 2011).

1.2 Genetische Architektur der Schizophrenie

SCZ ist eine genetisch komplexe, polygene Erkrankung, an deren Entstehung ein großes frequenzüberspannendes Spektrum an genetischen Varianten beteiligt ist (Abbildung 1). Dieses beinhaltet sowohl häufigere Varianten mit einer niedrigeren Penetranz, als auch seltenere Varianten mit einer höheren Penetranz (McCarthy et al., 2008; Sullivan et al., 2012).

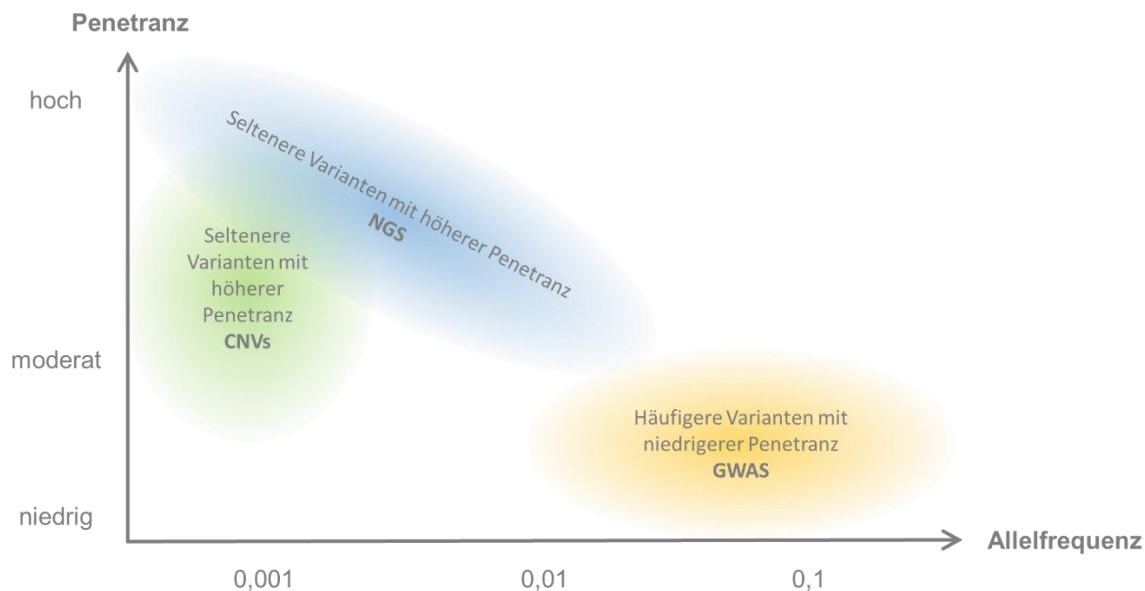


Abbildung 1 Genetische Architektur der Schizophrenie

Bereits identifizierte genetische Risikofaktoren für Schizophrenie sind einerseits Varianten mit höherer Allelfrequenz in der Allgemeinbevölkerung und niedrigerer Penetranz, welche in der Regel mittels genomweiter Assoziationsstudien (GWAS) identifiziert werden, und andererseits Varianten mit sehr niedriger Allelfrequenz in der Allgemeinbevölkerung und höherer Penetranz. Letztere sind vor allem Kopienzahlvarianten (CNVs) sowie Einzelnukleotid-Varianten und kleine Insertionen und Deletionen, welche mittels *Next Generation Sequencing* (NGS) Technologien identifiziert werden. Die Abbildung wurde modifiziert nach McCarthy et al. (2008) und Sullivan et al. (2012).

Häufigere Varianten mit niedrigerer Penetranz

Häufigere Varianten mit einer niedrigeren Penetranz, sogenannte Einzelnukleotidpolymorphismen (*single nucleotide polymorphisms*; SNPs), die mit einer Erkrankung assoziiert sind, können mittels Genomweiter Assoziationsstudien (GWAS) identifiziert werden. Der Fokus in bisher publizierten Studien lag dabei auf SNPs mit Allelfrequenzen $\geq 5\%$ in der Allgemeinbevölkerung. In einer GWAS werden die Allelfrequenzen von SNPs zwischen zwei Kohorten (z.B. einer Patienten- und einer Kontrollkohorte) verglichen. Dazu werden die Probanden mittels DNA-Microarray genomweit genotypisiert. Diese DNA-Microarrays enthalten hunderttausende SNPs, die über das gesamte Genom verteilt sind. Anschließend wird mittels statistischer Analysen nach einer Assoziation zwischen SNPs und dem untersuchten Phänotyp gesucht. Bei einem Signifikanzniveau von $p < 5 \times 10^{-8}$

spricht man von einer genomweit signifikanten Assoziation (Barsh et al., 2012) und die assoziierten SNPs gelten als Phänotyp-relevant.

In der größten bisher publizierten SCZ GWAS (76.755 Patienten mit SCZ und 243.649 Kontrollen) wurden häufige Varianten in 287 Loci identifiziert, welche genomweit-signifikant mit SCZ assoziiert waren (Trubetskoy et al., 2022). Die assoziierten Varianten in dieser Studie waren in Genen angereichert, die in den Neuronen des zentralen Nervensystems exprimiert sind. Frühere große SCZ GWAS (Pardiñas et al., 2018, 40.675 Patienten und 64.643 Kontrollen; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014, 36.989 Patienten und 113.075 Kontrollen) konnten außerdem eine Anreicherung der genetischen Assoziationen in Genen, die im Gehirn und in Immunsystem-relevanten Geweben exprimiert werden (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), sowie in Genen, die unter starkem Selektionsdruck stehen (Pardiñas et al., 2018), nachweisen. Weiterhin konnte eine Assoziation mit dem Gen *DRD2*, welches den Dopamin-Rezeptor D2 kodiert, identifiziert werden (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Dopamin-D2-Rezeptoren gelten als wichtigster Angriffspunkt für klassische Neuroleptika (Rey, 2011). Somit konnte gezeigt werden, dass genetische Studien ein vielversprechender Ansatz zur Entwicklung neuer Therapiestrategien sein können.

Der Anteil an der phänotypischen Varianz, welcher durch häufige Varianten erklärt werden kann, wird derzeit auf etwa 20-30% geschätzt (Pardiñas et al., 2018; Trubetskoy et al., 2022). Eine der Herausforderungen von GWAS ist, dass ein großer Teil der identifizierten SNPs intergenisch lokalisiert ist und somit eine Zuordnung zu einem potentiellen Suszeptibilitätsgen schwierig ist. Bei der Exom-Sequenzierung hingegen werden pathogene Varianten in kodierenden Bereichen von Genen identifiziert.

Seltenere Varianten mit höherer Penetranz

Zu den selteneren genetischen Risikofaktoren mit höherer Penetranz zählen unter anderem Kopienzahlvarianten (*copy number variants*; CNVs). CNVs sind unbalancierte, strukturelle chromosomale Veränderungen, die zu einem Verlust (Deletion) oder einem Zugewinn (Duplikation) von genetischem Material im Vergleich zur Referenzsequenz führen. Die derzeit am häufigsten verwendete Methode zur Identifizierung submikroskopischer CNVs sind Microarray-basierte Analysen, mit denen das gesamte Genom auf CNVs untersucht werden kann (Feuk et al., 2006). Zunehmend werden kleinere CNVs auch über Exom- bzw. Genomsequenzierungsdaten nachgewiesen (Khan et al., 2018; Szatkiewicz et al., 2020).

In der bisher größten genomweiten SCZ CNV-Analyse, in der 21.094 Patienten mit SCZ und 20.227 Kontrollen untersucht wurden, wurden genomweit signifikant mit SCZ assoziierte CNVs in acht genomischen Loci identifiziert: Deletionen in den chromosomalen Regionen 1q21.1,

2p16.3 (*NRXN1*), 3q29, 15q13.3, 16p11.2 (distal) und 22q11.2 sowie Duplikationen in den chromosomalen Regionen 1q21.1, 7q11.23 und 16p11.2 (proximal). Die ORs dieser CNVs betrugen zwischen 3,8 (Deletionen und Duplikationen in 1q21.1) und 67,7 (Deletionen in 22q11.2). Insgesamt waren in 1,4% der untersuchten Patienten mit SCZ CNVs in einem der acht genomweit signifikanten Loci nachweisbar (Marshall et al., 2017).

Aufgrund technischer Limitationen fokussierten sich die meisten bisher durchgeführten Microarray-basierten CNV-Studien auf Deletionen und Duplikationen mit einer Mindestgröße von 100 kb (Degenhardt et al., 2013; The International Schizophrenia Consortium, 2008; Vacic et al., 2011). Infolgedessen konnten bisher nur relativ große CNVs, die sich über mehrere Gene erstreckten, detektiert werden. Dies erschwert die Identifizierung des Phänotyp-relevanten Gens.

Durch die Einführung von *Next Generation Sequencing* (NGS) Technologien rückten seltenere, höher penetrante Varianten, welche nicht mittels GWAS oder Microarray-basierter genomweiter CNV-Analysen nachweisbar sind, in den Fokus. Diese Technologien ermöglichten die kosten- und zeiteffiziente Sequenzierung ganzer Exome in größeren Kohorten und außerdem die Identifizierung proteinkodierender Varianten auf Einzelbasenebene und kleinerer (<100 kb) CNVs, welche einzelne Gene direkt betreffen. Dies erleichtert die Identifizierung neuer krankheitsrelevanter Gene.

Die exomweite Sequenzierung ermöglicht die Identifizierung neuer Kandidatengene und die Generierung weiterer genetischer Evidenz für bereits vorbeschriebene Gene. Die Identifizierung potentiell kausaler Varianten mit höherer Penetranz ist darüber hinaus wichtig für funktionelle Analysen wie z.B. Zebrafisch-Studien oder die Untersuchung induzierter pluripotenter Stammzellen (*induced pluripotent stem cells*; iPSCs) zur Überprüfung der phänotypischen Relevanz des betroffenen Kandidatengens. Seltene Varianten haben im Vergleich zu häufigeren Varianten oft auch ausgeprägtere funktionelle Folgen und können somit im Tiermodell oder in iPSCs einen besseren Einblick in den Krankheitsmechanismus bieten (Raychaudhuri, 2011). Zusätzlich ist bei selteneren Varianten mit größerem Effekt auch eine Translation in die Klinik (z.B. genetische Testung im diagnostischen Kontext) wahrscheinlicher.

Aufgrund der großen Anzahl von pathogenen Varianten pro Exom (~20.000; Ng et al. 2009) ist es jedoch weiterhin schwierig, zwischen Phänotyp-neutralen und Phänotyp-relevanten Varianten zu unterscheiden. Im Folgenden werden verschiedene Strategien zur Identifizierung seltener genetischer Risikovarianten mittels Exom-Sequenzierung vorgestellt.

1.3 Strategien zur Identifizierung seltener genetischer Risikovarianten

In Abbildung 2 sind die wesentlichen Strategien zum Nachweis von selteneren genetischen Varianten graphisch dargestellt. Diese werden im Folgenden einzeln erläutert.

Trio Sequenzierung

Die ersten Studien, welche die Exom-Sequenzierung zur Identifizierung seltener genetischer Risikovarianten für die SCZ verwendeten, fokussierten sich auf *de novo* Mutationen mittels Trio Sequenzierung (Girard et al., 2011; Xu et al., 2012, 2011). *De novo* Mutationen werden nicht von den Eltern vererbt, sondern treten in Folge einer Mutation in der mütterlichen oder väterlichen Keimzelle bei der Indexperson auf (Abbildung 2A) und sind aus drei Hauptgründen von besonderem Interesse für die Erforschung der genetischen Grundlage der SCZ:

(i) Trotz einer starken Selektion gegen SCZ assoziierte Varianten durch eine verminderte Anzahl Nachkommen bei Patienten mit SCZ bleibt die Inzidenz der SCZ über die Zeit gleich. Dies führt zu der Annahme, dass genetische Varianten, die aus dem Genpool selektiert werden, durch *de novo* Mutationen ersetzt werden (Power et al., 2013). Daher ist zunächst jedes Gen, bei welchem eine *de novo* Mutation in einem sporadischen Fall mit SCZ identifiziert wurde, ein potentielles Kandidatengen.

(ii) Studien konnten zeigen, dass ein erhöhtes väterliches Alter zum Zeitpunkt der Zeugung ein Risikofaktor für eine Erkrankung mit SCZ ist (Malaspina D et al., 2001; Miller et al., 2011). Die Anzahl der *de novo* Mutationen bei einem Nachkommen steigt ebenfalls mit zunehmendem väterlichen Alter (Kong et al., 2012).

(iii) Wegen der geringen Anzahl von durchschnittlich 0,75-1 *de novo* Mutation pro Exom (Fromer et al., 2014; Gulsuner et al., 2013; Xu et al., 2011) ist die Interpretation der Daten weniger komplex als die Analyse aller genetischen Varianten im Exom eines Individuums.

Aus diesen Gründen sind Gene, in denen eine *de novo* Mutation bei einem Patienten mit SCZ beschrieben wurde, für die Forschung interessante Kandidatengene. Bisher wurden > 10 Studien, in denen Patienten mit SCZ und deren nicht erkrankte Eltern exomweit (Ambalavanan et al., 2016; Fromer et al., 2014; Girard et al., 2011; Guipponi et al., 2014; Gulsuner et al., 2013; Kranz et al., 2015; McCarthy et al., 2014; Smedemark-Margulies et al., 2016; Suddaby et al., 2019; Takata et al., 2014; Wang et al., 2015; Xu et al., 2011, 2012; Balakrishna & Curtis, 2020; Howrigan et al., 2020; Rees et al., 2020) oder genomweit (Tang et al., 2017) sequenziert wurden, publiziert. Insgesamt wurden über 4.000 Trios bzw. Quads analysiert und *de novo* Mutationen in mehr als 2.700 Genen identifiziert.

Die publizierten SCZ *de novo* Studien zeigten unter anderem auch, dass nicht-synonyme *de novo* Mutationen in Patienten mit SCZ im Vergleich zu gesunden Probanden angereichert waren (Girard et al., 2011; Xu et al., 2012). Einige *de novo* Studien wiesen außerdem auf eine Krankheitsrelevanz von Genen hin, die in der fetalen Gehirnentwicklung exprimiert werden (Gulsuner et al., 2013; Wang et al., 2015; Xu et al., 2012). Andere Studien lieferten Hinweise für eine Beteiligung chromatinregulierender Gene in der Pathogenese der SCZ (McCarthy et al., 2014; Takata et al., 2014).

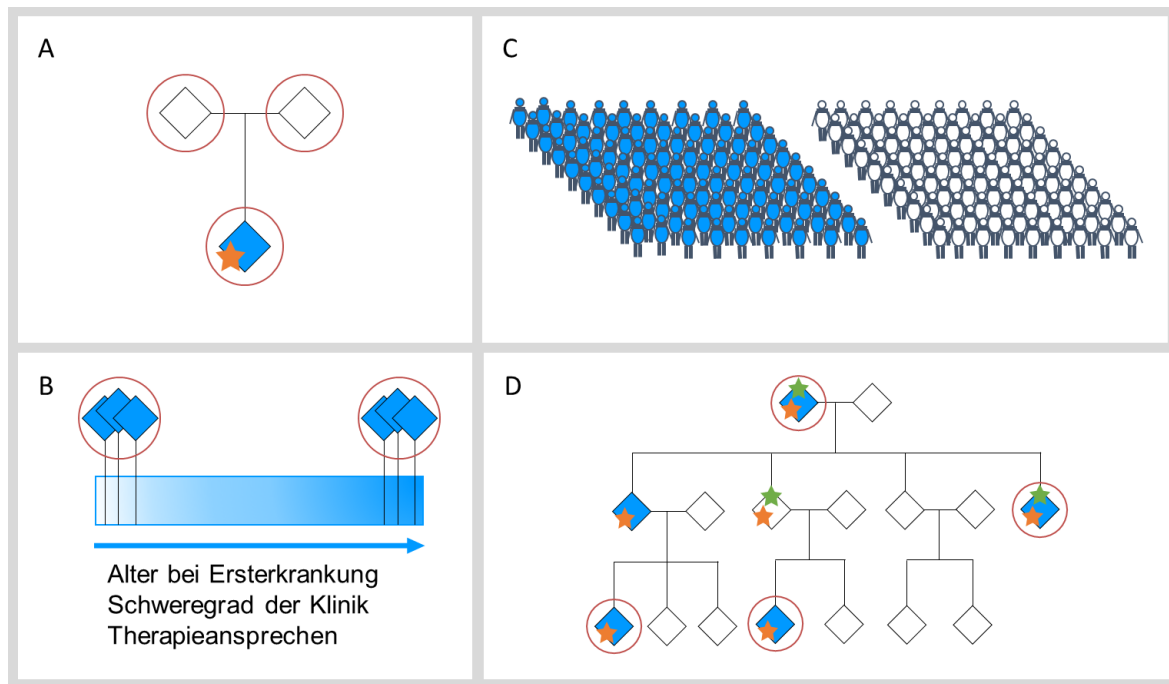


Abbildung 2 Mögliche Strategien zur Identifizierung seltener genetischer Risikovarianten mittels Exom-Sequenzierung. A: Trio Sequenzierung; B: Sequenzierung von Patienten mit maximal bzw. minimal Ausprägung eines klinischen Phänotyps; C: Fall-Kontroll-Studie; D: Familien-basierte Studie. Blaue Symbole: Patienten; Weiße Symbole: Gesunde Probanden. Verschiedenfarbige Sterne stellen pathogene Varianten in unterschiedlichen Genen dar (für eine übersichtlichere Darstellung nicht in B und C abgebildet, jedoch werden auch bei diesen Studiendesigns bei einer polygenen Erkrankung pathogene Varianten in zahlreichen Genen erwartet); Rote Kreise markieren für die Exom-Sequenzierung ausgewählte Individuen. Die Abbildung wurde modifiziert nach Ludwig et al. (2019).

Sequenzierung von Patienten mit maximal bzw. minimal Ausprägung eines klinischen Phänotyps

Eine zweite mögliche Methode zur Identifizierung von selteneren Varianten mit höherer Penetranz ist die Sequenzierung von Patienten mit maximaler bzw. minimaler Ausprägung eines klinischen Phänotyps (zum Beispiel Patienten mit sehr jungem Erkrankungsalter, besonders schwerem klinischen Verlauf oder schlechtem Therapieansprechen; Abbildung 2B). Verschiedene Studien zeigten eine erhöhte genetische Belastung bei diesen Patienten (Ahn et al., 2014; Joobar et al., 2005; Nicolson et al., 2003).

Die exomweite Sequenzierung von Patienten mit einer erwarteten, erhöhten genetischen Belastung führt zu einer Anreicherung genetischer Risikovarianten in der Kohorte. Weiterhin bedingt die Homogenisierung der Untersuchungskohorte eine Verringerung der klinischen und der genetischen Heterogenität. Dadurch werden pathogene Varianten in der Kohorte angereichert. Weiterhin wird auch die Wahrscheinlichkeit erhöht, Varianten in denselben krankheitsrelevanten Genen oder Stoffwechselwegen zu finden.

Fall-Kontroll-Studie

Eine weitere Möglichkeit zur Identifizierung seltener Varianten ist die Durchführung von Fall-Kontroll-Studien (Abbildung 2C).

Purcell et al. (2014) analysierten die Exome von 2.536 Patienten mit SCZ und 2.543 Kontrollen und beschrieben eine signifikant erhöhte polygene Last von hauptsächlich sehr seltenen, proteintrunkierenden Varianten bei Patienten mit SCZ im Vergleich zu Kontrollen. Dies konnte auch in einer nachfolgenden Studie mit Exomen von 12.332 Individuen (4.946 Patienten mit SCZ, 6.242 gesunden Kontrollen und 1.144 Patienten mit anderen psychiatrischen Erkrankungen) bestätigt werden (Genovese et al., 2016). In der größten bisher publizierten Exom-Sequenzierungsstudie, einer Meta-Analyse mit 24.248 Patienten mit SCZ und 97.322 Kontrollen, konnten 10 Gene mit einer exomweit signifikanten Assoziation ($p < 2,14 \times 10^{-6}$) von sehr seltenen pathogenen Varianten mit SCZ identifiziert werden (Singh et al., 2022).

Aufgrund der hohen Anzahl an detektierten Varianten pro Exom ist die Unterscheidung von Phänotyp-neutralen und Phänotyp-relevanten Genen schwierig. Wahrscheinlich bedingt durch die klinische und genetische Heterogenität der SCZ konnten bisher kaum rekurrente/identische Mutationen in demselben Gen identifiziert werden. Zu den wenigen Ausnahmen gehören die pathogenen Varianten c.3182_3184delAAG (ENST00000409080) in dem Gen *ITGA6* (Ambalavanan et al., 2016; Xu et al., 2012), c.419C>T (ENST00000303391) in dem Gen *MECP2* (Cohen et al., 2002; Howrigan et al., 2020) und c.4582-2delAG (ENST00000262519) in dem Gen *SETD1A* (Guipponi et al., 2014; Takata et al., 2014; Singh et al., 2016; Kummeling et al., 2021). Es konnte jedoch eine Anreicherung seltener Varianten

in bestimmten Stoffwechselwegen festgestellt werden. Die angereicherten Varianten waren vor allem in Gehirn-exprimierten Genen, insbesondere in Genen, die in Neuronen exprimiert werden, und in Genen mit verschiedenen molekularen Funktionen, die die Bildung, Struktur und Funktion von Synapsen betreffen (Genovese et al., 2016; Purcell et al., 2014; Singh et al., 2022).

Familien-basierte Studie

Ein weiterer vielversprechender Ansatz ist die Analyse von sogenannten multiplex Familien. In diesen sind über mehrere Generationen hinweg Familienmitglieder erkrankt (Abbildung 2D). In multiplex Familien ist es wahrscheinlich, dass mindestens ein genetischer Risikofaktor mit großem Effekt zur Krankheitsentstehung beiträgt und innerhalb der Familie mit der Erkrankung co-segregiert. Die Fokussierung auf Varianten, die alle erkrankten Familienmitglieder gemeinsam haben, ermöglicht es, die Anzahl der Kandidatengene zu reduzieren und gleichzeitig die Wahrscheinlichkeit zu erhöhen, dass die identifizierten Gene für den Phänotyp relevant sind. Die Rekrutierung solcher multiplex Familien ist eine Herausforderung. Einerseits, da sie selten sind und andererseits, da häufig nicht alle Familienmitglieder für eine Studie zur Verfügung stehen. Oft sind die Angehörigen über verschiedene Orte verteilt, oder die Erkrankung wird innerhalb der Familie aus z.B. Angst vor Stigmatisierung nicht thematisiert. Zu Beginn der vorliegenden Arbeit war nur eine Studie publiziert, in der SCZ multiplex Familien mit Fokus auf seltenere und höher penetrante Einzelnukleotid-Varianten (SNVs) und kleine Insertionen und Deletionen exomweit sequenziert wurden (Timms et al., 2013). In dieser Studie wurden fünf Familien analysiert, in denen jeweils über zwei Generationen mindestens vier Individuen an SCZ, schizoaffektiver Störung oder einer nicht spezifizierten Psychose erkrankt waren. Es wurden insgesamt 22 Protein-verändernde Varianten in 21 Genen identifiziert, die bei allen erkrankten Familienmitgliedern innerhalb eines Stammbaums auftraten (Timms et al., 2013).

2 Zielsetzung der Arbeit

Schizophrenie ist eine schwere psychiatrische Erkrankung mit einer geschätzten Heritabilität von 60-80%, an deren Entstehung genetische Varianten und Umgebungsfaktoren beteiligt sind. Die der Erkrankung zugrundeliegenden Pathomechanismen sind bisher nur teilweise aufgeklärt. Zu Beginn der vorliegenden Arbeit hatte die Mehrzahl der molekulargenetischen Forschungsarbeiten die Identifizierung häufiger Varianten zum Ziel. Die bisher identifizierten krankheitsrelevanten Varianten erklären etwa 20-30% der geschätzten Heritabilität. Folglich wurde ein großer Teil der genetischen Risikofaktoren noch nicht entdeckt. Neuere *Next Generation Sequencing* Technologien wie z.B. die Exom-Sequenzierung ermöglichen die Identifizierung seltenerer, höher penetranter Varianten, welche mit früheren Genotypisierungsmethoden nicht identifizierbar waren.

Es wird erwartet, dass die Aufklärung der genetischen Architektur der Schizophrenie und das damit einhergehende bessere Verständnis der krankheitsrelevanten biologischen Mechanismen wesentlich zur Entwicklung von Strategien für Diagnostik, Therapie und Prävention beitragen wird.

Ziel dieser Arbeit war es, mit einem breiten Spektrum an Methoden zur Aufklärung der genetischen Architektur der Schizophrenie beizutragen. Der Fokus der Arbeit lag dabei auf der Analyse seltener Varianten und der Identifizierung neuer Kandidatengene in dem aktuell weltweit größten Kollektiv von Schizophrenie multiplex Familien ($n = 54$) mittels Exom-Sequenzierung. In diesen Familien waren mindestens drei Mitglieder von einer Schizophrenie oder einer anderen psychiatrischen Erkrankung betroffen.

3 Material & Methoden

3.1 Geräte

Autoklav

- Systec D-150, Systec GmbH
- Varioklav® 135 S Dampfsterilisator, H+P Labortechnik GmbH

DNA-Bank/Verwaltung

- 2DBarcoded 0,5ml Sample Storage Tube, FluidX Ltd.
- 2D CYPHER Tubes, Thermo Fisher Scientific GmbH
- 2D CYPHER 1,2ml Cluster Tube Racks, Thermo Fisher Scientific GmbH
- LidLock low base rack96, FluidX Ltd.
- TPE Septum Cap, FluidX Ltd.
- Impression™ Whole Rack 2D & 1D Code Scanner, FluidX Ltd.
- SmartScan Solo™ 2D Barcode Reader, Thermo Fisher Scientific GmbH
- SmartScan 96 2D Barcode Reader, Thermo Fisher Scientific GmbH
- XCAP SemiAutomatic Septum Cap Sealer for Sample Storage Tubes, FluidX Ltd.
- XDC96 Automatic Whole Rack Septum Capping & Decapping Systems, FluidX Ltd.

DNA-Fragmentierung

- Bioruptor®, Diagenode

Eismaschine

- AF80, Scotsman® Ice Systems

Gel-Elektrophoresekammern und Spannungsgeräte

- PowerPac Basic, BioRad Laboratories GmbH
- Sub-Cell Model 96, BioRad Laboratories GmbH
- TapeStation 2200, Agilent Technologies, Inc.
- WIDE MINI SUB CELL® GT, BioRad Laboratories GmbH

Gel-Dokumentation

- GelDoc™ XR System, BioRad Laboratories GmbH

Genotypisierungs- und Sequenzierungssysteme

- 3130xl Genetic Analyzer, Life Technologies GmbH
- cBot, Illumina
- HiSeq 2500, Illumina
- MassARRAY® System, Agena Bioscience

Nanodispenser

- MassARRAY™ Nanodispenser, SAMSUNG Techwin Co. Ltd. für Sequenom® GmbH

Nukleinsäure-Messgeräte

- NanoDrop® 1000 Spectrophotometer, Thermo Fisher Scientific
- NanoDrop® 8000 Spectrophotometer, Thermo Fisher Scientific
- Qubit 2.0 Fluorometer, Invitrogen by Life Technologies

Pipettierhilfen

- Cell Mate II, Matrix Technologies Corp., Thermo Fisher Scientific GmbH
- e-line Multistep Pipette, Sartorius AG
- Eppendorf Research Plus®, verschiedene Größen, Eppendorf GmbH
- Transferpette® S8, verschiedene Größen, BRAND GmbH & Co. KG

Pipettierroboter

- Biomek® NX MC Laboratory Automation Workstation, Beckman Coulter GmbH
- Biomek® NX S8 Laboratory Automation Workstation, Beckman Coulter GmbH

Reinstwasser-System

- Milli-Q A10 Synthesis, Merck KGaA

Schüttel- und Mischgeräte

- High-Speed Microplate Shaker, Illumina
- Magnetrührer ARE Aluminum Hot Plate Stirrer, Velp scientifica
- Plattenschüttler MS 3 digital, IKA®-Werke GmbH & CO. KG
- Überkopfschüttler REAX 2, Heidolph Instruments GmbH & Co. KG
- Vortex Genie 2, Scientific Industries Inc.

Thermocycler

- LightCycler 480 II, Roche Molecular Systems, Inc.
- vapo.protect Mastercycler pro, Eppendorf GmbH
- vapo.protect Mastercycler pro 384, Eppendorf GmbH

Waage

- TE3102-S, Sartorius AG

Zentrifugen

- Megafuge 1.0 R, Heraeus GmbH
- Mini star silverline, VWR International GmbH

3.2 Chemikalien, Puffer und Lösungen

- Bromphenol Blau, Sigma-Aldrich Chemie GmbH
- Dimethylsulfoxid; DMSO (C₂H₆SO), Sigma-Aldrich Chemie GmbH
- Elution Buffer, QIAGEN N.V.
- Ethanol absolut; EtOH (C₂H₅OH), CHEMSOLUTE®, Th. Geyer GmbH & Co. KG
- Ethidium Bromid; EtBr (C₂₁H₂₀N₃Br) 1%, Merck KGaA
- Ladepuffer: 10ml 10X TBE, 10ml 0,1%Bromphenol Blau, 40ml 20% Ficoll, 40ml H₂O_{bidest}
- Natriumhypochlorit (NaOCl) (12%), Carl Roth GmbH + Co. KG
- 0,4% Natriumhypochlorit-Lösung: 15,4ml NaOCl, 484,6ml H₂O
- 10X TBE Puffer, AppliChem GmbH

3.3 Ergänzende Materialien zur Analyse von Nukleinsäuren

- Biozym LE Agarose, Biozym Scientific GmbH
- dNTPs [10mM], Labomedic GmbH
- LightCycler 480 SYBR Green Master, Roche Molecular Systems, Inc.
- Molecular biology reagent, Sigma-Aldrich Chemie GmbH
- PCR Puffer 10X Extra Buffer mit 15mM MgCl₂, VWR International GmbH
- PNK Puffer 10X, New England Biolabs, Inc.
- Qubit dsDNA BR Assay Kit, Thermo Fisher Scientific GmbH
- Qubit dsDNA HS Assay Kit, Thermo Fisher Scientific GmbH
- TapeStation 2200 High Sensitivity D1000 ladder, Agilent Technologies, Inc.
- TapeStation 2200 High Sensitivity D1000 sample buffer, Agilent Technologies, Inc
- TapeStation 2200 High Sensitivity D1000 ScreenTape, Agilent Technologies, Inc
- TapeStation 2200 D1000 ladder, Agilent Technologies, Inc.
- TapeStation 2200 D1000 sample buffer, Agilent Technologies, Inc.
- TapeStation 2200 D1000 ScreenTape, Agilent Technologies, Inc.
- VWR-Taq DNA-Polymerase [5U/μl], VWR International GmbH

3.4 Kommerzielle Systeme

- AMPure XP, Beckman Coulter GmbH
- 5X Big Dye Terminator Cycle sequencing Kit 3.1, Life Technologies GmbH
- CleanSEQ, Agencourt, Beckman Coulter GmbH
- DNA Längenstandard 100bp, AppliChem GmbH

- DNA Längenstandard 1kb, AppliChem GmbH
- Infinium PsychArray-24 v1.2 Kit, Illumina
- iPLEX™ Gold Reagent Kit, Sequenom® GmbH
- SpectroCHIP® Arrays & Clean Resin Kit, Sequenom® GmbH

3.5 Software und Datenbanken

- Assay Design 3.1, Sequenom® GmbH
- Biomek® Software 3.2, Beckman-Coulter GmbH
- CADD - Combined Annotation Dependent Depletion v1.3 (<https://cadd.gs.washington.edu/>)
- CLC bio Biomedical Genomics Workbench 2.5.1, QIAGEN
- 2D CYPHER™ Pilot Databases, Thermo Fisher Scientific GmbH
- Copy Number Inference From Exome Reads (CoNIFER; Krumm et al., 2012)
- 3130xl DataCollection v3.0, Life Technologies GmbH
- ENSEMBL genome browser (<http://www.ensembl.org/index.html>)
- ExAC browser (<http://exac.broadinstitute.org/>)
- ExomeDepth (Plagnol et al., 2012)
- eXome-Hidden Markov Model (XHMM; Fromer et al., 2012)
- Exome Variant Server (<http://evs.gs.washington.edu/EVS>)
- Genebook (<http://atgu.mgh.harvard.edu/~spurcell/genebook/genebook.cgi?user=guest&cmd=overview#>)
- GenomeStudio™ v2.0.2, Illumina® Inc.
- 1000 Genomes Project (<http://1000genomes.org>)
- HGNC (<http://genenames.org>)
- LightCycler® 480 Software (Roche, Basel, Schweiz)
- MySequenom, Assay Design Suite (<https://mysequenom.com/Home>)
- NanoDrop® ND-100 v3.3.0, Peqlab Biotechnology GmbH
- NanoDrop® ND-8000 v2.2.1, Peqlab Biotechnology GmbH
- NCBI (<http://www.ncbi.nlm.nih.gov>)
- OMIM (<https://omim.org/>)
- PLINK v1.90 (<http://www.cog-genomics.org/plink2/>)
- Primer3 v0.4.0 (<https://bioinfo.ut.ee/primer3-0.4.0/>)
- Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed/>)
- RT-Workstation 3.3, Sequenom® GmbH
- SeqMan II Version 5.0, DNASTAR Inc.
- SmartScan™ 96, Thermo Fisher Scientific GmbH
- SpectroAQUIRE, Version 3.3.1.2, Sequenom® GmbH

- SpectroPOINT, Sequenom® GmbH
- ToppNet (<https://toppgene.cchmc.org/network.jsp>)
- Typer 3.4/ 4.0, Sequenom® GmbH
- UCSC (<http://genome.ucsc.edu/>)
- Varbank (<https://varbank.ccg.uni-koeln.de/>)
- VEGAS: Versatile Gene-based Association Study (<http://gump.qimr.edu.au/VEGAS/>)

3.6 Patienten- und Kontrollkollektive

Alle Untersuchungen wurden in Übereinstimmung mit der Deklaration von Helsinki des Weltärztebundes (World Medical Association, 2013) durchgeführt. Alle Teilnehmer wurden ausführlich aufgeklärt und willigten schriftlich in die Studienteilnahme ein. Die Studien wurden von den jeweiligen Ethikkommissionen der teilnehmenden Rekrutierungszentren genehmigt.

Die Bonner Arbeitsgruppe kollaboriert seit mehr als 20 Jahren mit rekrutierenden Kolleginnen und Kollegen aus ganz Deutschland: Frau PD Dr. Ina Giegling, Herr Prof. Wolfgang Maier, Herr Prof. Andreas Reif, Frau Prof. Marcella Rietschel, Herr Prof. Dan Rujescu und Frau Prof. Sibylle Schwab. Aus Datenschutzgründen werden die verschiedenen Rekrutierungsstandorte hier nicht einzeln aufgeführt.

Von allen Patienten lag mindestens die psychiatrische Diagnose vor. Diese wurde von erfahrenen Psychiatern basierend auf den Kriterien des Diagnostischen und Statistischen Handbuchs Psychischer Störungen (*Diagnostic and Statistical Manual of Mental Disorders*; DSM-III, DSM-III-R, oder DSM-IV) der Amerikanischen Psychiatrischen Vereinigung (American Psychiatric Association, 1980, 1987, 1994) oder des Klassifikationssystems ICD-10 der Weltgesundheitsorganisation (*International Statistical Classification of Diseases and Related Health Problems*) (Dilling et al., 2015) gestellt. Zusätzlich wurde ein Teil der Patienten hinsichtlich Kataphasie und Katatonie phänotypisiert.

Es wurden unterschiedliche Patientenkollektive verwendet:

- 1) 54 multiplex Familien
- 2) 141 Index Patienten

Weiterhin wurden folgende populationsbasierte deutsche Kontrollkollektive verwendet: 145 Exome (Krawczak et al., 2006; Nöthlings & Krawczak, 2012) sowie Genotypisierungsdaten von ~1.000 deutschen Kontrollen (Blutspender).

Multiplex Familien

Die Analysen der multiplex Familien wurden im Rahmen des vom Bundesministerium für Bildung und Forschung im Rahmen des e:Med Programms geförderten Konsortiums IntegraMent (*Integrated Understanding of Causes and Mechanisms in Mental Disorders*) durchgeführt.

Insgesamt wurden 54 multiplex Familien exomweit sequenziert. 12 Familien wurden in Jakarta/Indonesien und 42 Familien in verschiedenen Zentren in Deutschland rekrutiert.

Zu Beginn der Studie lag bereits DNA der Probanden vor. Von $n = 11$ Individuen war keine aus venösem Blut extrahierte genomische DNA mehr vorhanden. Alternativ wurde aus Zellkultur oder Genomamplifikation (*whole genome amplification*, WGA) gewonnene DNA verwendet.

In 53 Familien hatten mindestens drei Familienmitglieder eine der folgenden Diagnosen: SCZ, Psychose, schizophreniforme Störung, schizoaffektive oder affektive Störung (im Folgenden als "betroffen" bezeichnet). In einer Familie hatte das dritte erkrankte Familienmitglied die Diagnose einer Anpassungsstörung. In einigen Familien gab es weitere Angehörige, die andere psychiatrische Störungen hatten, wie z.B. Alkoholmissbrauch. Diese wurden im Folgenden als Individuen mit einer „anderen psychiatrischen Erkrankung“ bezeichnet, Individuen ohne psychiatrische Diagnose als „gesund“. Exemplarische Stammbäume sind in Abbildung 3 dargestellt.

Bei den 54 Familien waren Informationen zur psychiatrischen Diagnose von durchschnittlich $n = 6,4$ (3-27) Individuen pro Familie vorhanden. Davon waren insgesamt $n = 187$ betroffen, $n = 138$ gesund und $n = 23$ hatten die Diagnose einer anderen psychiatrischen Erkrankung.

Zu Beginn der Arbeit wurde eine Pilotstudie mit drei deutschen multiplex Familien zur Etablierung der Methoden und Analysestrategien durchgeführt. In diesen drei Familien wurden insgesamt 20 Individuen untersucht ($n = 9$ Patienten mit SCZ, $n = 3$ Patienten mit einer unipolaren Depression und $n = 8$ gesunde Familienmitglieder). In zwei Familien wurde bei den SCZ Patienten zusätzlich eine periodische Katatonie diagnostiziert, während bei den SCZ Patienten der dritten Familie zusätzlich eine Kataphasie diagnostiziert wurde. Zu den Symptomen der periodischen Katatonie zählen sowohl hyperkinetische als auch akinetische Störungen des psychomotorischen Verhaltens (Stöber et al., 2000), während die Hauptsymptome der Kataphasie formale Denk- und Sprachstörungen sind (Pfuhmann et al., 1998). Sowohl die periodische Katatonie als auch die Kataphasie gehören nach der Klassifikation nach Leonhard zu den unsystematischen Schizophrenie Formen (Leonhard, 1999).

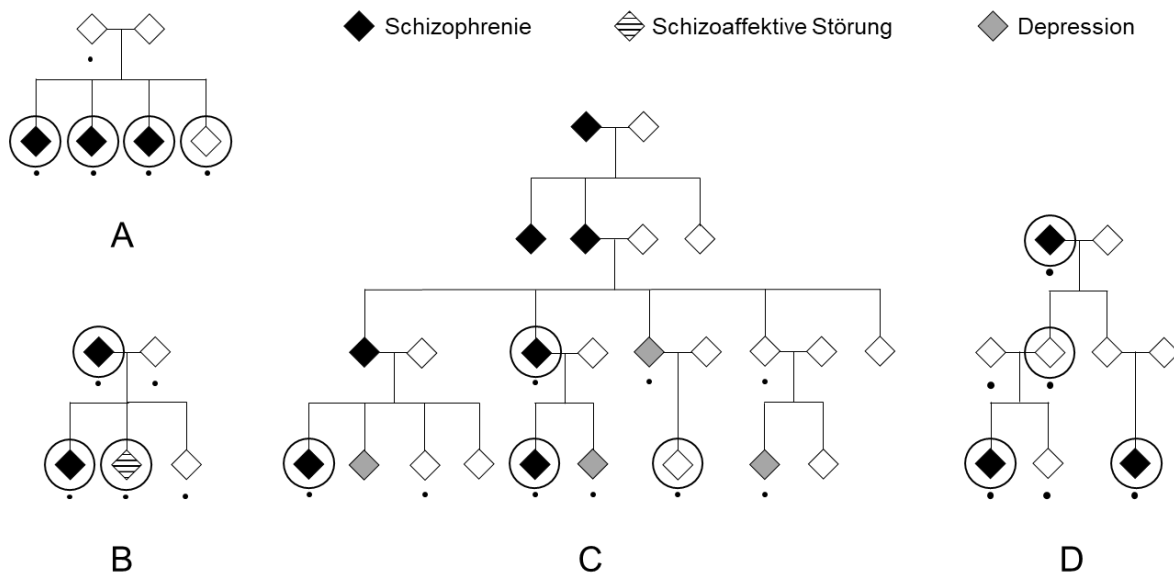


Abbildung 3 Exemplarische Stammbäume von vier multiplex Familien. DNA war von allen mit einem Punkt markierten Individuen vorhanden. Eingekreiste Individuen wurden für die exomweite Sequenzierung ausgewählt. Schwarze Symbole stellen Individuen mit Schizophrenie dar, gestreifte Symbole Individuen mit schizoaffectiver Störung und graue Symbole Individuen mit einer affektiven Erkrankung (Depression). Die betroffenen Individuen waren im jeweiligen Stammbaum über A) eine, B) zwei, C) drei und mehr, oder D) zwei nicht aufeinanderfolgende Generationen verteilt. Zum Schutz der Privatsphäre der Familien wird in der Abbildung nicht zwischen männlichen und weiblichen Familienmitgliedern unterschieden und die Reihenfolge der Geburt wurde vertauscht.

Studiendesign

Zunächst wurde eine Pilotstudie mit drei deutschen multiplex Familien zur Etablierung der Methoden und Analysestrategien durchgeführt. In diesen Familien wurden sowohl seltene SNVs und Indels als auch seltene CNVs untersucht. Anschließend wurden die übrigen 51 Familien analysiert. Hierbei lag der Fokus auf seltenen CNVs. Eine Übersicht über das Studiendesign ist in Abbildung 4 dargestellt.

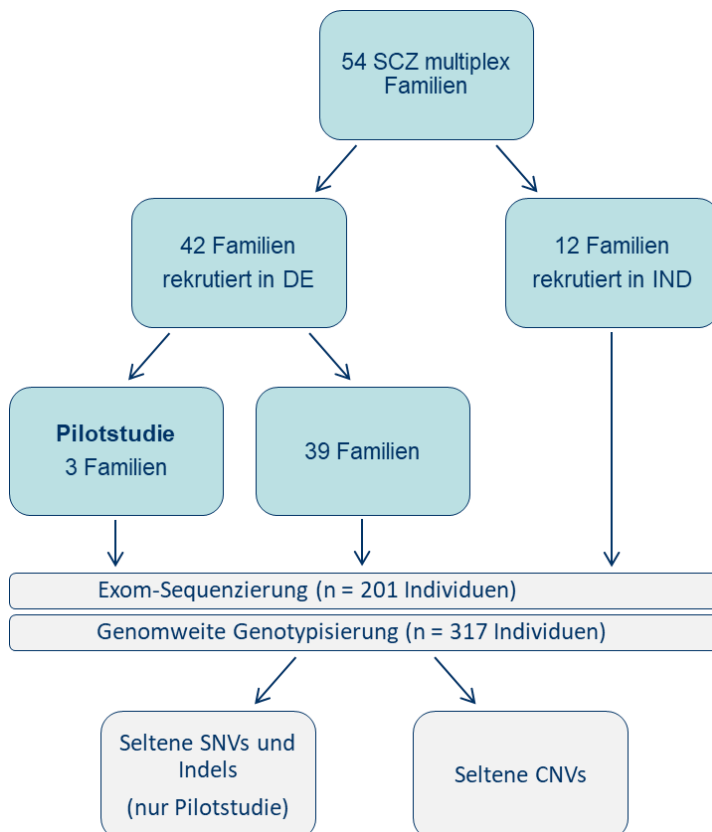


Abbildung 4 Übersicht über exomweit sequenzierte multiplex Familien. SCZ: Schizophrenie; DE: Deutschland; IND: Indonesien; SNVs: Einzelnukleotid-Varianten; Indels: Insertionen und Deletionen; CNVs: Kopienzahlvarianten

Indexe (Exom-Sequenzierung)

Der Indexe-Datensatz beinhaltete die Exome von 141 Patienten, bei denen eine SCZ diagnostiziert wurde. Die in Deutschland rekrutierten Indexpatienten waren unabhängig von (nicht verwandt mit) den exomweit sequenzierten Familien.

75 Patienten dieses Kollektivs wurden mit einem jungen Ersterkrankungsalter (durchschnittliches Alter bei Symptombeginn: 16 Jahre, durchschnittliches Alter bei Diagnose SCZ: 18 Jahre) und einem chronifizierenden Krankheitsverlauf diagnostiziert. Bei den anderen 66 Patienten wurde zusätzlich zur SCZ eine periodische Katatonie diagnostiziert.

3.7 Genomweite Genotypisierung

Insgesamt 317 Individuen der Familien und 141 Indexpatienten wurden zusätzlich auf dem Illumina Infinium PsychArray-24 BeadChip genomweit genotypisiert (Illumina, San Diego, CA, USA). Die Daten wurden im Department of Genomics des Life and Brain Centers (Universität Bonn) generiert und zur Verfügung gestellt.

Basierend auf den genomweiten Genotypisierungsdaten wurde bei allen Probanden das phänotypische mit dem genotypischen Geschlecht abgeglichen sowie bei den multiplex Familien die im Stammbaum abgebildete Verwandtschaft der Familienmitglieder überprüft. Zur Untersuchung der Verwandtschaft wurde die Übereinstimmung der analysierten Allele „Identity by descent“ zwischen zwei Familienmitgliedern mittels PLINK v1.9 (<https://www.cog-genomics.org/plink/1.9>; Chang et al. 2015) berechnet.

3.8 Ausschluss von Schizophrenie-assoziierten

Kopienzahlvarianten

Die genomweiten Genotypisierungsdaten (s. Abschnitt 3.7) wurden weiterhin dazu verwendet, mittels Illumina GenomeStudio bekannte SCZ-assoziierte CNVs bei den exomweit sequenzierten multiplex Familien auszuschließen. Dabei wurden folgende Loci untersucht: 1q21.1; 2p16.3 (*NRXN1*); 3q29; 7q11.2; 7q36.3; 8q11.23; 8q22.2; 9p24.3, 15q11.2; 15q11.2-13.1; 15q13.2-13.3; 15q13.3; 16p11.2 distal; 16p11.2 proximal; 22q11.2; Xq28 distal (Marshall et al., 2017).

3.9 Exom-Sequenzierung

3.9.1 Probenvorbereitung

Die Exom-Sequenzierung ermöglicht die Analyse aller Protein-kodierenden Bereiche des menschlichen Genoms. Insgesamt wurden n = 201 Individuen aus 54 SCZ multiplex Familien exomweit sequenziert. Aus Kostengründen wurden nicht alle Familienmitglieder, von denen genug DNA vorhanden war, sequenziert, sondern nur diejenigen, die zur Identifizierung der potentiell krankheitsrelevanten seltenen Varianten nötig waren.

In 53 Familie wurden jeweils drei „betroffene“, welche möglichst weit entfernt miteinander verwandt waren, exomweit sequenziert. In diesen Familien waren n = 128 Individuen an SCZ oder einer Psychose erkrankt, n = 14 an einer schizoaffektiven Störung, n = 8 an einer bipolaren Störung, n = 13 an einer unipolaren Depression, n = 15 an einer anderen psychiatrischen Erkrankung und n = 23 Individuen waren gesund.

In einer Familie wurden zwei „betroffene“ und ein drittes Familienmitglied mit einer anderen psychiatrischen Erkrankung sequenziert.

Zur Qualitätsbestimmung der DNA wurden jeweils 5µl unverdünnter DNA zusammen mit 5µl Ladepuffer auf ein 0,8%iges Agarosegel aufgetragen. Nach einer 30-minütigen Gelelektrophorese bei 80V wurde die DNA auf dem Gelbild auf Degradierung überprüft. Degradierete DNA wurde von der exomweiten Sequenzierung ausgeschlossen.

Anschließend wurde die DNA in Elution Buffer (QIAGEN) auf 30ng/µl verdünnt. Die Konzentrationsbestimmung der doppelsträngigen DNA erfolgte mittels Qubit 2.0 Fluorometer (Invitrogen by Life technologies) nach Herstellerprotokoll.

Die verdünnte DNA wurde mittels Bioruptor Ultraschallgerät (Diagenode) fragmentiert. Die Größe der DNA-Fragmente wurde mittels TapeStation 2200 (Agilent Technologies) überprüft.

3.9.2 Sequenzierung

Die genomischen Zielsequenzen wurden mit dem SureSelect Human All Exon v5 System von Agilent Technologies angereichert. Die Cluster-Generierung erfolgte mittels Illumina cBot unter Verwendung des PE_HiSeq_Cluster_Kit_v4_cBot_recipe_v9.0 Protokolls. Die Sequenzierung wurde unter Verwendung eines 2x125bp-Paired-End-Leseprotokolls auf einem Illumina HiSeq2500 durchgeführt.

3.9.3 Bioinformatische Auswertung und Filtern der detektierten Varianten

Zur Identifizierung von Varianten aus den durch die Sequenzierung generierten Rohdaten müssen die erhaltenen DNA-Fragmente (*Reads*) zunächst mit dem menschlichen Referenzgenom (in der vorliegenden Studie hg19) abgeglichen (*aligned*) werden. Dadurch werden *fastq* Dateien in *BAM* Dateien konvertiert, anhand dieser schließlich das *Variant Calling* erfolgen kann. Das *Alignment* und das *Variant Calling* wurden mit zwei verschiedenen Datenanalyseplattformen durchgeführt: Der vom Cologne Center for Genomics entwickelten Varbank (<https://varbank.ccg.uni-koeln.de>) in Zusammenarbeit mit Dr. Holger Thiele (Cologne Center for Genomics, Universität Köln) und der CLC bio Biomedical Genomics Workbench 2.5.1 von QIAGEN. Die verwendeten Einstellungen und Filter zur Qualitätskontrolle sind in den Tabellen A1 und A2 im Anhang aufgeführt (s. auch Abbildung 5).

Die drei Familien der Pilotstudie wurden mit CLC und Varbank ausgewertet. Hierbei wurde die Summe aus den identifizierten Varianten der beiden Datenanalyseplattformen in die Folgeanalysen eingeschlossen. Bei den übrigen 51 Familien lag der Fokus auf der Untersuchung seltener CNVs. Dafür wurden die mit der Varbank generierten *BAM* Dateien verwendet (s. Abschnitt 3.10).

Zum Einschluss in die folgenden Analysen mussten die identifizierten Varianten folgende Filterkriterien erfüllen:

1. Die Variante musste im kodierenden Bereich des Gens liegen und somit einen potentiellen Effekt auf die Struktur des vom Gen kodierten Proteins haben oder die Spleißstelle ($\pm 5\text{bp}$ von der Exongrenze entfernt) betreffen.

2. Die Variante musste in der Allgemeinbevölkerung sehr selten sein und die Frequenz des selteneren Allels (*minor allele frequency*, MAF) war in folgenden Kohorten $\leq 0,1\%$: (i) 1000 Genomes Projekt (1000 Genomes Project Consortium et al. 2015; ~ 2.500 Individuen), (ii) Exome Variant Server (<http://evs.gs.washington.edu/EVS>; ~ 6.500 Individuen) und (iii) Exome Aggregation Consortium (ExAC; (Lek et al. 2016) ohne psychiatrische Kohorten (~ 45.000 Individuen)). Hierbei wurden bei den deutschen Familien ausschließlich die angegebenen Allelfrequenzen in der europäischen Bevölkerung verwendet (ExAC: nicht-finnische Europäer; ~ 21.000 Individuen) und bei den indonesischen Familien die Allelfrequenzen in der süd- (~ 8.300 Individuen in ExAC) und der ostasiatischen Bevölkerung (~ 2.800 Individuen in ExAC).

Da keine der öffentlich zugänglichen Datenbanken deutsche Probanden enthält, wurde die Allelfrequenz zusätzlich in 145 Exomen von Kontrollprobanden (Krawczak et al., 2006; Nöthlings & Krawczak, 2012) sowie in einem unabhängigen Kollektiv von 141 SCZ Indexpatienten aus Deutschland überprüft. Zusätzlich wurden ausgewählte Varianten in 1.000 deutschen Kontrollprobanden zur Frequenzanalyse genotypisiert.

3. Die Varianten mussten einen Combined Annotation Dependent Depletion (CADD) Score v1.3 ≥ 15 haben (Kircher et al., 2014). Mit Hilfe des CADD Scores kann die Pathogenität von Einzelnukleotid-Varianten und Indels *in silico* geschätzt werden. Die Scores reichen von 1 bis 99, wobei höhere Werte anzeigen, dass eine Variante mit größerer Wahrscheinlichkeit pathogen ist.

4. Die für die Downstream-Analyse betrachteten Varianten mussten in mindestens zwei „betroffenen“ Familienmitgliedern identifiziert werden (unter Berücksichtigung von Phenokopie und genetischer Heterogenität).

Ausschlusskriterien waren Homozygotie und Auftreten der spezifischen Variante in > 3 unabhängigen Familien. In einer Familie der Pilotstudie wurden alle Varianten ausgeschlossen, die in einem gesunden Familienmitglied identifiziert wurden.

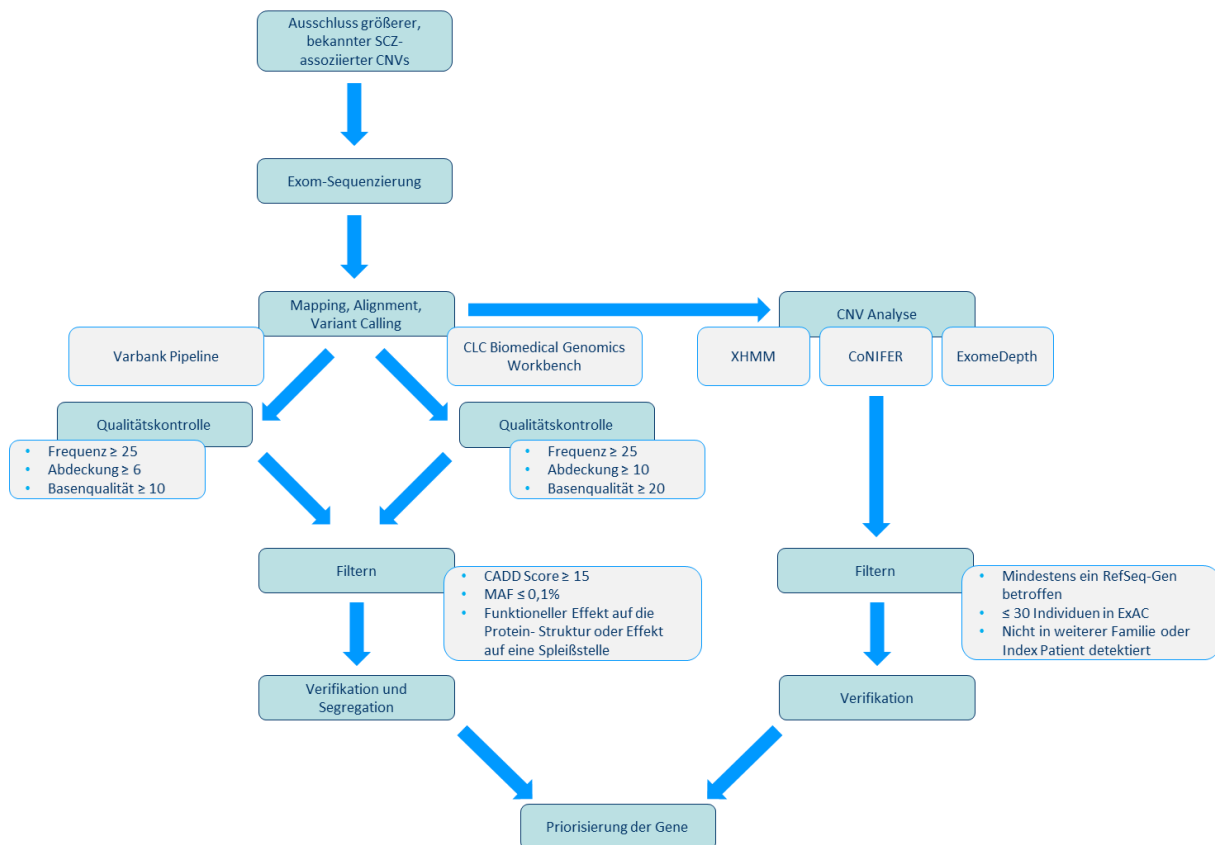


Abbildung 5 Arbeitsablauf der Exom-Sequenzierungsstudie

Vor der Exom-Sequenzierung wurde das Vorhandensein von bekannten Schizophrenie (SCZ)-assoziierten Kopienzahlvarianten (CNVs) ausgeschlossen (Marshall et al., 2017). Detektion und Filtern der Varianten erfolgte über zwei verschiedene Datenanalyseplattformen: (i) die Varbank pipeline v2.14-v2.25 des Cologne Center for Genomics (<http://varbank.ccg.uni-koeln.de>) und (ii) die CLC bio Biomedical Genomics Workbench 2.5.1 (QIAGEN, Aarhus, Dänemark; <https://www.qiagenbioinformatics.com/>). Um in die nachfolgenden Analysen einbezogen zu werden, musste die Frequenz des selteneren Allels (*minor allele frequency*, MAF) der identifizierten Variante in folgenden Kohorten $\leq 0.1\%$ sein: (i) 1000 Genomes Projekt (1000 Genomes Project Consortium et al., 2015), (ii) Exome Variant Server (<http://evs.gs.washington.edu/EVS>), (iii) Exome Aggregation Consortium (Lek et al., 2016) ohne psychiatrische Kohorten (~45.000 Individuen), (iv) 145 Exome von deutschen Kontrollen und zusätzlich für die drei Familien der Pilotstudie in den Genotypisierungsdaten von 1.000 deutschen Kontrollen. Weiterhin mussten die Varianten einen Combined Annotation Dependent Depletion (CADD) Score v1.3 ≥ 15 haben (Kircher et al., 2014). Außerdem wurde eine auf Exom Daten basierende CNV-Analyse durchgeführt. Dafür wurden folgende Algorithmen verwendet: (i) eXome-Hidden Markov Model (XHMM; Fromer et al., 2012); (ii) Copy Number Inference From Exome Reads (CoNIFER; Krumm et al., 2012); und (iii) ExomeDepth (Plagnol et al., 2012). In die nachfolgenden Analysen einbezogen wurden CNVs, die (i) mindestens ein RefSeq-Gen (<https://www.ncbi.nlm.nih.gov/refseq/>) betrafen, (ii) ≤ 30 mal im ExAC Browser (exac.broadinstitute.org/) vorkamen und (iii) nicht in einer weiteren Familie oder einem der 141 Index Patienten identifiziert wurden.

3.10 Genomweite Analyse von Kopienzahlvarianten

3.10.1 Verwendete Algorithmen zur Detektion von Kopienzahlvarianten

Eine auf Exomdaten basierende CNV-Analyse wurde in 51 Familien durchgeführt. In drei Familien war eine CNV-Analyse nicht möglich, da von diesen keine genomische DNA vorhanden war.

Zur Detektion der CNVs wurden drei verschiedene Algorithmen verwendet: eXome-Hidden Markov Model (XHMM; Fromer et al., 2012), Copy Number Inference From Exome Reads (CoNIFER; Krumm et al., 2012) und ExomeDepth (Plagnol et al., 2012). Die CNVs wurden *in silico* anhand der Abdeckung der proteinkodierenden Zielsequenzen durch sequenzierte Fragmente (*Reads*) identifiziert.

Die für die CNV-Analyse benötigten Exom *BAM* Dateien wurden aus der Varbank (<https://varbank.ccg.uni-koeln.de/>) heruntergeladen. Für jedes Individuum werden in diesem Dateiformat die *Alignment*-Positionen der einzelnen *Reads* an ein Referenzgenom, in unserem Fall hg19, gespeichert. Da die Qualität der Ergebnisse von XHMM und CoNIFER mit der Anzahl der analysierten Individuen steigt, wurden zusätzlich zu den Mitgliedern der SCZ multiplex Familien ein hausinterner Datensatz von 141 SCZ Index Patienten analysiert. In ExomeDepth wird die Abdeckung des Exoms jedes Individuums gegen ein Referenzset getestet. Dieses bestand aus 188 exomweit sequenzierten Familienmitgliedern und den 141 SCZ Indexpatienten. Die Familienmitglieder von denen nur aus WGA gewonnene DNA vorhanden war ($n = 9$ Individuen) wurden aus dem Referenzset ausgeschlossen. Die Exom Daten aller analysierten Individuen wurden unter denselben Bedingungen im Labor sowie mit denselben bioinformatischen Verfahren generiert.

Mit folgenden Ausnahmen wurden die Standardeinstellungen von XHMM, CoNIFER, und ExomeDepth verwendet: (i) Bei XHMM, wurden die Parameter `-maxMeanTargetRD` und `-maxMeanSampleRD` auf 1.500 festgelegt; (ii) Bei CoNIFER wurden, die ersten sechs Komponenten der Singulärwertzerlegung, die den bedeutendsten Teil der Varianz erklärten, extrahiert; (iii) Bei ExomeDepth wurde als Grenze für die Berechnung der Read-Abdeckung die Länge der Zielsequenzen des Exomanreicherungsverfahrens verwendet.

3.10.2 Filtern der detektierten Kopienzahlvarianten

Die CNV Detektion ist vulnerabel für falsch-positive Befunde. Daher wurden stringente Filterkriterien angewendet (Abbildung 5). Eingeschlossen in die Analyse wurden nur Deletionen und Duplikationen, welche folgende Kriterien erfüllten:

1. Die CNVs mussten von allen drei Algorithmen (XHMM, CoNIFER, und ExomeDepth) detektiert werden.
2. Die CNVs lagen nicht auf dem X- oder dem Y-Chromosom.
3. Die CNVs wurden innerhalb einer Familie in mindestens zwei Familienmitgliedern mit SCZ, bipolarer Störung oder schizoaffektiver Störung identifiziert.
4. Mindestens ein RefSeq Gen (<https://www.ncbi.nlm.nih.gov/refseq/>) war von der CNV betroffen.

Der Fokus der Studie lag auf seltenen CNVs. Daher wurden alle Deletionen und Duplikationen ausgeschlossen, die mit einer in (i) > 30 Individuen im ExAC Browser (Lek et al. 2016; 60.706 Individuen; exac.broadinstitute.org/), oder (ii) > einer weiteren Familie oder einem der SCZ Patienten aus dem hausinternen Datensatz detektierten CNV zumindest teilweise überlappten.

3.11 Experimentelle Verifikation

3.11.1 Verifikation der detektierten Einzelnukleotid-Varianten und Indels

Alle identifizierten SNVs und Indels wurden zunächst in den Readbrowsern der verwendeten Datenanalyseplattformen Varbank und CLC visuell überprüft. Zusätzlich wurden alle 35 potentiell echten Varianten der Pilotstudie experimentell überprüft. Dies diente der Überprüfung der Qualität der etablierten Filter- und Auswertestrategie. Die Varianten wurden experimentell entweder mittels Agena Bioscience MassARRAY System genotypisiert oder Sanger sequenziert.

Genotypisierung mittels Agena Bioscience MassARRAY System

Um die Sonden für die Genotypisierung zu designen wurde das Tool Assay Design Suite von Agena Bioscience verwendet (www.agenacx.com). Für fünf Varianten konnten keine Sonden designet werden. Die übrigen 30 Varianten wurden mittels Agena Bioscience Mass Array System und iPLEX™ Gold Reagent Kit nach Herstellerprotokoll genotypisiert (Primerliste auf Anfrage erhältlich). Drei Varianten konnten nicht erfolgreich genotypisiert werden.

Sanger Sequenzierung

Die insgesamt acht Varianten, für die das Primerdesign oder die Genotypisierung nicht erfolgreich waren, wurden mittels Sanger Sequenzierung verifiziert (Primerliste auf Anfrage erhältlich). Die dafür benötigten Primer wurden mittels Primer3 software (<http://primer3.ut.ee/>) entworfen. Die Sequenzierung erfolgte mittels 3130xl Genetic Analyzer (Applied Biosystems) und die Auswertung der Sequenzierungsdaten mittels SeqMan II (DNASTAR). Für eine Variante (c.1924G>T im Gen *KRBA1*) war aus technischen Gründen weder die Genotypisierung noch die Sanger Sequenzierung erfolgreich. Da diese Variante jedoch in den Readbrowsern von Varbank und CLC bei allen betroffenen Familienmitgliedern klar identifizierbar war, wurde sie ebenfalls in die Downstream-Analyse eingeschlossen.

3.11.2 Verifikation der detektierten Kopienzahlvarianten mittels qPCR

Alle CNVs, die in die Studie eingeschlossen wurden, wurden mittels quantitativer Polymerase-Kettenreaktion (qPCR) verifiziert. Eine qPCR basiert auf dem Prinzip einer herkömmlichen PCR, wobei zusätzlich ein in die DNA interkalierender Farbstoff verwendet wird. Dieser wird nach jedem Amplifikationszyklus gemessen, um die amplifizierte DNA zu quantifizieren.

Für jede CNV Region wurden mittels Primer3 software (<http://primer3.ut.ee/>) zwei bis vier Primerpaare entworfen (Primerliste auf Anfrage erhältlich). Zusätzlich wurden zwei Haushaltsgene (*BCN1*, *CFTR*), welche immer in zwei Kopien im Genom vorliegen, zur Kontrolle untersucht. Pro Primerpaar wurde ein Mix aus 493.3µl Millipore Wasser, 3.34µl Vorwärtsprimer (100µM) und 3.34µl Rückwärtsprimer (100µM) erstellt. Die zu untersuchende DNA wurde auf 10ng/µl in Millipore Wasser verdünnt. Zusätzlich wurden auch Kontrollen mit bekannter Kopienzahl (CN = 2) verwendet. Pro Reaktion wurden 2µl DNA, 3µl Primermix und 5µl SYBR™ Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) benötigt.

Die qPCR Reaktion erfolgte in einem LightCycler® 480 II (Roche, Basel, Schweiz). Die einzelnen Schritte des verwendeten Cyclex-Programms sind in Tabelle 1 aufgeführt. Die Kopienzahlen wurden mittels $2^{-\Delta\Delta CT}$ Methode berechnet (Livak & Schmittgen, 2001).

Tabelle 1 Cyclex-Programm für qPCR

Reaktionsschritt	Temperatur	Zeit
Prä-Inkubation (1 Zyklus)	95°C	10 s
Amplifikation (40 Zyklen)	95°C	15 s
	60°C	1 min
	72°C	15 s
Schmelzkurve (1 Zyklus)	95°C	5 s
	65°C	1 min
	97°C	-
Abkühlung	40°C	10 s

3.12 Priorisierung der detektierten Gene

Aufgrund der genetischen Heterogenität und der Komplexität des untersuchten Phänotyps, wurden durch die Einzelnukleotid- und CNV-Analysen – wie erwartet – zahlreiche Kandidatengene identifiziert. Es wurden vier verschiedene Strategien angewendet, um die Gene weiter hinsichtlich ihrer Krankheitsrelevanz zu priorisieren: (i) Abgleich der eigenen Daten mit bereits publizierten Studien zu Entwicklungsstörungen des Gehirns, (ii) Gen-basierte Tests, (iii) Protein-Protein Interaktionsanalysen und (iv) Genexpressionsanalysen.

3.12.1 Genetische Evidenz – Abgleich mit publizierten Studien

Zunächst wurde ein Abgleich der eigenen Daten mit weiteren Studien zu Entwicklungsstörungen des Gehirns, u.a. SCZ, Intelligenzminderung (ID) und Autismus-Spektrum Störung (ASD), durchgeführt.

Dazu wurden ein hausinterner Datensatz mit 141 Patienten mit SCZ und verschiedene öffentlich zugängliche Datenbanken, wie Pubmed (<https://pubmed.ncbi.nlm.nih.gov/>) und Genebook, eine Datenbank, in der Ergebnisse aus genetischen SCZ Studien enthalten sind (Purcell et al. 2014), verwendet.

Unsere Daten wurden mit Genen abgeglichen, welche: (i) im weltweit größten CNV-Datensatz bei Patienten mit SCZ (21.094 Patienten und 20.227 Kontrollen) in einem der acht genomweit signifikanten Loci lagen (397 Gene; Marshall et al. 2017), (ii) in einer der bisher größten publizierten SCZ GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014 (36.989 Patienten und 113.075 Kontrollen); Pardiñas et al. 2018 (40.675 Patienten und 64.643 Kontrollen) Trubetskoy et al., 2022 (76.755 Patienten und 243.649 Kontrollen)) von einem genomweit signifikanten SNP betroffen waren (insgesamt 239 Gene) und (iii) in anderen SCZ Exom- und Genom-Sequenzierungsstudien als potentiell krankheitsursächlich identifiziert wurden (~ 3000 Gene) (Girard et al., 2011; Xu et al., 2011, 2012; Timms et al., 2013; Gulsuner et al., 2013; Guipponi et al., 2014; Fromer et al., 2014; Purcell et al., 2014; Takata et al., 2014; McCarthy et al., 2014; Kranz et al., 2015; Wang et al., 2015; Ambalavanan et al., 2016; Watanabe et al., 2016; Egawa et al., 2016; Homann et al., 2016; Zhou et al., 2016; Smedemark-Margulies et al., 2016; Hornig et al., 2017; John et al., 2017; Tang et al., 2017; Kos et al., 2017; O'Brien et al., 2018; Senormanci et al., 2018; Xue et al., 2018; John et al., 2018, 2019; Suddaby et al., 2019; Li et al., 2021; Howrigan et al., 2020; Rees et al., 2020; Balakrishna & Curtis, 2020; Chen et al., 2021; Toyama et al., 2022; Shang et al., 2024).

Zusätzlich wurden die Gene in der Pilotstudie auch auf eine Überlappung mit bekannten Katatonie-assoziierten Risikoloci untersucht (Stöber et al., 2000).

3.12.2 Genetische Evidenz – Gen-basierte Tests

Weiterhin wurden auch Assoziationen von häufigen Varianten in den Kandidatengenen mit SCZ untersucht. Dazu wurden in Zusammenarbeit mit Dr. Till Andlauer (Technische Universität München; Max-Planck-Institut für Psychiatrie, München) Gen-basierte Tests basierend auf einem der weltweit größten SCZ GWAS-Datensätze (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014; 36.989 Patienten und 113.075 Kontrollen) durchgeführt. Die Analysen wurden mittels VEGAS2 v2 (Mishra & Macgregor, 2015) unter Verwendung des 1000 Genomes Phase 3 European Referenz Panels, des Referenzgenoms hg19 ohne flankierende Regionen und aller SNPs innerhalb der Gen-Definition des hg19 Referenzgenoms durchgeführt. Der Haupthistokompatibilitätskomplex (Chromosom 6: 27.000.000-35.000.000) wurde von der Analyse ausgeschlossen. Die Ergebnisse wurden mittels Bonferroni Methode für multiples Testen korrigiert. Jeweils ein Gen, welches in der Pilotstudie (*FRG2C*), bzw. in der genomweiten CNV-Analyse (*RNASE2*) identifiziert wurde, wurde aus der Analyse ausgeschlossen, da es nicht Teil der VEGAS2 Datenbank war.

3.12.3 Protein-Protein Interaktionsanalysen

Die Protein-Protein Interaktionsanalysen (PPI-Analysen) wurden in Zusammenarbeit mit Christian Hänig, Prof. Erich Wanker (Max-Delbrück-Centrum Berlin) und Dr. Sandra Van der Auwera (Universität Greifswald; Deutsches Zentrum für Neurodegenerative Erkrankungen Rostock/Greifswald) durchgeführt.

Basierend auf der *network medicine* Hypothese interagieren Proteine, die für dieselbe Erkrankung relevant sind, tendenziell häufiger miteinander und pathogene genetische Veränderungen in interagierenden Proteinen führen häufig zu ähnlichen Phänotypen (Barabási et al., 2011).

In den PPI-Analysen wurde untersucht, ob es in den Proteinen, welche in den beiden hier beschriebenen Studien (Pilotstudie und CNV-Studie) identifiziert wurden, eine Anreicherung von Interaktionen mit Proteinen gibt, die in publizierten SCZ Studien beschrieben wurden.

Dazu wurde ein Proteinnetzwerk mit in der Literatur beschriebenen SCZ Kandidatengenen erstellt. Dieses enthielt: (i) Gene, die in einem bekannten SCZ-assoziierten CNV-Locus lagen (Rees et al., 2014; Marshall et al., 2017), (ii) Gene, in denen bei einem Patienten mit SCZ eine *de novo* Mutation identifiziert wurde (Girard et al., 2011; Xu et al., 2011, 2012; Gulsuner et al., 2013; Guipponi et al., 2014; Fromer et al., 2014; Takata et al., 2014; McCarthy et al., 2014; Kranz et al., 2015; Wang et al., 2015; Ambalavanan et al., 2016; Smedemark-Margulies et al., 2016; Tang et al., 2017; Suddaby et al., 2019) und (iii) Gene, in denen ein genomweit

signifikanter SNP in zwei der größten bisher publizierten SCZ GWAS identifiziert wurde (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Pardiñas et al., 2018). Diese Liste enthielt zum Zeitpunkt der Durchführung der Pilotstudie 1.240 Gene und zum Zeitpunkt der Durchführung der genomweiten CNV-Analyse 1.271 Gene.

Protein-Protein Interaktionsanalysen – Familien-Pilotstudie

In der Pilotstudie wurden zunächst mit Hilfe der *Human Integrated Protein-Protein Interaction rEference* (HIPPIE) Datenbank (v2.2; <http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/>) zwei Netzwerke erstellt, die ausschließlich *high confidence* PPIs mit einem HIPPIE Score $\geq 0,73$ (Schaefer et al., 2012) enthielten: Das erste Netzwerk (cCandidate) enthielt alle in der Pilotstudie identifizierten Proteine, die in der HIPPIE Datenbank vorhanden waren und deren in der HIPPIE Datenbank verfügbare Interaktionen. Dieses Netzwerk umfasste insgesamt 323 Proteine (davon 26 in der Pilotstudie identifizierte), die durch 350 Interaktionen verbunden waren.

Ein zweites Netzwerk, cHIPPIE_without_cCandidate, enthielt alle in HIPPIE verfügbaren Interaktionen, mit Ausnahme der Interaktionen in cCandidate. Dies waren 12.460 Proteine, die durch 69.420 Interaktionen verbunden waren.

Um eine Anreicherung der oben genannten Literatur-Gene in den erstellten Netzwerken zu überprüfen, wurden ein Chi-Quadrat-Test (McHugh 2013) und ein Exakter Fisher-Test durchgeführt. Zur graphischen Darstellung der PPI Netzwerke wurde die Software Cytoscape (Shannon et al., 2003) verwendet.

Protein-Protein Interaktionsanalysen – CNV-Studie

In der genomweiten CNV-Analyse wurden ebenfalls zwei Netzwerke mittels HIPPIE Datenbank erstellt:

Network_Literature enthielt alle Interaktionen zwischen Proteinen, die von den 1.271 bekannten SCZ Kandidatengenomen aus der Literatur kodiert werden, und allen in HIPPIE verfügbaren Proteinen. Dieses Netzwerk beinhaltete 5.054 PPIs. Das zweite Netzwerk, Network_HIPPIE_minus_Literature, enthielt alle anderen 29.918 in HIPPIE gelisteten PPIs.

In dieser Studie wurden ebenfalls ein Chi-Quadrat-Test (McHugh 2013) und ein Exakter Fisher-Test durchgeführt, um die Anreicherung von PPIs für Proteine, die von den neu identifizierten Genen kodiert werden, in den Netzwerken Network_Literature und Network_HIPPIE_minus_Literature zu vergleichen.

In beiden Studien wurden außerdem die identifizierten Kandidatenproteine zusätzlich mit dem Tool ToppNet (Chen et al. 2007; Chen et al. 2009; <https://toppgene.cchmc.org/network.jsp>) priorisiert. Dazu wurde der k-Step Markov Algorithmus mit $k = 6$ Schritten verwendet.

3.12.4 Genexpressionsanalysen

Die Expressionsanalysen wurden in Kooperation mit Dr. Lalit Kaurani und Prof. André Fischer (Universität Göttingen) durchgeführt und setzten sich aus drei verschiedenen Methoden zusammen: Zuerst wurde die Expression der identifizierten Kandidatengene in verschiedenen Gehirngeweben in der *Genotype-Tissue Expression* (GTEx) Datenbank (GTEx Consortium, 2013) untersucht. Diese Datenbank enthält post mortem Daten von mehreren hundert Individuen aus über 50 verschiedenen Geweben.

Des Weiteren wurden die Kandidatengene auf differenzielle Genexpression zwischen Patienten mit SCZ und Kontrollen in 15 publizierten Microarray- und RNA-Sequenzierungsstudien untersucht (Catts et al., 2014; Chen et al., 2013; de Baumont et al., 2015; Fillman et al., 2013; Hwang et al., 2013; Iwamoto et al., 2004, 2005; Kohen et al., 2014; Lanz et al., 2015; Maycox et al., 2009; Narayan et al., 2008; Pietersen, Mauney, Kim, Lim, et al., 2014; Pietersen, Mauney, Kim, Passeri, et al., 2014; Wu et al., 2012; Zhao et al., 2015). Insgesamt 703 Individuen wurden in diesen Studien untersucht. Die *Gene Expression Omnibus* Datenbank (Edgar et al., 2002) wurde verwendet, um die Daten aus den jeweiligen Studien zu extrahieren. Das Tool GEO2R (Barrett et al., 2013) wurde verwendet, um die differenziell exprimierten Gene zu identifizieren. Die differenzielle Expression wurde als $\log_2$ fold change berechnet. Gene mit einem p-Wert $< 0,05$ wurden als signifikant differenziell exprimiert betrachtet.

Zuletzt wurde eine qPCR in hausinternem post mortem Gehirngewebe (Brodmann Areal 9) durchgeführt. Dazu wurde RNA mittels TRIzol Reagent (Thermo Fisher Scientific) extrahiert und anschließend eine cDNA (*complementary DNA*, komplementäre DNA) Synthese mittels Transcriptor High Fidelity cDNA Synthesis Kit (Roche) durchgeführt. Die qPCR wurde auf einem Roche 480 Light Cycler mit *GAPDH* als Kontrolle zur Normalisierung durchgeführt.

4 Ergebnisse

Die Ergebnisse dieser Arbeit gliedern sich in zwei Hauptteile (Abbildung 6): Die Familien-Pilotstudie und die CNV-Studie.

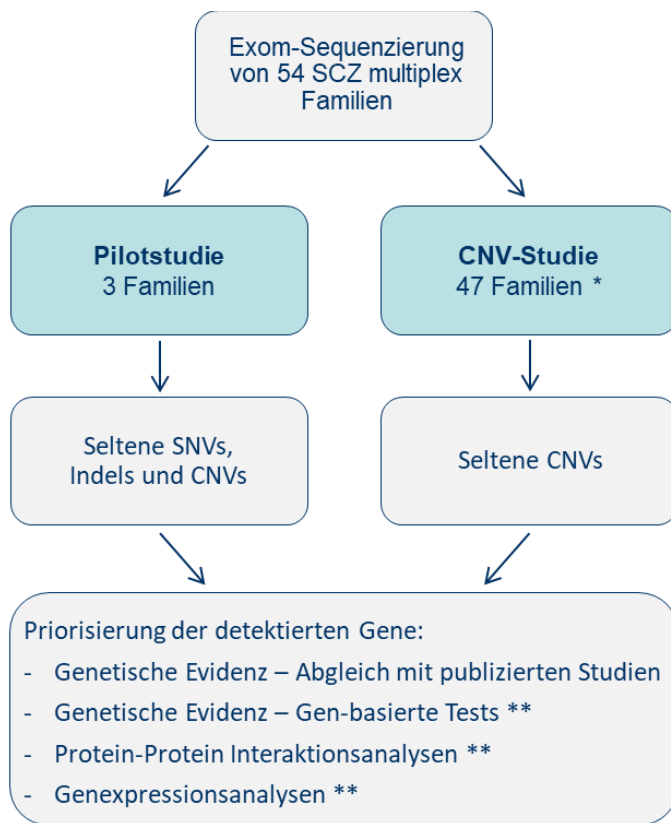


Abbildung 6 Gliederung der Ergebnisse. Die Ergebnisse dieser Arbeit setzen sich aus zwei Hauptteilen zusammen (Pilotstudie und CNV-Studie). In der Pilotstudie lag der Fokus auf seltenen SNVs und Indels sowie Kopienzahlvarianten (CNVs), in der CNV-Studie wurden ausschließlich Kopienzahlvarianten untersucht. SCZ: Schizophrenie; SNVs: Einzelnukleotid-Varianten; Indels: Insertionen und Deletionen; *Vier Familien wurden von der CNV-Studie ausgeschlossen; **Die Gen-basierten Tests, Protein-Protein Interaktionsanalysen und Genexpressionsanalysen wurden in Zusammenarbeit mit verschiedenen Kooperationspartnern durchgeführt.

Insgesamt wurden exomweite Daten für 54 Familien generiert. In einer Familie wurde eine bekannte krankheitsrelevante CNV (Deletion in der chromosomalen Region 2p16.3; *NRXN1*) identifiziert. Die Daten dieser Familie wurden in den weiteren Analysen nicht berücksichtigt. Drei Familien wurden von der CNV-Analyse ausgeschlossen, da von diesen keine genomische DNA vorhanden war.

Die Gen-basierten Tests wurden in Kooperation mit Dr. Till Andlauer (Technische Universität München; Max-Planck-Institut für Psychiatrie, München), die Protein-Protein Interaktionsanalysen mit Christian Hänig, Prof. Erich Wanker (Max-Delbrück-Centrum Berlin) und Dr. Sandra Van der Auwera (Universität Greifswald; Deutsches Zentrum für Neurodegenerative Erkrankungen Rostock/Greifswald) und die Expressionsanalysen mit Dr. Lalit Kaurani und Prof. André Fischer (Universität Göttingen) durchgeführt.

4.1 Familien-Pilotstudie

In der Pilotstudie wurden insgesamt 12 Individuen aus drei multiplex Familien exomweit sequenziert (n = 9 Patienten mit SCZ, n = 3 gesund). Diese wurden sowohl auf Einzelnukleotid-Varianten als auch auf CNVs untersucht.

Identifizierte Einzelnukleotid-Varianten und Indels

Nach Qualitätskontrolle waren im Datensatz 31 Einzelnukleotid-Varianten und 4 Indels nachweisbar (Tabelle 2; Informationen zu den Allelfrequenzen sowie die Ergebnisse der *in silico* Vorhersagen sind in den Tabellen A3 und A4 dargestellt). Von diesen 35 Varianten wurden 34 experimentell verifiziert (n = 27 durch Genotypisierung mittels Massenarray-System von Agena Bioscience und n = 7 mittels Sanger Sequenzierung). Aus technischen Gründen war die Verifikation der Variante c.1924G>T im Gen *KRBA1* auf Chromosom 7 nicht möglich. Da diese Variante jedoch in den Readbrowsern von Varbank und CLC visuell eindeutig zu identifizieren war, wurde diese als richtig positiv eingeschätzt und in die weiteren Analysen eingeschlossen.

Tabelle 2 Identifizierte Einzelnukleotid-Varianten und Indels

Chromosom	Start	Gen	Mutation	AS-Austausch
1	237 730 032	<i>RYR2</i>	c.3380A>G	p.E1127G
3	75 714 805	<i>FRG2C</i>	c.463_464delinsA	p.G155Sfs*11
5	128 990 110	<i>ADAMTS19</i>	c.2270C>T	p.A757V
7	56 046 077	<i>GBAS</i>	c.267C>A	p.Y89*
X	150 844 501	<i>PASD1</i>	c.2209_2209delC	p.P737Lfs*17
X	69 749 756	<i>TEX11</i>	c.2659G>A	p.E887K
1	247 614 645	<i>OR2B11</i>	c.640C>T	p.P214S
2	127 958 737	<i>CYP27C1</i>	c.349C>T	p.R117W
9	86 571 258	<i>C9orf64</i>	c.158A>C	p.H53P
9	86 616 880	<i>RM11</i>	c.980_997delAGAAAGAACAGATGGAAA	p.Q327_T333delinsP
13	100 622 624	<i>ZIC5</i>	c.1306G>C	p.G436R
14	53 003 456	<i>TXNDC16</i>	c.373A>G	p.I125V
16	2 813 480	<i>SRRM2</i>	c.2951C>T	p.P984L
19	17 395 054	<i>ANKLE1</i>	c.1244C>T	p.T541I
19	9 639 109	<i>ZNF426</i>	c.1498G>A	p.G500R
20	1 161 977	<i>TMEM74B</i>	c.286C>T	p.R96*
22	25 753 280	<i>LRP5L</i>	c.380C>T	p.A127V
X	153 297 785	<i>MECP2</i>	c.250C>T	p.R84W
3	197 409 416	<i>KIAA0226</i>	c.1916G>A	p.R639Q
6	136 268 632	<i>PDE7B</i>	c.52G>A	p.D18N
6	153 345 582	<i>RGS17</i>	c.259A>G	p.K87E
7	150 713 551	<i>ATG9B</i>	c.2490G>A	p.M830I
7	97 923 411	<i>BAIAP2L1</i>	c.1435G>A	p.G479R
7	102 401 858	<i>FAM185A</i>	c.442G>A	p.G148S
7	149 427 353	<i>KRBA1</i>	c.1924G>T †	p.G642C
7	150 698 668	<i>NOS3</i>	c.1465G>A	p.G489S
10	70 333 117	<i>TET1</i>	c.1022C>T	p.A341V
12	53 227 659	<i>KRT79</i>	c.386T>C	p.V129A
13	95 243 187	<i>TGDS</i>	c.229_232del(AT)2insAT	p.I77Mfs*2
14	38 061 099	<i>FOXA1</i>	c.890C>T	p.P297L
15	72 557 506	<i>PARP6</i>	c.244G>A	p.V82I
17	71 232 045	<i>C17orf80</i>	c.424G>C	p.A142P
17	42 168 803	<i>HDAC5</i>	c.1225C>T	p.R409W
20	23 350 822	<i>GZF1</i>	c.1880A>G	p.E627G
20	42 157 324	<i>L3MBTL1</i>	c.824C>T	p.S275L

Die kodierende Position basiert auf dem längsten proteinkodierenden Ensembl-Transkript des jeweiligen Gens (GRCh37). AS: Aminosäure; † Variante aus technischen Gründen nicht experimentell verifiziert.

Die 34 experimentell verifizierten Varianten wurden in allen 20 Individuen der drei Familien mittels Massensystem oder Sanger Sequenzierung getestet. Keine wurde in einem Mitglied der übrigen beiden Familien identifiziert. Ebenso waren die Varianten nicht in den gesunden Ehepartnern nachweisbar. In einer Familie waren 12 der 17 identifizierten Varianten auch in mindestens einem der gesunden Familienmitglieder vorhanden, und 16 Varianten wurden in mindestens einem der Familienmitglieder identifiziert, bei denen eine schwere Depression diagnostiziert wurde.

Genomweite Analyse von Kopienzahlvarianten

In den drei untersuchten Familien wurden insgesamt zwei CNVs (eine Duplikation auf Chromosom 3p21.31 und eine Duplikation auf Chromosom 8q21.13) detektiert, welche die Filterkriterien erfüllten.

Duplikationen auf Chromosom 3p21.31

Duplikationen in der chromosomalen Region 3p21.31 wurden in allen drei Patienten mit SCZ, sowie dem obligaten Konduktor einer Familie identifiziert und mittels qPCR verifiziert (Abbildung 7 und Tabelle 3). Die verifizierte, 51,67 kb umfassende Region (chr3:47.795.268-47.846.934, hg19) betraf in allen vier untersuchten Familienmitgliedern partiell die Gene *SMARCC1* und *DHX30*. Diese CNV wurde weder in einer weiteren multiplex Familie noch in dem hausinternen Datensatz identifiziert.

In ExAC wurden zwei nicht-finnisch europäische (NFE, inkl. psychiatrischer Kohorte) Individuen mit einer partiellen Duplikation von *SMARCC1* und *DHX30* beschrieben (Abbildung 7). Der ExAC Datensatz beinhaltet Kohorten mit psychiatrischer Diagnose, unter anderem Schizophrenie. Ein Ausschluss dieser Kohorten war aus technischen Gründen nicht möglich.

Die Wahrscheinlichkeit der Triplo-Sensitivität (pTriplo) betrug für das Gen *SMARCC1* 0,89 und für *DHX30* 1,0, was auf eine Intoleranz der Gene gegenüber Duplikationen hinweist (Collins et al., 2022).

Duplikationen auf Chromosom 8q21.13

In der chromosomalen Region 8q21.13 wurden Duplikationen bei allen drei SCZ Patienten einer Familie identifiziert und mittels qPCR verifiziert (Abbildung 8 und Tabelle 4). Ein gesundes Familienmitglied konnte aufgrund mangelnder DNA nicht mittels qPCR untersucht werden. Allerdings war die Duplikation *in silico* durch die verwendeten CNV-Algorithmen bei dieser Person nicht identifizierbar.

Die verifizierte Region umfasste bei allen drei CNV Trägern 87,87 kb (chr8:80.876.034-80.963.908, hg19) und betraf partiell die beiden Gene *MRPS28* und *TPD52*. Eine Duplikation in derselben Region konnte auch in einem der SCZ Patienten aus dem hausinternen Datensatz detektiert und verifiziert werden (verifizierte Region chr8:80.876.034-80.963.908, hg19). Es wurde keine CNV in dieser Region in einer der anderen untersuchten Familien gefunden.

In ExAC wurde eine Duplikation in einem NFE Individuum beschrieben, welche ebenfalls die beiden Gene *MRPS28* und *TPD52* partiell betraf (Abbildung 8). Die Wahrscheinlichkeit der Triplo-Sensitivität betrug für *MRPS28* 0.29 und für *TPD52* 0,17, was auf eine geringe Intoleranz der beiden Gene gegenüber Duplikationen hinweist (Collins et al., 2022).

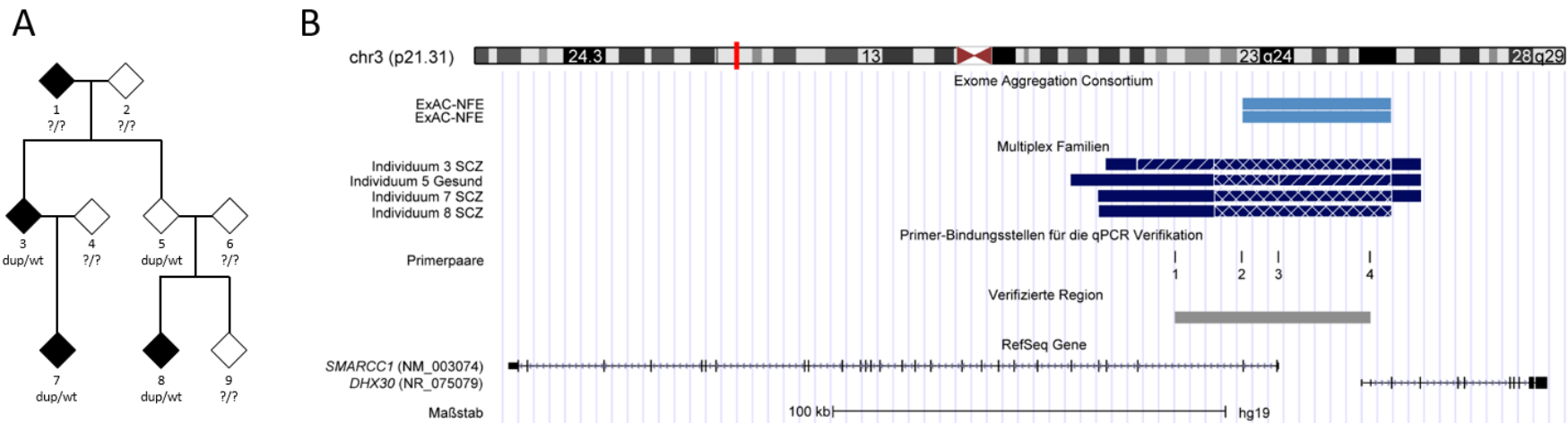


Abbildung 7 Duplikationen auf Chromosom 3p21.31

A) Stammbaum der Familie, in der Duplikationen in der chromosomalen Region 3p21.32 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; dup: Duplikation; wt: Wildtyp; ??/? nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012), XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in drei Familienmitgliedern mit Schizophrenie (SCZ; 3, 7, 8) und einem gesunden Familienmitglied (5) detektiert wurden. Die Nummerierung der einzelnen Individuen erfolgte entsprechend des Stammbaums. ■ = CoNIFER; ▨ = XHMM; ▩ = ExomeDepth. Zwei partielle Duplikationen von *SMARCC1* und *DHX30* wurden in den Daten des Exome Aggregation Consortium (Lek et al., 2016) beschrieben (hellblau). NFE: nicht-finnischer Europäer. Vier Primerpaare (1-4) wurden für die Verifikation mittels quantitativer PCR (qPCR) verwendet. Die durch qPCR verifizierte Region ist in grau abgebildet. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19.

Tabelle 3 Duplikationen auf Chromosom 3p21.31

	CNV-Algorithmus	Start	Stop	Länge in kb	Gene
Individuum 3 SCZ	CoNIFER	47.779.504	47.859.652	80,148	SMARCC1, DHX30
	XHMM	47.787.366	47.852.229	64,863	
	ExomeDepth	47.807.067	47.852.229	45,162	
	Verifizierte Region	47.795.268	47.846.934	51,666	
Individuum 5 Gesund	CoNIFER	47.770.466	47.859.652	89,186	SMARCC1, DHX30
	XHMM	47.807.067	47.852.229	45,162	
	ExomeDepth	47.807.067	47.823.671	16,604	SMARCC1
	Verifizierte Region	47.795.268	47.846.934	51,666	SMARCC1, DHX30
Individuum 7 SCZ	CoNIFER	47.777.486	47.859.652	82,166	SMARCC1, DHX30
	XHMM	47.807.067	47.852.229	45,162	
	ExomeDepth	47.807.067	47.852.229	45,162	
	Verifizierte Region	47.795.268	47.846.934	51,666	
Individuum 8 SCZ	CoNIFER	47.777.486	47.852.229	74,743	SMARCC1, DHX30
	XHMM	47.807.067	47.852.229	45,162	
	ExomeDepth	47.807.067	47.852.229	45,162	
	Verifizierte Region	47.795.268	47.846.934	51,666	

Aufgeführt sind Loci der Kopienzahlvarianten (CNV), welche von den drei Algorithmen CoNIFER (Krumm et al., 2012), XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in drei Familienmitgliedern mit Schizophrenie (SCZ; 3, 7, 8) und einem gesunden Familienmitglied (5) detektiert wurden. Die Nummerierung der einzelnen Individuen erfolgte entsprechend des Stammbaums (Abbildung 7A). Verifizierte Region: Region, welche mittels quantitativer PCR verifiziert wurde. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. kb: Kilobasen.

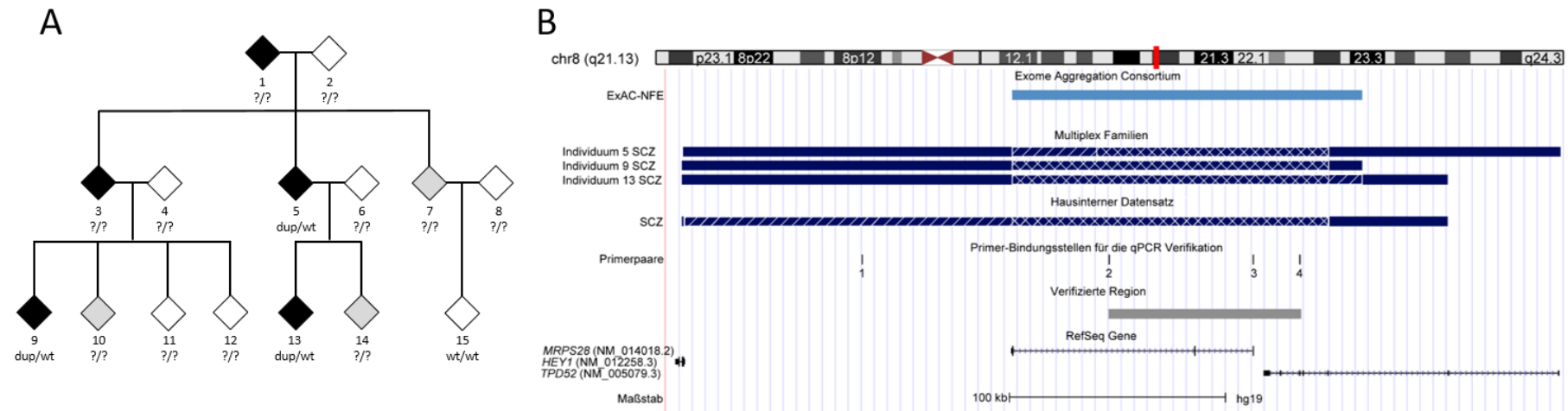


Abbildung 8 Duplikationen auf Chromosom 8q21.13

A) Stammbaum der Familie, in der Duplikationen in der chromosomalen Region 8q21.13 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; graue Symbole: Individuen mit anderer psychiatrischer Erkrankung; dup: Duplikation; wt: Wildtyp; ?? nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012),XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in drei Familienmitgliedern mit Schizophrenie (SCZ; 5, 9, 13) und einem SCZ Patienten aus dem hausinternen Datensatz detektiert wurden. Die Nummerierung der einzelnen Individuen erfolgte entsprechend des Stammbaums. ■ = CoNIFER; ▨ = XHMM; ▩ = ExomeDepth. Eine partielle Duplikation von *MRPS28* und *TPD52* wurde in den Daten des Exome Aggregation Consortium (ExAC; Lek et al., 2016) in einem NFE Individuum beschrieben (hellblau). NFE: nicht-finnische Europäer. Vier Primerpaare (1-4) wurden für die Verifikation mittels quantitativer PCR (qPCR) verwendet. Die durch qPCR verifizierte Region ist in grau abgebildet. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19.

Tabelle 4 Duplikationen auf Chromosom 8q21.13

	CNV-Algorithmus	Start	Stop	Länge in kb	Gene
Individuum 5 SCZ	CoNIFER	80.679.209	81.083.559	404,35	HEY1, MRPS28, TPD52
	XHMM	80.831.217	80.977.122	145,905	MRPS28, TPD52
	ExomeDepth	80.870.594	80.977.122	106,528	
	Verifizierte Region	80.876.034	80.963.908	87,874	
Individuum 9 SCZ	CoNIFER	80.678.862	80.992.733	313,871	HEY1, MRPS28, TPD52
	XHMM	80.831.217	80.977.122	145,905	MRPS28, TPD52
	ExomeDepth	80.831.217	80.977.122	145,905	
	Verifizierte Region	80.876.034	80.963.908	87,874	
Individuum 13 SCZ	CoNIFER	80.678.862	81.032.212	353,35	HEY1, MRPS28, TPD52
	XHMM	80.831.217	80.992.733	161,516	MRPS28, TPD52
	ExomeDepth	80.831.217	80.977.122	145,905	
	Verifizierte Region	80.876.034	80.963.908	87,874	
Hausinterner Datensatz					
SCZ	CoNIFER	80.678.862	81.032.212	353,35	HEY1, MRPS28, TPD52
	XHMM	80.679.210	80.977.122	297,912	
	ExomeDepth	80.831.217	80.977.122	145,905	MRPS28, TPD52
	Verifizierte Region	80.876.034	80.963.908	87,874	

Aufgeführt sind Loci der Kopienzahlvarianten (CNV), welche von den drei Algorithmen CoNIFER (Krumm et al., 2012), XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in drei Familienmitgliedern mit Schizophrenie (SCZ; 5, 9, 13) und einem SCZ Patienten aus dem hausinternen Datensatz detektiert wurden. Die Nummerierung der einzelnen Individuen erfolgte entsprechend des Stammbaums (Abbildung 8A). Verifizierte Region: Region, welche mittels quantitativer PCR verifiziert wurde. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. kb: Kilobasen.

Tabelle 5 Zusätzliche genetische Evidenz für die 39 Kandidatengene der Pilotstudie

Gen	Katatonie Locus	CNV-Studie	GWAS	Publizierte SCZ NGS Studien	Genebook NS strict p	Genebook NS broad p	Hausinterner Datensatz	Gene-based test p-Wert	Gene-based test p-Wert; korrigiert
<i>ADAMTS19</i>	-	-	-	Howrigan et al. 2020; c.61G>A	0.21	0.34	2 Patienten c.1118G>A, c.3176G>A	0.0239	
<i>ANKLE1</i>	-	-	-	-	0.66	0.13	-	0.5335	
<i>ATG9B</i>	-	-	-	-	1	1	-	0.4515	
<i>BAIAP2L1</i>	-	-	-	-	1	1	-	0.2458	
<i>C17orf80</i>	-	-	-	-	0.66	0.16	-	0.9321	
<i>C9orf64</i>	-	-	-	-	1	0.7	-	0.0265	
<i>CYP27C1</i>	-	-	-	-	0.55	1	-	0.1818	
<i>FAM185A</i>	-	-	-	-	-	-	-	0.2827	
<i>FOXA1</i>	-	-	-	-	0.83	1	-	0.2657	
<i>FRG2C</i>	-	-	-	-	-	-	-	- *	
<i>GBAS</i>	-	-	-	-	1	1	-	0.0312	
<i>GZF1</i>	-	-	-	Toyama et al. 2022; c.2021A>G	0.74	0.11	-	0.9600	
<i>HDAC5</i>	-	-	-	Howrigan et al. 2020; c.3275T>C	0.53	0.41	-	0.4216	
<i>KIAA0226</i>	-	-	-	-	0.61	0.3	1 Patient c.1484G>A	0.4575	
<i>KRBA1</i>	-	-	-	-	-	-	-	0.5325	
<i>KRT79</i>	-	-	-	-	0.11	0.002	-	0.6933	
<i>L3MBTL1</i>	-	-	-	John et al. 2018; c.635T>G	1	1	-	0.5684	
<i>LRP5L</i>	-	-	-	-	0.41	1	-	0.8701	
<i>MECP2</i>	-	-	-	McCarthy et al. 2014 / Howrigan et al. 2020; c.568C>T / c.419C>A	0.5	0.24	1 Patient c.253C>T	0.0002	0.0057
<i>NOS3</i>	-	-	-	Fromer et al. 2014; c.2821T>G	0.38	1	-	0.2927	
<i>OR2B11</i>	-	-	-	-	-	1	-	0.7433	
<i>PARP6</i>	-	-	-	Howrigan et al. 2020; c.1514A>G	0.84	1	-	0.0979	

Gen	Katatonie Locus	CNV-Studie	GWAS	Publizierte SCZ NGS Studien	Genebook NS strict p	Genebook NS broad p	Hausinterner Datensatz	Gene-based test p-Wert	Gene-based test p-Wert; korrigiert
<i>PASD1</i>	-	-	-	-	-	0.41	-	0.5005	
<i>PDE7B</i>	-	-	-	-	1	1	-	0.0381	
<i>RGS17</i>	-	-	-	-	0.52	0.39	-	0.0093	
<i>RMI1</i>	-	-	-	-	0.03	0.02	-	0.1988	
<i>RYR2</i>	-	-	-	Ambalavanan et al. 2015 / Howrigan et al. 2020 / Toyama et al. 2022; c.2236_2238delinsTAC / c.14072A>G / c.4652A>G	0.07	0.22	2 Patienten c.12541G>A, c.13178T>C	0.0129	
<i>SRRM2</i>	-	-	-	Toyama et al. 2022; c.6233G>A	1	1	1 Patient c.6278C>T	0.6773	
<i>TET1</i>	-	-	-	-	0.53	0.18	1 Patient c.1597_1598delCT	0.0084	
<i>TEX11</i>	-	-	-	-	-	0.37	-	0.4765	
<i>TGDS</i>	-	-	-	-	1	1	-	0.9301	
<i>TMEM74B</i>	-	-	-	-	-	0.67	-	0.4535	
<i>TXNDC16</i>	-	-	-	-	1	1	-	0.0046	
<i>ZIC5</i>	-	-	-	-	0.69	0.35	-	0.3896	
<i>ZNF426</i>	-	-	-	Fromer et al. 2014; c.212G>A	0.21	1	-	0.0949	
<i>DHX30</i>	-	-	-	Howrigan et al. 2020; c.2830G>A	-	1	-	0.0899	
<i>SMARCC1</i>	-	-	-	-	0.33	0.09	1 Patient c.2231C>G	0.1469	
<i>MRPS28</i>	-	-	-	-	0.23	0.42	1 Patient (Duplikation)	0.5924	
<i>TPD52</i>	-	-	-	-	1	1		0.3307	

Um zusätzliche genetische Evidenz für die Kandidatengene zu erhalten, wurde überprüft, ob diese auch in anderen Datensätzen identifiziert wurden. Zu diesen zählten der hausinterne Datensatz mit 141 Patienten mit Schizophrenie (SCZ) und verschiedene öffentlich zugängliche Datensätze. Dazu zählten (i) bekannte Katatonie-assoziierte Risikoloci (Stöber et al., 2000), (ii) bekannte SCZ-assoziierte Kopienzahlvarianten (CNVs; Marshall et al. 2017), (iii) Gene, in denen in einer der größten genomweiten Assoziationsstudien (GWAS) ein genomweit signifikanter Einzelnukleotid-Polymorphismus beschrieben wurde (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Pardiñas et al., 2018; Trubetskoy et al., 2022) und (iv) Gene, die in anderen SCZ Exom- und Genom-Sequenzierungsstudien identifiziert wurden (Girard et al., 2011; Xu et al., 2011, 2012; Timms et al., 2013; Gulsuner et al., 2013; Guipponi et al., 2014; Fromer et al., 2014; Takata et al., 2014; McCarthy et al., 2014; Kranz et al., 2015; Wang et al., 2015; Ambalavanan et al., 2016; Watanabe et al., 2016; Egawa et al., 2016; Homann et al., 2016; Zhou et al., 2016; Smedemark-Margulies et al., 2016; Hornig et al., 2017; John et al., 2017; Kos et al., 2017; Li et al., 2021; Howrigan et al., 2020; Balakrishna & Curtis, 2020; Rees et al., 2020; Chen et al., 2021; Toyama et al., 2022; John et al., 2018, 2019; Shang et al., 2024). NGS: *next generation sequencing*; NS (*nonsynonymous*) strict: trunkierende und *missense* Varianten, die im Genebook (Purcell et al. 2014) von fünf Algorithmen als pathogen klassifiziert wurden (<http://atgu.mgh.harvard.edu/~spurcell/genebook/genebook.cgi?user=guest&cmd=overview#>); NS broad: trunkierende und *missense* Varianten, die im Genebook von mindestens einem Algorithmus als pathogen klassifiziert wurden. Außerdem wurden Gen-basierte Tests basierend auf den Daten der PGC SCZ GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) durchgeführt, - * nicht vorhanden in VEGAS (Mishra & Macgregor, 2015); - nicht beschrieben/keine Daten vorhanden. In der Spalte „Gene-based test p-Wert“ sind alle p-Werte vor der Korrektur für multiples Testen aufgeführt. Der in fett hervorgehobene p-Wert blieb auch nach der Korrektur signifikant.

Genetische Evidenz

Die 39 Kandidatengene der Pilotstudie wurden in verschiedenen öffentlich zugänglichen Datensätzen sowie einem unabhängigen hausinternen Datensatz überprüft. Zusätzlich wurden Gen-basierte Tests durchgeführt (Tabelle 5). In sieben Genen (*ADAMTS19*, *KIAA0226*, *MECP2*, *RYR2*, *SRRM2*, *TET1* und *SMARCC1*) wurden seltene SNVs im hausinternen Datensatz detektiert und in zwei Genen (*MRPS28* und *TPD52*) eine Duplikation. In elf der 39 Gene (*ADAMTS19*, *GZF1*, *HDAC5*, *L3MBTL1*, *MECP2*, *NOS3*, *PARP6*, *RYR2*, *SRRM2*, *ZNF426* und *DHX30*) wurden Mutationen in anderen SCZ Exom- und Genom-Sequenzierungsstudien beschrieben. Für neun der 38 untersuchten Kandidatengene (das Gen *FRG2C* war nicht Teil der VEGAS2 Datenbank) wurde bei den Gen-basierten Tests eine Assoziation ($p < 0,05$) zwischen häufigen Varianten und einem Risiko für SCZ identifiziert. Jedoch blieb die Assoziation nur für ein Gen – *MECP2* – auch nach der Korrektur für multiples Testen mittels Bonferroni Methode signifikant (p korrigiert = 0,0057; Tabelle 5).

Protein-Protein Interaktionsanalysen

Mit Hilfe der HIPPIE-Datenbank wurden zwei PPI-Netzwerke erstellt: cCandidate enthielt alle Proteine, die von Kandidatengenen aus der vorliegenden Studie kodiert werden und deren Interaktionen. Das zweite Netzwerk, cHIPPIE_without_cCandidate, enthielt alle in HIPPIE verfügbaren PPIs ohne die Interaktionen aus dem Netzwerk cCandidate. Anschließend wurde überprüft, ob cCandidate im Vergleich zu cHIPPIE_without_cCandidate einen erhöhten Anteil an Proteinen, die von bereits bekannten SCZ Kandidatengenen aus der Literatur kodiert werden, enthielt. Es konnte mit beiden verwendeten Tests eine signifikante Anreicherung nachgewiesen werden (Chi-Quadrat-Test (McHugh, 2013); $p = 0.0022$; Exakter Fisher-Test; $p = 0.0039$).

Anschließend wurde mittels der Software Cytoscape (Shannon et al., 2003) eine Abbildung aller PPIs zwischen den von Kandidatengenen aus der vorliegenden Studie kodierten Proteinen und den Proteinen, welche von bekannten SCZ Kandidatengenen aus der Literatur kodiert werden, erstellt. Dieses Netzwerk enthielt insgesamt 48 Proteine, welche durch 43 Interaktionen verbunden waren (Abbildung 9). Drei der Kandidatengene aus der vorliegenden Studie wiesen mehr als drei Interaktionen auf. Unter diesen waren auch die chromatinregulierenden Gene *HDAC5* und *SMARCC1* (Shahbazian & Grunstein, 2007; Alver et al., 2017).

Für eine zweite, unabhängige Netzwerk-basierte Priorisierung der Kandidatengene wurde das Tool ToppNet verwendet (Chen et al., 2007, 2009). Die höchsten Scores erhielten hierbei die Proteine *HDAC5*, *FOXA1*, *SRRM2*, *MECP2* und *SMARCC1* (Tabelle 6).

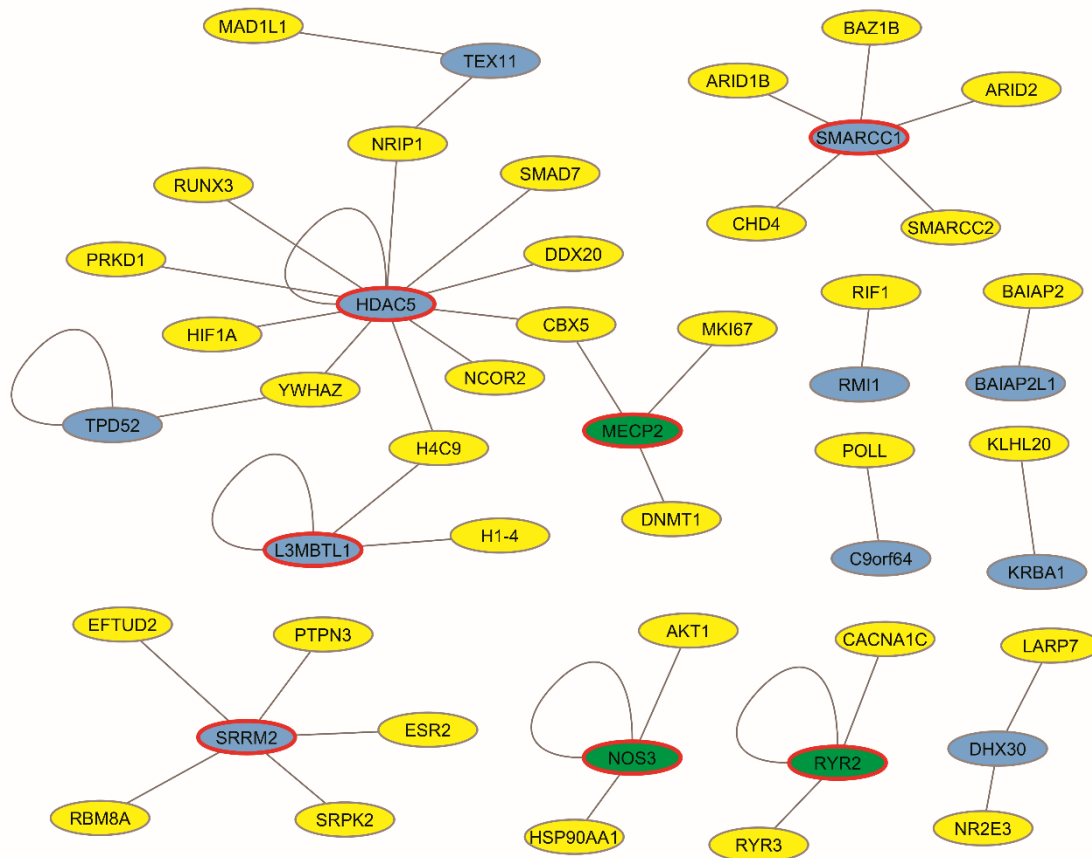


Abbildung 9 Protein-Protein Interaktionsanalysen

Schematische Darstellung der Interaktionen zwischen den Proteinen der in der vorliegenden Studie identifizierten Kandidatengene und Proteinen von Genen, die in anderen Studien mit Schizophrenie (SCZ) assoziiert wurden. Insgesamt sind 43 Interaktionen zwischen 48 Proteinen dargestellt. Knoten weisen auf Proteine hin und Kanten auf Protein-Protein Interaktionen. Blau: Proteine, die von neu identifizierten SCZ Kandidatengenen kodiert werden. Gelb: Proteine, die von bereits beschriebenen SCZ Kandidatengenen kodiert werden; Grün: Proteine, die von bereits beschriebenen SCZ Kandidatengenen kodiert werden, die zusätzlich auch in der Pilotstudie identifiziert wurden; Die rote Umrandung hebt Proteine mit mehr als zwei Interaktionen hervor.

Tabelle 6 Priorisierung der Kandidatengene mittels ToppNet

Rang	Gen ID	Gen	Anzahl PPIs	Score
1	10014	<i>HDAC5</i>	358	3.542E-4
2	3169	<i>FOXA1</i>	295	3.273E-4
3	23524	<i>SRRM2</i>	281	2.606E-4
4	6599	<i>SMARCC1</i>	168	1.752E-4
5	4204	<i>MECP2</i>	162	1.730E-4
6	22907	<i>DHX30</i>	124	1.081E-4
7	26013	<i>L3MBTL1</i>	90	9.269E-5
8	56159	<i>TEX11</i>	72	7.042E-5
9	28957	<i>MRPS28</i>	77	6.092E-5
10	7163	<i>TPD52</i>	52	5.878E-5
11	4846	<i>NOS3</i>	55	5.362E-5
12	6262	<i>RYR2</i>	28	5.358E-5
13	55971	<i>BAIAP2L1</i>	41	4.781E-5
14	84267	<i>C9orf64</i>	16	3.250E-5
15	338785	<i>KRT79</i>	26	2.777E-5
16	55028	<i>C17orf80</i>	26	2.718E-5
17	57544	<i>TXNDC16</i>	25	2.599E-5
18	84626	<i>KRBA1</i>	23	2.516E-5
19	80010	<i>RMI1</i>	27	2.383E-5
20	79088	<i>ZNF426</i>	17	2.057E-5
21	80312	<i>TET1</i>	17	1.785E-5
22	26575	<i>RGS17</i>	18	1.582E-5
23	64412	<i>GZF1</i>	18	1.391E-5
24	56965	<i>PARP6</i>	10	7.821E-6
25	23483	<i>TGDS</i>	6	7.453E-6
26	139135	<i>PASD1</i>	2	3.202E-6
27	285973	<i>ATG9B</i>	3	2.878E-6
28	27115	<i>PDE7B</i>	2	2.846E-6
29	339761	<i>CYP27C1</i>	2	2.545E-6
30	85416	<i>ZIC5</i>	2	2.291E-6
31	91355	<i>LRP5L</i>	2	1.932E-6

Die Tabelle zeigt die neu identifizierten Kandidatengene der Pilotstudie priorisiert mittels ToppNet (Chen et al., 2007, 2009) nach ihrer Konnektivität zu bekannten Schizophrenie Kandidatengenen aus der Literatur. ID: Entrez Gene ID (<http://www.ncbi.nlm.nih.gov/gene>). Gelistet sind nur Gene mit mindestens einer Protein-Protein Interaktion in der ToppNet Datenbank.

Genexpression

Von den 39 Kandidatengenen waren basierend auf dem GTEx (GTEx Consortium, 2013) Datensatz 28 Gene im Gehirn exprimiert (> 1 Fragment pro Kilobase Transkript pro eine Million gemappter Reads) und 25 davon im Hippocampus (Tabelle A5).

Im zweiten Teil der Expressionsanalyse waren 29 der 39 Gene in mindestens einer der 15 verfügbaren publizierten Studien (Tabelle A7; Catts et al., 2014; Chen et al., 2013; de Baumont et al., 2015; Fillman et al., 2013; Hwang et al., 2013; Iwamoto et al., 2004, 2005; Kohen et al.,

2014; Lanz et al., 2015; Maycox et al., 2009; Narayan et al., 2008; Pietersen, Mauney, Kim, Lim, et al., 2014; Pietersen, Mauney, Kim, Passeri, et al., 2014; Wu et al., 2012; Zhao et al., 2015) zwischen Patienten mit SCZ und Kontrollen differenziell exprimiert.

Im hausinternen Datensatz (In.house) waren drei Gene (*FOXA1*, *L3MBTL1* und *PARP6*) zwischen Patienten mit SCZ und Kontrollen differenziell exprimiert (Abbildung 10, Tabelle A6).

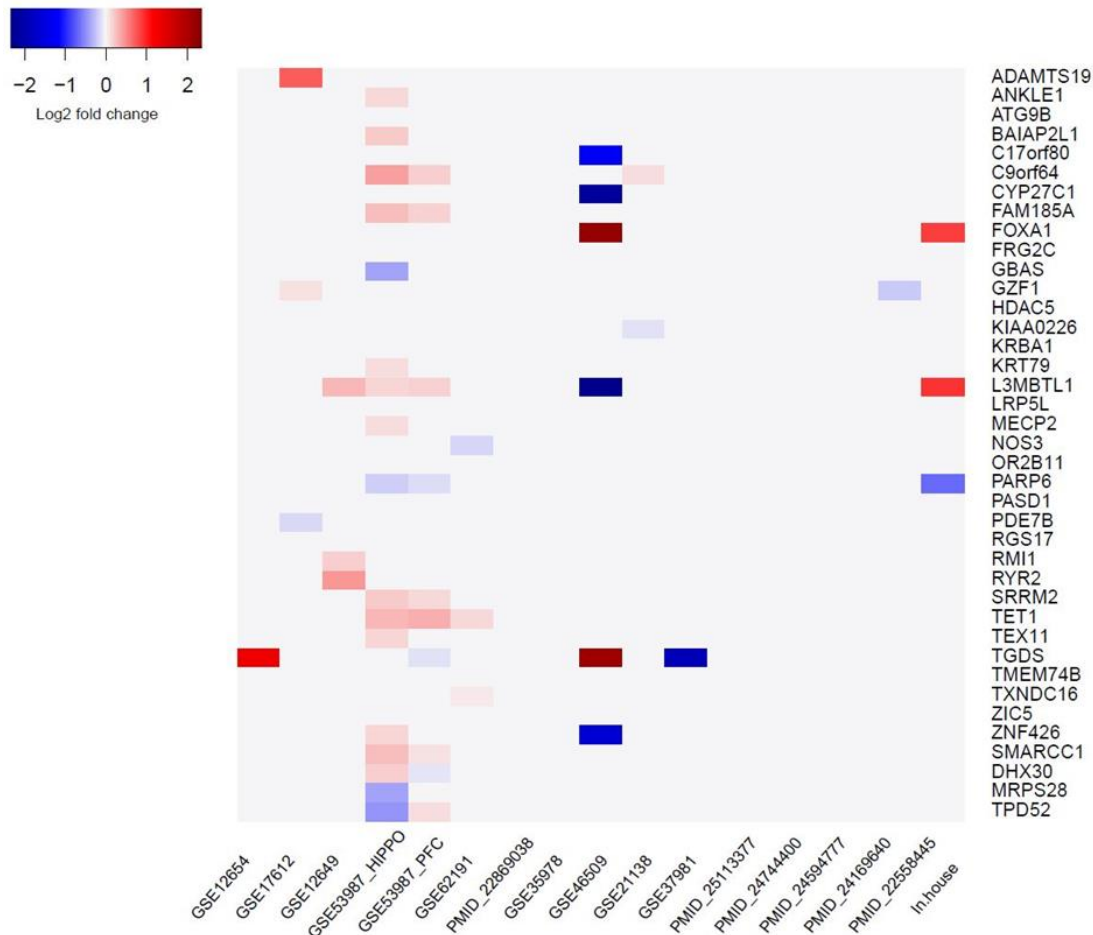


Abbildung 10 Differentielle Genexpression bei Patienten mit Schizophrenie und Kontrollen

Abgebildet ist die differentielle Genexpression bei Patienten mit Schizophrenie und gesunden Kontrollen in 15 verschiedenen Studien (Tabelle A7; Catts et al., 2014; Chen et al., 2013; de Baumont et al., 2015; Fillman et al., 2013; Hwang et al., 2013; Iwamoto et al., 2004, 2005; Kohen et al., 2014; Lanz et al., 2015; Maycox et al., 2009; Narayan et al., 2008; Pietersen, Mauney, Kim, Lim, et al., 2014; Pietersen, Mauney, Kim, Passeri, et al., 2014; Wu et al., 2012; Zhao et al., 2015) sowie einem internen Datensatz (In.house). In einer Studie (GSE53987; Lanz et al., 2015) wurden zwei verschiedene Hirnregionen (HIPPO; Hippocampus und frontaler Cortex; PFC) analysiert. Die Farben zeigen den log2-fold change an. log2-fold change > 0: Die Expression war bei den Patienten signifikant höher als bei den Kontrollprobanden. log2-fold change < 0: Die Expression war bei den Patienten signifikant niedriger als bei den Kontrollprobanden.

4.2 CNV-Studie

Insgesamt wurden in den 47 analysierten multiplex Familien acht CNVs identifiziert, die die Filterkriterien erfüllten: Fünf Deletionen in den chromosomalen Regionen 1p36.13, 14q11.2, 14q22.2, 14q24.3 und 16q23.2 sowie drei Duplikationen in den Regionen 6p12.1, 9q33.2 und 12q24.31. In einer weiteren Familie wurde eine Deletion in einem bekannten, mit SCZ assoziierten Locus (2q16.3-p16.2; *NRXN1*) identifiziert. Diese CNV wurde von den folgenden Analysen ausgeschlossen.

Keine der CNVs kam in mehr als einer der 47 Familien vor, jedoch wurden in einer Familie zwei unterschiedliche CNVs (Deletion in 14q11.2 und Duplikation in 6p12.1) identifiziert. Sechs CNVs (Deletionen in 1p36.13, 14q11.2, 14q22.2 und 14q24.3 und Duplikationen in 9q33.2 und 12q24.31) wurden mittels qPCR verifiziert. Eine CNV (Deletion in 16q23.2) wurde visuell basierend auf den Daten der genomweiten Genotypisierung bestätigt. Die Duplikation in 6p12.1 konnte aufgrund nicht ausreichender DNA und einer zu geringen Größe (~39kb) weder experimentell noch visuell überprüft werden.

Tabelle 7 Durch genomweite CNV-Analyse identifizierte Loci

Locus	XHMM	CoNIFER	ExomeDepth	Verifiziert
1p36.13 CLCNKA CLCNKB	Deletion Deletion	Deletion Deletion	Deletion Deletion	Deletion Deletion
14q11.2 <i>EDDM3B</i> <i>RNASE6</i> <i>RNASE1</i> <i>RNASE3</i> <i>RNASE2</i> <i>METTL17</i>	2n 2n 2n Deletion Deletion 2n	Deletion Deletion Deletion Deletion Deletion Deletion	2n 2n 2n Deletion 2n 2n	- - - Deletion Deletion -
14q22.2 <i>GMFB</i> <i>CGRRF1</i> <i>SAMD4A</i>	2n Deletion Deletion	Deletion Deletion Deletion	2n Deletion 2n	- Deletion 2n
14q24.3 <i>HEATR4</i> <i>ACOT1</i> <i>ACOT2</i>	2n 2n Deletion	Deletion Deletion Deletion	2n 2n Deletion	- - Deletion
16q23.2 <i>PKD1L2</i> <i>BCMO1</i>	Deletion Deletion	Deletion Deletion	Deletion Deletion	Deletion Deletion
6p12.1 <i>KIAA1586</i> <i>ZNF451</i>	Duplikation Duplikation	Duplikation Duplikation	Duplikation Duplikation	- -
9q33.2 <i>RAB14</i> <i>GSN</i>	2n Duplikation	Duplikation Duplikation	2n Duplikation	- Duplikation
12q24.31 <i>RSRC2</i> <i>ZCCHC8</i>	Duplikation Duplikation	Duplikation Duplikation	Duplikation Duplikation	Duplikation Duplikation

Aufgeführt sind alle Gene, in denen von mindestens einem der drei Algorithmen (eXome-Hidden Markov Model (XHMM; Fromer et al., 2012), Copy Number Inference From Exome Reads (CoNIFER; Krumm et al., 2012) und ExomeDepth (Plagnol et al., 2012)) eine CNV detektiert wurde. Fett: Gene, die von CNVs betroffen waren, die von allen drei Algorithmen detektiert und/oder experimentell/visuell bestätigt wurden. 2n: Keine Deletion oder Duplikation. - nicht Teil der experimentellen Verifikation.

Deletionen auf Chromosom 1p36.13

In der chromosomalen Region 1p36.13 wurden bei allen drei untersuchten erkrankten Individuen einer Familie (Elternteil mit SCZ und zwei Nachkommen mit SCZ und schizoaffektiver Störung) Deletionen identifiziert (Abbildung 11). Mittels qPCR konnte gezeigt werden, dass die Deletion in allen drei Familienmitgliedern mindestens das Gen *CLCNKA* betraf. In zwei Familienmitgliedern (Elternteil und Nachkomme mit SCZ) wurde eine 17,3 kb (chr1:16.354.549-16.371.832; hg19) große Deletion nachgewiesen, die die Gene *CLCNKA* und *CLCNKB* zum Teil betraf. Im dritten Familienmitglied betraf die Deletion nur das Gen *CLCNKA*. In einem untersuchten gesunden Familienmitglied konnte die Deletion nicht nachgewiesen werden. Von dem zweiten, gesunden Elternteil war keine DNA verfügbar, sodass nicht ausgeschlossen werden konnte, dass die kleinere CNV in dieser Region von dem

gesunden Elternteil vererbt worden war. Weder in den ExAC Daten, noch in dem hausinternen Datensatz wurde eine Deletion in den Genen *CLCNKA* und *CLCNKB* identifiziert.

Deletionen auf Chromosom 14q11.2

In der chromosomalen Region 14q11.2 wurden bei beiden untersuchten Familienmitgliedern mit SCZ (Elternteil und Nachkomme) Deletionen identifiziert (Abbildung 12). Mittels qPCR konnte gezeigt werden, dass die mindestens 27,9 kb (chr14:21.360.151-21.388.066; hg19) große Deletion beim Elternteil die Gene *RNASE3* und *RNASE2*, beim Nachkommen das Gen *RNASE3* betraf. In einem untersuchten gesunden Familienmitglied konnte die Deletion nicht nachgewiesen werden. Von einem dritten Familienmitglied mit SCZ war keine DNA vorhanden, sodass keine CNV-Analyse durchgeführt werden konnte. Weiterhin war keine DNA von dem zweiten, gesunden Elternteil verfügbar, sodass nicht ausgeschlossen werden konnte, dass die kleinere CNV in dieser Region von dem gesunden Elternteil vererbt worden war. Weder in den ExAC Daten, noch in dem hausinternen Datensatz wurde eine Deletion in den Genen *RNASE3* und *RNASE2* identifiziert.

Deletionen auf Chromosom 14q22.2

In der chromosomalen Region 14q22.2 wurde eine Deletion identifiziert, welche von einem Elternteil mit Anpassungsstörung an zwei erkrankte Nachkommen (SCZ und bipolare Störung) vererbt wurde (Abbildung 13). Mittels qPCR konnte nachgewiesen werden, dass die Deletion bei allen drei Familienmitgliedern mindestens das Gen *CGRRF1* betraf, jedoch nicht das Gen *SAMD4A*, in welchem zwei der CNV-Algorithmen ebenfalls eine Deletion detektierten. Weder in den ExAC Daten, noch in dem hausinternen Datensatz wurde eine Deletion in dem Gen *CGRRF1* identifiziert.

Deletionen auf Chromosom 14q24.3

In der chromosomalen Region 14q24.3 wurden in einer Familie Deletionen in zwei Familienmitgliedern mit SCZ sowie zwei gesunden Familienmitgliedern nachgewiesen (Abbildung 14). Die experimentelle Verifikation bestätigte eine Deletion, die mindestens 5,8 kb (chr14:74.034.455-74.040.304; hg19) umfasste und das Gen *ACOT2* betraf. Da die Familie nicht rekontaktiert werden konnte, konnte nicht überprüft werden, ob bei den beiden gesunden Familienmitgliedern später eine psychiatrische Erkrankung aufgetreten war. Das dritte Familienmitglied mit SCZ konnte aufgrund nicht ausreichender DNA nicht auf CNVs untersucht werden.

Zusätzlich zu den vier Familienmitgliedern konnte in derselben Region eine 5,8 kb große Deletion (chr14:74.035.975-74.041.755; hg19) in einem der Patienten mit SCZ aus dem hausinternen Datensatz detektiert werden. Diese konnte jedoch aufgrund nicht ausreichender

DNA nicht experimentell verifiziert werden. Im ExAC Datensatz wurde eine Deletion im Gen *ACOT2* beschrieben (<http://exac.broadinstitute.org>; Lek et al., 2016).

Deletionen auf Chromosom 16q23.2

In der chromosomalen Region 16q23.2 wurde in einer Familie eine Deletion bei zwei von drei Nachkommen mit SCZ und einem Elternteil mit schizotypischer Persönlichkeitsstörung identifiziert (Abbildung 15). Die Verifikation mittels genomweiter Genotypisierungsdaten zeigte, dass die Deletion 143,1 kb (chr16:81.134.821-81.277.960; hg19) umfasste und die beiden Gene *PKD1L2* und *BCMO1* betraf. Von dem dritten Familienmitglied mit SCZ war keine DNA verfügbar, sodass keine CNV-Analyse möglich war. Es konnte keine Deletion in dem hausinternen Datensatz in den Genen *PKD1L2* und *BCMO1* detektiert werden. In den ExAC Daten wurden insgesamt 22 Deletionen in diesen beiden Genen identifiziert (<http://exac.broadinstitute.org>; Lek et al., 2016).

Duplikationen auf Chromosom 6p12.1

In der chromosomalen Region 6p12.1 wurden Duplikationen bei einem Elternteil mit SCZ und einem Nachkommen mit SCZ nachgewiesen. In derselben Familie wurde auch die Deletion in der chromosomalen Region 14q11.2 nachgewiesen. Die von allen drei CNV-Algorithmen detektierte Region umfasste 39,5 kb (chr6:56.915.525-56.955.019; hg19); und betraf die Gene *KIAA1586* und *ZNF451* (Abbildung 16). In dem untersuchten gesunden Familienmitglied konnte die Duplikation nicht identifiziert werden. Von einem dritten Familienmitglied mit SCZ war keine DNA vorhanden, sodass keine CNV-Analyse durchgeführt werden konnte. Es konnte keine Duplikation in dem hausinternen Datensatz in den Genen *KIAA1586* und *ZNF451* detektiert werden. In den ExAC Daten wurden insgesamt 28 Duplikationen in diesen beiden Genen identifiziert (<http://exac.broadinstitute.org>; Lek et al., 2016).

Duplikationen auf Chromosom 9q33.2

In der chromosomalen Region 9q33.2 wurden Duplikationen in zwei Individuen einer Familie identifiziert. Die Duplikation wurde von einem Elternteil mit einer psychotischen Störung an einen Nachkommen mit SCZ vererbt (Abbildung 17). Die Verifikation mittels qPCR zeigte, dass die Duplikation mindestens das Gen *GSN* betraf. Einer der CNV-Algorithmen detektierte die Duplikation auch in dem Gen *RAB14*. Dies konnte jedoch nicht durch die experimentelle Verifikation bestätigt werden. Die CNV konnte weder in dem gesunden Familienmitglied, noch in zwei Familienmitgliedern mit einer affektiven Störung (Suizidversuch und Depression mit Wahnvorstellungen) nachgewiesen werden. Weder in den ExAC Daten, noch in dem hausinternen Datensatz wurde eine Duplikation in dem Gen *GSN* identifiziert.

Duplikationen auf Chromosom 12q24.31

In der chromosomalen Region 12q24.31 wurden Duplikationen in zwei Geschwistern mit SCZ identifiziert. Die Duplikation konnte weder in dem untersuchten gesunden Elternteil, noch in dem Geschwister mit Bulimie nachgewiesen werden (Abbildung 18). Der andere Elternteil hatte eine positive Familienanamnese für SCZ, jedoch war von diesem keine DNA für eine CNV-Analyse verfügbar. Die Verifikation mittels qPCR zeigte, dass die Duplikation mindestens 5,1kb (chr12:122.984.725-122.990.137; hg19), umfasste und zum Teil die Gene *ZCCHC8* und *RSRC2* betraf. Im hausinternen Datensatz konnte keine Duplikation in dieser Region detektiert werden. Im ExAC Datensatz wurden eine Duplikation im Gen *ZCCHC8* und zwei Duplikationen, welche *ZCCHC8* und *RSRC2* betrafen, identifiziert.

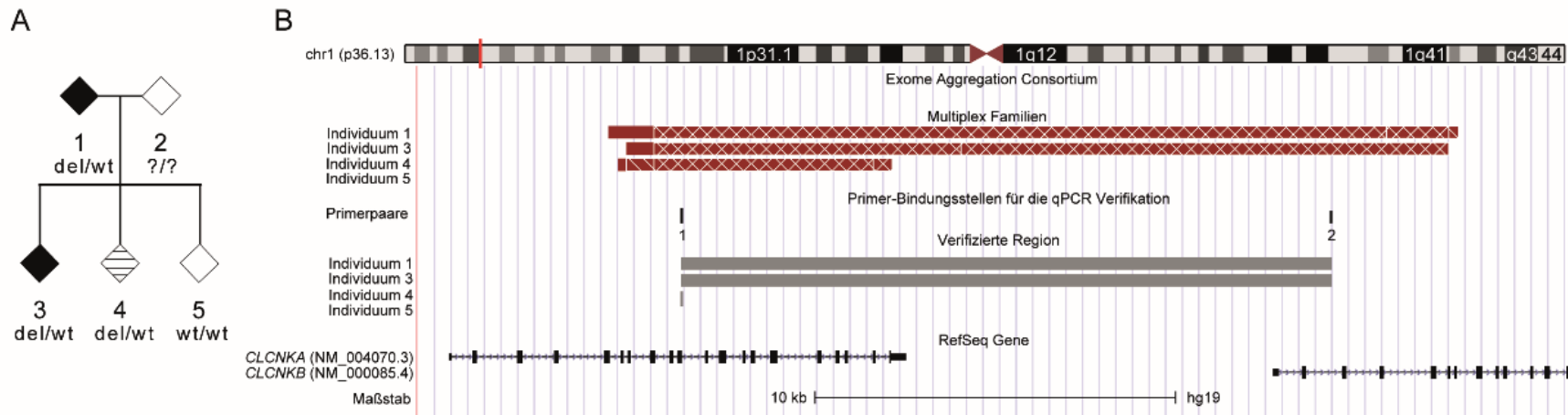


Abbildung 11 Deletionen auf Chromosom 1p36.13

A) Stammbaum der Familie, in der Deletionen in der chromosomalen Region 1p36.13 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; gestreifte Symbole: Individuen mit schizoaffectiver Störung; del: Deletion; wt: Wildtyp; ?/? nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012),XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in zwei Familienmitgliedern mit Schizophrenie (Individuen 1 und 3) und einem Familienmitglied mit schizoaffectiver Störung (Individuum 4) detektiert wurden. ■ = CoNIFER; ▨ = XHMM; ▩ = ExomeDepth. Zwei Primerpaare (1, 2) wurden für die Verifikation mittels quantitativer PCR (qPCR) verwendet. Die durch qPCR verifizierten Regionen sind in grau abgebildet. Die Deletion umfasste die Gene *CLCNKA* und *CLCNKB* in den Individuen 1 und 3, in Individuum 4 nur das Gen *CLCNKA*. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. Die Abbildung wurde mit dem UCSC genome browser (<http://genome.ucsc.edu/index.html>) erstellt.

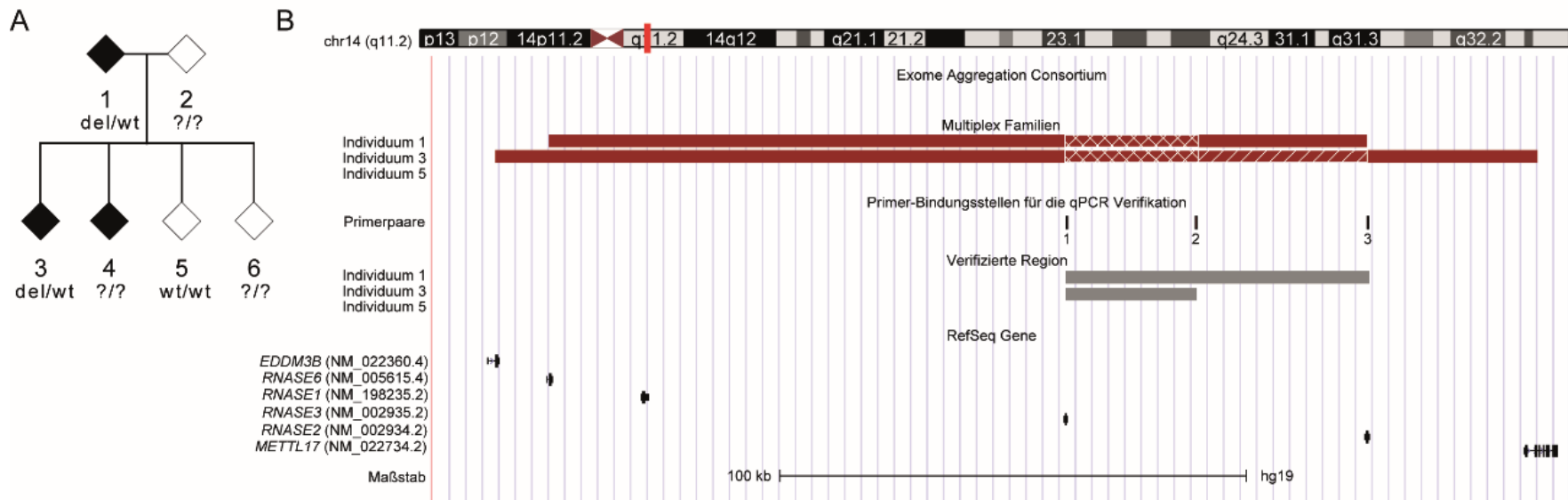


Abbildung 12 Deletionen auf Chromosom 14q11.2

A) Stammbaum der Familie, in der Deletionen in der chromosomalen Region 14q11.2 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; del: Deletion; wt: Wildtyp; ?/? nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012),XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in zwei Familienmitgliedern mit Schizophrenie (Individuen 1 und 3) detektiert wurden. ■ = CoNIFER; ▨ = XHMM; ▩ = ExomeDepth. Drei Primerpaare (1-3) wurden für die Verifikation mittels quantitativer PCR (qPCR) verwendet. Die durch qPCR verifizierte Regionen sind in grau abgebildet. Die Deletion umfasste die Gene *RNASE3* und *RNASE2* in Individuum 1, in Individuum 3 nur das Gen *RNASE3*. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. Die Abbildung wurde mit dem UCSC genome browser (<http://genome.ucsc.edu/index.html>) erstellt.

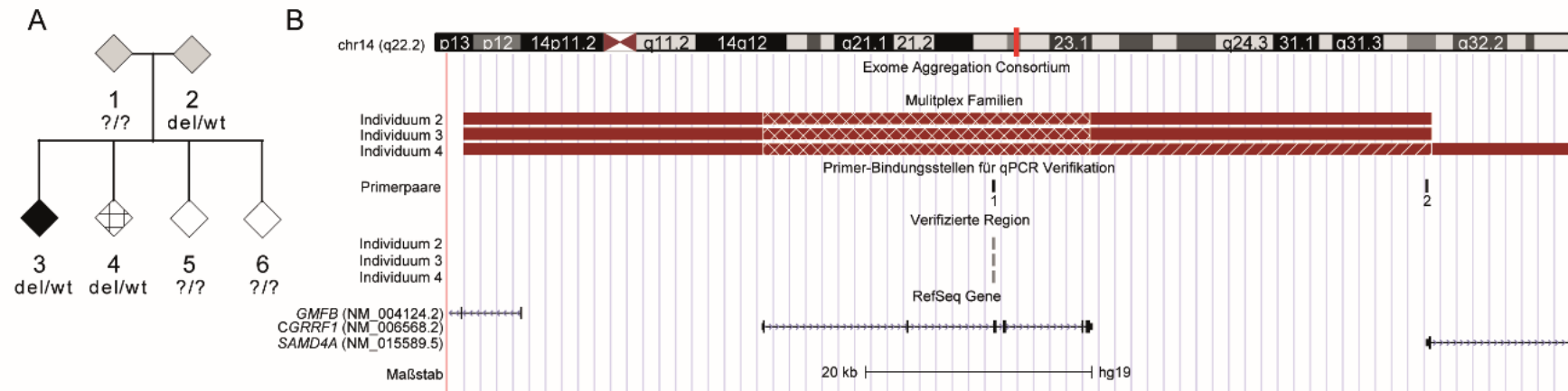


Abbildung 13 Deletionen auf Chromosom 14q22.2

A) Stammbaum der Familie, in der Deletionen in der chromosomalen Region 14q22.2 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; karierte Symbole: Individuen mit bipolarer Störung; graue Symbole: Individuen mit anderer psychiatrischer Erkrankung; del: Deletion; wt: Wildtyp; $?/?$ nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012),XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in jeweils einem Familienmitglied mit Schizophrenie (Individuum 3), bipolarer Störung (Individuum 4) und Anpassungsstörung (Individuum 2) detektiert wurden. ■ = CoNIFER; ▨ = XHMM; ▩ = ExomeDepth. Zwei Primerpaare (1, 2) wurden für die Verifikation mittels quantitativer PCR (qPCR) verwendet. Die durch qPCR verifizierten Regionen sind in grau abgebildet. Die Deletion umfasste in allen drei Familienmitgliedern das Gen *CGRRF1*, jedoch nicht das Gen *SAMD4A*. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. Die Abbildung wurde mit dem UCSC genome browser (<http://genome.ucsc.edu/index.html>) erstellt.



A) Stammbaum der Familie, in der Deletionen in der chromosomalen Region 14q24.3 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; del: Deletion; wt: Wildtyp; ?? nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012), XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in zwei Familienmitgliedern mit Schizophrenie (Individuen 3 und 4), zwei gesunden Familienmitgliedern (Individuen 6 und 7) und einem der 141 Patienten mit Schizophrenie (hausinterner Datensatz) detektiert wurden. ■ = CoNIFER; ▨ = XHMM; ▩ = ExomeDepth. Zwei Primerpaare (1, 2) wurden für die Verifikation mittels quantitativer PCR (qPCR) verwendet. Die durch qPCR verifizierten Regionen sind in grau abgebildet. Die Deletion umfasste in allen vier Familienmitgliedern mindestens das Gen *ACOT2*. Im Datensatz des Exome Aggregation Consortium (ExAC) wurde eine Deletion im Gen *ACOT2* beschrieben. Zwei weitere Deletionen betrafen die angrenzenden Gene *HEATR4* und *ACOT1* (<http://exac.broadinstitute.org>; Lek et al., 2016). NFE: nicht-finnische Europäer. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. Die Abbildung wurde mit dem UCSC genome browser (<http://genome.ucsc.edu/index.html>) erstellt.

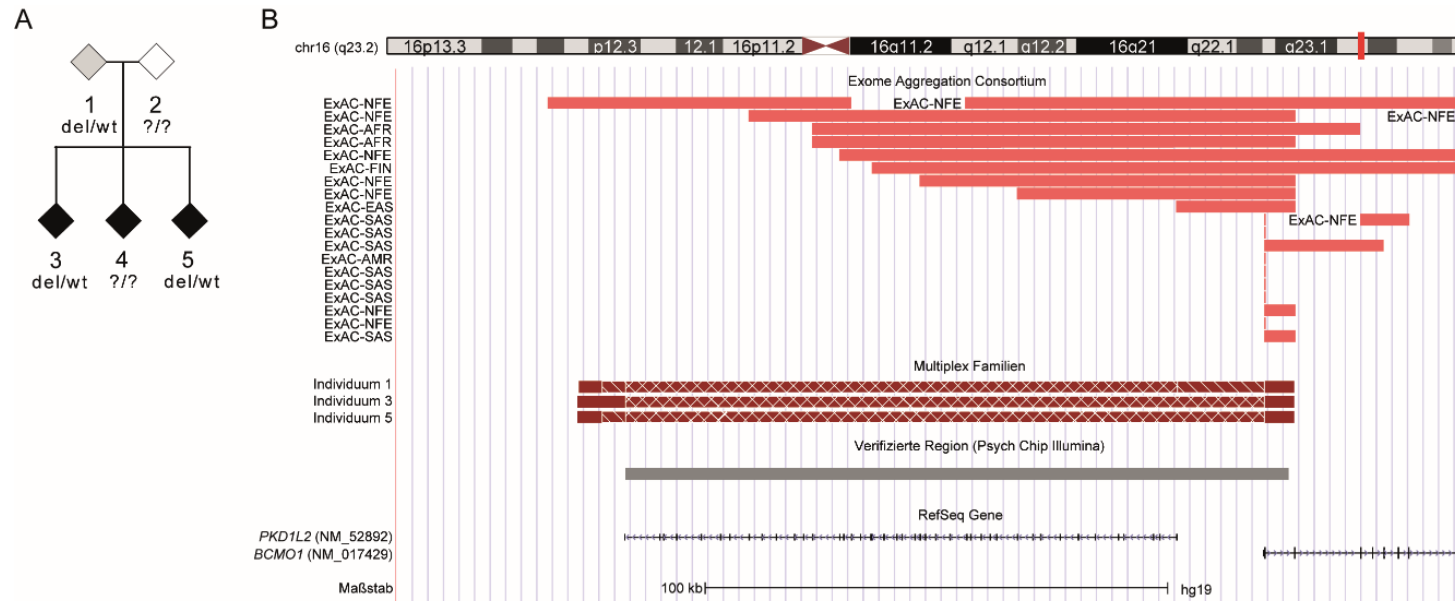


Abbildung 15 Deletionen auf Chromosom 16q23.2

A) Stammbaum der Familie, in der Deletionen in der chromosomalen Region 16q23.2 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; graue Symbole: Individuen mit anderer psychiatrischer Erkrankung; del: Deletion; wt: Wildtyp; ?? nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012), XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in zwei Familienmitgliedern mit Schizophrenie (Individuen 3 und 5), und einem Familienmitglied mit schizotypischer Persönlichkeitsstörung (Individuum 1) detektiert wurden. ■ = CoNIFER; ▨ = XHMM; ▩ = ExomeDepth. Die Deletion betraf bei allen drei Familienmitgliedern die Gene *PKDL2* und *BCMO1* wie mittels Illumina GenomeStudio visuell bestätigt wurde (grauer Bereich). Im Datensatz des Exome Aggregation Consortium (ExAC) wurden insgesamt 22 Deletion in *PKDL2* und/oder *BCMO1* beschrieben (<http://exac.broadinstitute.org>; Lek et al., 2016). NFE: nicht-finnische Europäer; AFR: Afrikaner; FIN: Finnen; EAS: Ostasiaten; SAS: Südasiaten; AMR: Latino. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. Die Abbildung wurde mit dem UCSC genome browser (<http://genome.ucsc.edu/index.html>) erstellt.

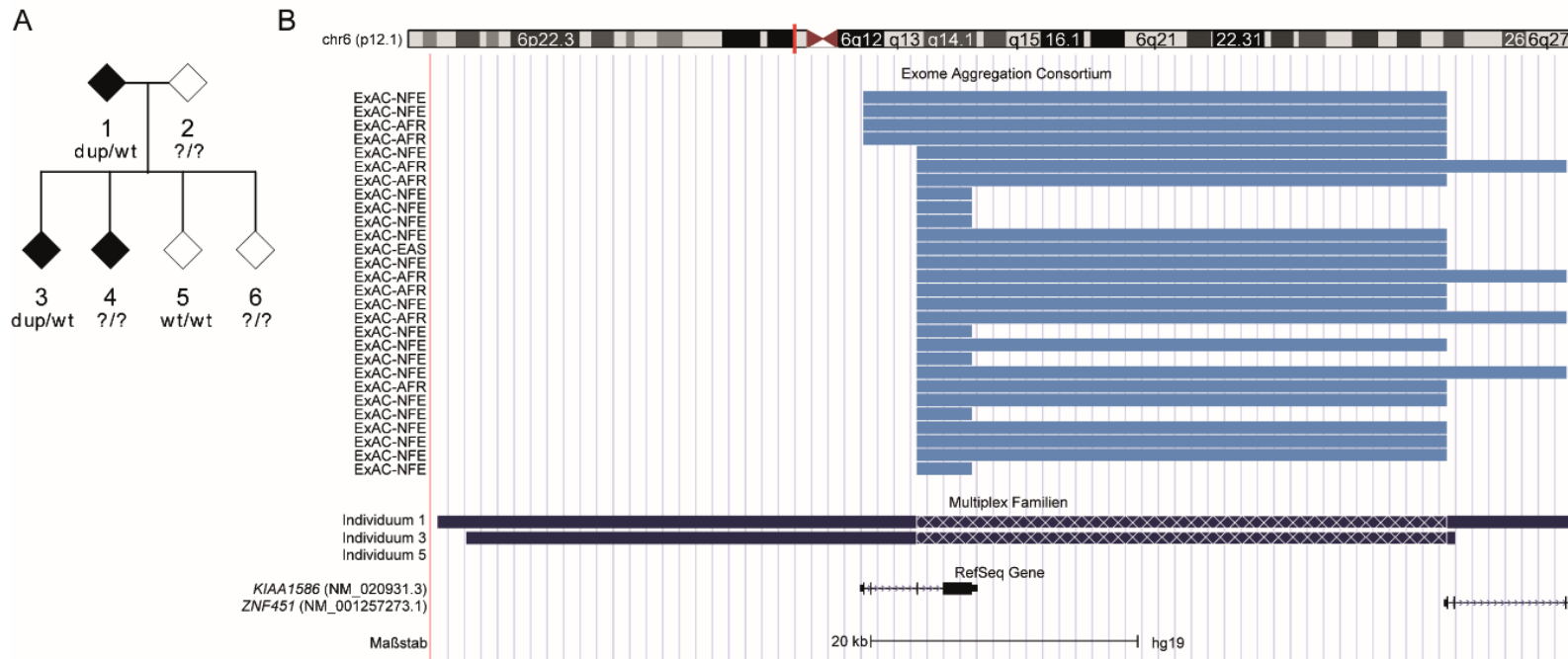


Abbildung 16 Duplikationen auf Chromosom 6p12.1

A) Stammbaum der Familie, in der Duplikationen in der chromosomalen Region 6p12.1 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; dup: Duplikation; wt: Wildtyp; ?/? nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012),XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in zwei Familienmitgliedern mit Schizophrenie (Individuen 1 und 3), detektiert wurden. ■ = CoNIFER; ▨ = XHMM; ▩ = ExomeDepth. Die Duplikation betraf bei beiden Familienmitgliedern die Gene *KIAA1586* und *ZNF451*.

Im Datensatz des Exome Aggregation Consortium (ExAC) wurden insgesamt 28 Duplikationen in *KIAA1586* und/oder *ZNF451* beschrieben (<http://exac.broadinstitute.org>; Lek et al., 2016). NFE: nicht-finnische Europäer; AFR: Afrikaner; EAS: Ostasiaten. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. Die Abbildung wurde mit dem UCSC genome browser (<http://genome.ucsc.edu/index.html>) erstellt.

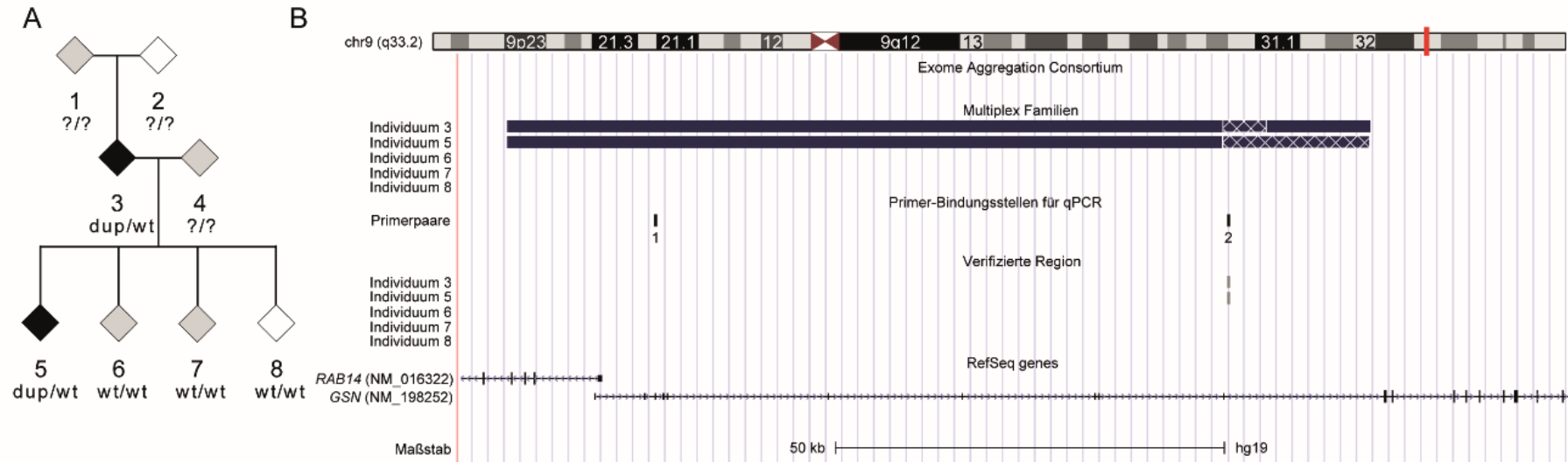


Abbildung 17 Duplikationen auf Chromosom 9q33.2

A) Stammbaum der Familie, in der Duplikationen in der chromosomalen Region 9q33.2 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; graue Symbole: Individuen mit anderer psychiatrischer Erkrankung; dup: Duplikation; wt: Wildtyp; ?? nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012),XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in zwei betroffenen (psychotische Störung und Schizophrenie) Familienmitgliedern (Individuen 3 und 5), detektiert wurden. ■ = CoNIFER; ▨ = XHMM; ▩ = ExomeDepth. Zwei Primerpaare (1, 2) wurden für die Verifikation mittels quantitativer PCR (qPCR) verwendet. Die durch qPCR verifizierte Regionen sind in grau abgebildet. Die Experimentelle Verifikation zeigte, dass die Duplikation bei beiden Individuen nur Primerpaar 2 im Gen GSN umfasste. Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. Die Abbildung wurde mit dem UCSC genome browser (<http://genome.ucsc.edu/index.html>) erstellt.

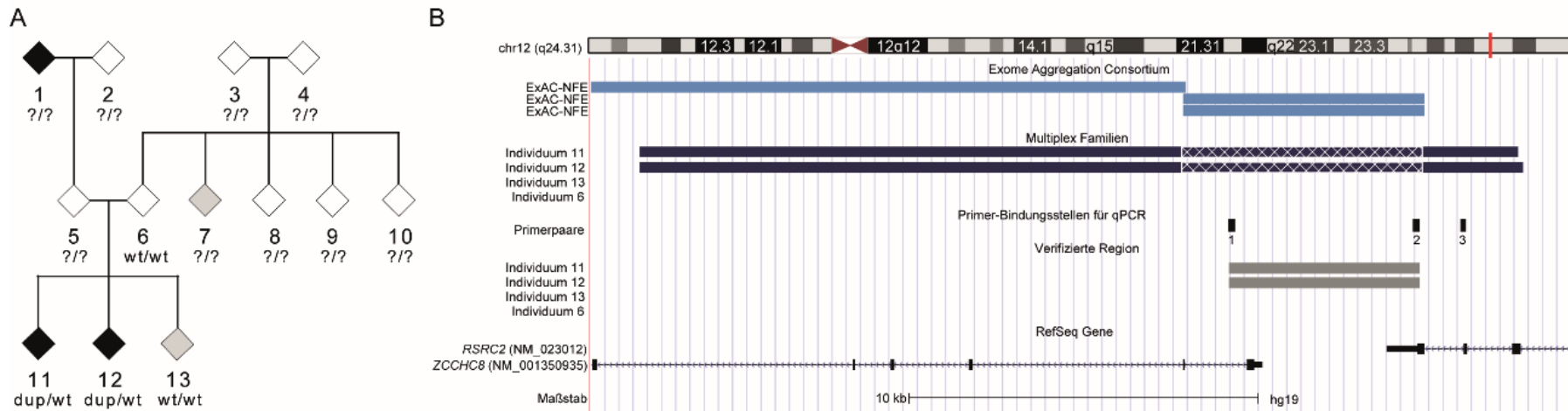


Abbildung 18 Duplikationen auf Chromosom 12q24.31

A) Stammbaum der Familie, in der Duplikationen in der chromosomalen Region 12q24.31 identifiziert wurden. Schwarze Symbole: Individuen mit Schizophrenie; graue Symbole: Individuen mit anderer psychiatrischer Erkrankung; dup: Duplikation; wt: Wildtyp; $?/?$ nicht untersucht. Um die Identifizierung der Familienmitglieder zu erschweren, wurde nicht zwischen männlichen und weiblichen Individuen unterschieden und die Reihenfolge der Geburt verändert.

B) Dargestellt sind die Positionen der Kopienzahlvarianten, welche von den drei Algorithmen CoNIFER (Krumm et al., 2012), XHMM (Fromer et al., 2012) und ExomeDepth (Plagnol et al., 2012) in zwei Familienmitgliedern mit Schizophrenie (Individuen 11 und 12), detektiert wurden. ■ = CoNIFER; ■ = XHMM; ■ = ExomeDepth. Drei Primerpaare (1-3) wurden für die Verifikation mittels quantitativer PCR (qPCR) verwendet. Die durch qPCR verifizierte Regionen sind in grau abgebildet. Die Experimentelle Verifikation zeigte, dass die Duplikation bei beiden Individuen nur die Primerpaare 1 und 2 umfasste und die Gene *RSRC2* und *ZCCHC8* betraf. Im Datensatz des Exome Aggregation Consortium (ExAC) wurden insgesamt 3 Duplikationen in *RSRC2* und/oder *ZCCHC8* beschrieben (<http://exac.broadinstitute.org>; Lek et al., 2016). NFE: nicht-finnische Europäer; Die chromosomalen Positionen basieren auf dem Referenzgenom GRCh37/hg19. Die Abbildung wurde mit dem UCSC genome browser (<http://genome.ucsc.edu/index.html>) erstellt.

Genetische Evidenz

Die 13 in dieser Studie identifizierten Kandidatengene (Tabelle 7, fett gedruckte Gene) wurden in verschiedenen öffentlich zugänglichen Datensätzen sowie einem unabhängigen hausinternen Datensatz überprüft. Zusätzlich wurden Gen-basierte Tests durchgeführt (Tabelle 8). In sechs Genen (*ACOT2*, *CLCNKA*, *CLCNKB*, *GSN*, *PKD1L2*, *RSRC2*) wurden seltene SNVs im hausinternen Datensatz detektiert und in zwei Genen (*PKD1L2* und *ZNF451*) wurden SNVs in einer SCZ *de novo* Studie beschrieben. Für vier der 12 untersuchten Kandidatengene (das Gen *RNASE2* war nicht Teil der VEGAS2 Datenbank) wurde bei den Gen-basierten Tests eine Assoziation ($p < 0,05$) zwischen häufigen Varianten und einem Risiko für SCZ identifiziert. Jedoch blieb die Assoziation nur für die beiden Gene *RSRC2* und *ZCCHC8* auch nach der Korrektur für multiples Testen mittels Bonferroni Methode signifikant (p korrigiert = 0,024 bzw. 0,036).

Tabelle 8 Zusätzliche genetische Evidenz für die 13 Kandidatengene der CNV-Studie

Gen	Hausinterner Datensatz	CNV-Studie	GWAS	Publizierte SCZ NGS Studien	Genebook NS strict P	Genebook NS broad P	Gene-based Test p-Wert	Gene-based test p-Wert korrigiert	ExAC CNV z Score	pTriplo / pHaplo
<i>ACOT2</i>	E121*	-	-	-	-	0.02	0.130		-	0,21
<i>BCMO1</i>	-	-	-	-	1	1.00	0.243		-2.42	-
<i>CGRRF1</i>	-	-	-	-	0.79	0.50	0.784		0.41	0,21
<i>CLCNKA</i>	D383Y, A485V	-	-	-	1.00	1.00	0.031		-	0,74
<i>CLCNKB</i>	A488D	-	-	-	1.00	0.29	0.719		-	0,70
<i>GSN</i>	Q443K, T644M, D645G, R689Q	-	-	-	1.00	0.33	0.369		0.55	0,63
<i>KIAA1586</i>	-	-	-	-	1.00	1.00	0.657		-2.47	0,14
<i>PKD1L2</i>	I87T, L173*	-	-	Howrigan et al. 2020; c.1566G>A	1.00	0.42	0.276		-	-
<i>RNASE2</i>	-	-	-	-	0.59	1.00	NA		-	0,07
<i>RNASE3</i>	-	-	-	-	0.48	0.30	0.018		-	0,12
<i>RSRC2</i>	S384A	-	-	-	0.78	1.00	0.002	0.024	0.93	0,42
<i>ZCCHC8</i>	-	-	-	-	1	0.33	0.003	0.036	-0.64	0,49
<i>ZNF451</i>	-	-	-	Howrigan et al. 2020; c.2362G>A	1	0.32	0.173		-2.47	0,40

Um zusätzliche genetische Evidenz für die Kandidatengene zu erhalten, wurde überprüft, ob diese auch in anderen Datensätzen identifiziert wurden. Zu diesen zählten der hausinterne Datensatz mit 141 Patienten mit Schizophrenie (SCZ) und verschiedene öffentlich zugängliche Datensätze wie (i) bekannte SCZ-assoziierte Kopienzahlvarianten (CNVs; Marshall et al. 2017), (ii) Gene, in denen in den größten bisher publizierten genomweiten Assoziationsstudien (GWAS) ((Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Pardiñas et al., 2018; Trubetsky et al., 2022) ein genomweit signifikanter mit SCZ assoziierter Einzelnukleotid-Polymorphismus beschrieben wurde und (iii) Gene, die in anderen SCZ Exom- und Genom-Sequenzierungsstudien identifiziert wurden (Girard et al., 2011; Xu et al., 2011, 2012; Timms et al., 2013; Gulsuner et al., 2013; Guipponi et al., 2014; Fromer et al., 2014; Takata et al., 2014; McCarthy et al., 2014; Kranz et al., 2015; Wang et al., 2015; Ambalavanan et al., 2016; Watanabe et al., 2016; Egawa et al., 2016; Homann et al., 2016; Zhou et al., 2016; Smedemark-Margulies et al., 2016; Hornig et al., 2017; John et al., 2017; Tang et al., 2017; Kos et al., 2017; O'Brien et al., 2018; Senormanci et al., 2018; Xue et al., 2018; John et al., 2018, 2019; Suddaby et al., 2019; Li et al., 2021; Balakrishna & Curtis, 2020; Rees et al., 2020; Howrigan et al., 2020; Chen et al., 2021; Toyama et al., 2022; Shang et al., 2024). NGS: *next generation sequencing*; NS (*nonsynonymous*) strict: trunkierende und *missense* Varianten, die im Genebook (Purcell et al. 2014) von fünf Algorithmen als pathogen klassifiziert wurden (<http://atgu.mgh.harvard.edu/~spurcell/genebook/genebook.cgi?user=guest&cmd=overview#>); NS broad: trunkierende und *missense* Varianten, die im Genebook von mindestens einem Algorithmus als pathogen klassifiziert wurden. Außerdem wurden Gen-basierte Tests basierend auf den Daten der PGC SCZ GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) durchgeführt, NA nicht vorhanden in VEGAS (Mishra & Macgregor, 2015); - nicht beschrieben/keine Daten vorhanden. Gene-based test p-Wert: alle p-Werte vor Korrektur für multiples Testen; Gene-based test p-Wert korrigiert: signifikante p-Werte nach Korrektur für multiples Testen. ExAC CNV z Score: Größere z Scores deuten auf eine größere Intoleranz gegenüber CNVs hin. pTriplo / pHaplo: Größere Werte deuten auf eine größere Intoleranz gegenüber Duplikationen / Deletionen hin (Collins et al., 2022).

Genexpression

Basierend auf dem GTEx Datensatz waren sieben der 13 Kandidatengene in allen 13 verfügbaren Gehirngeweben exprimiert (Transkripte pro 1 Million (Median) ≥ 1): *GSN*, *RSRC2*, *ACOT2*, *CGRRF1*, *KIAA1586*, *ZNF451* und *ZCCHC8*. Drei Gene (*RNASE3*, *PKD1L2* und *BCMO1*) waren in keinem der 13 Gehirngewebe exprimiert, ein Gen (*RNASE2*) war in fünf der 13 Gehirngewebe exprimiert (Amygdala, Hippocampus, Hypothalamus, Spinalnerv C1 und Substantia nigra). Zwei Gene (*CLCNKA*, *CLCNKB*) waren in zwei Gehirngeweben (Kleinhirnhemisphäre, Cerebellum) exprimiert (Abbildung 19, Tabelle 9; GTEx Consortium, 2017; gtexportal.org/home/multiGeneQueryPage).

Weiterhin waren 11 Kandidatengene (Tabelle 10) in mindestens einer von 15 publizierten Studien (Tabelle A7; Catts et al., 2014; Chen et al., 2013; de Baumont et al., 2015; Fillman et al., 2013; Hwang et al., 2013; Iwamoto et al., 2004, 2005; Kohen et al., 2014; Lanz et al., 2015; Maycox et al., 2009; Narayan et al., 2008; Pietersen, Mauney, Kim, Lim, et al., 2014; Pietersen, Mauney, Kim, Passeri, et al., 2014; Wu et al., 2012; Zhao et al., 2015) zwischen Patienten mit SCZ und Kontrollen differenziell exprimiert.

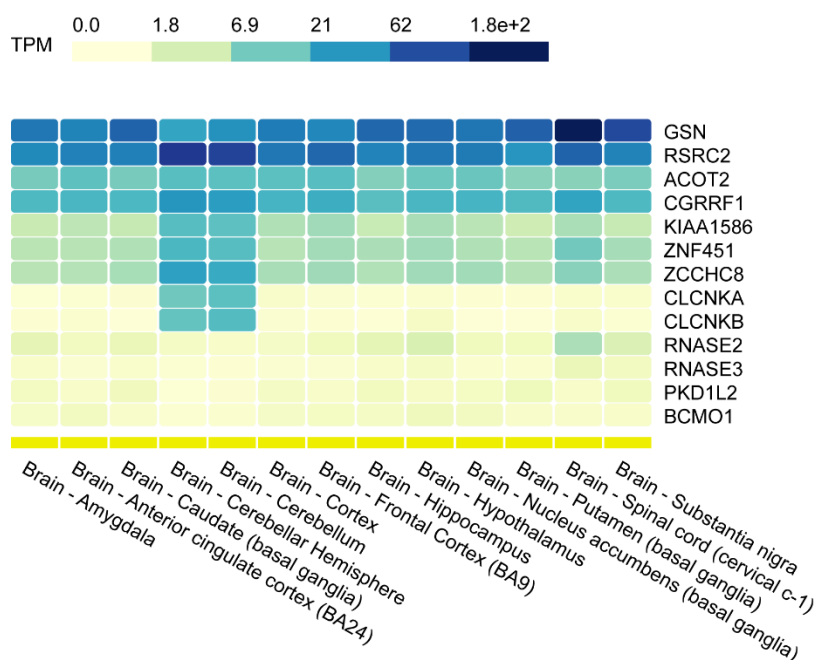


Abbildung 19 Genexpression der 13 Kandidatengene in verschiedenen Gehirngeweben

Die Intensität der Blautöne gibt die Höhe der Expression im jeweiligen Gewebe an, dunklere Farbtöne stellen eine höhere Expression dar. TPM: Median Transkripte pro eine Million. BA: Brodmann Areal. <https://gtexportal.org/home/multiGeneQueryPage>, GTEx Consortium, 2017).

Tabelle 9 Genexpression der 13 Kandidatengene in verschiedenen Gehirngeweben

Gen	Amygdala	Anterior cingulate cortex (BA24)	Caudate (basal ganglia)	Cerebellar Hemisphere	Cerebellum	Cortex	Frontal Cortex (BA9)	Hippocampus	Hypothalamus	Nucleus accumbens (basal ganglia)	Putamen (basal ganglia)	Spinal cord (cervical c-1)	Substantia nigra
<i>GSN</i>	34.59	28.6	45.28	16.69	23.25	32.17	27.52	42.93	41.4	53.41	47.56	177	66.52
<i>RSRC2</i>	26.53	30.02	30.27	90.18	74.58	34.22	43.08	28.9	34.53	32.62	22.19	45.3	29.14
<i>ACOT2</i>	6.641	8.817	6.655	9.614	9.151	9.097	9.653	5.789	7.433	7.872	5.433	5.394	6.386
<i>CGRRF1</i>	10.81	11.79	11.22	21.82	19.56	12.28	14.02	9.408	11.72	12.29	10.57	17.08	10.84
<i>KIAA1586</i>	2.289	2.856	2.455	9.627	8.963	3.425	4.102	2.35	3.779	2.973	2.083	3.685	2.389
<i>ZNF451</i>	2.922	3.037	3.47	11.4	9.782	3.111	4.008	3.624	4.166	3.398	3.014	7.03	3.897
<i>ZCCHC8</i>	3.03	3.194	3.82	18.32	15.18	3.722	4.223	3.362	4.163	4.061	3.248	5.456	3.58
<i>CLCNKA</i>	0.1031	0.1397	0.1626	7.351	9.013	0.2597	0.174	0.1352	0.1975	0.13	0.1236	0.2296	0.2457
<i>CLCNKB</i>	0.1266	0.211	0.05455	8.353	10.14	0.1749	0.1234	0.08243	0.3761	0.06713	0.05655	0.3106	0.1404
<i>RNASE2</i>	1.063	0.5326	0.9062	0.4291	0.3235	0.3912	0.612	1.103	1.647	0.6445	0.5829	3.575	1.605
<i>RNASE3</i>	0.3104	0.1843	0.248	0.1373	0.1184	0.1791	0.1841	0.3122	0.3964	0.2115	0.2082	0.8571	0.5269
<i>PKD1L2</i>	0.4935	0.3166	0.5985	0.09828	0.167	0.4113	0.3399	0.546	0.3809	0.4705	0.6999	0.3098	0.6381
<i>BCMO1</i>	0.3434	0.5341	0.2706	0.1472	0.1929	0.3893	0.4585	0.4381	0.6602	0.5244	0.2704	0.288	0.3279

Gelistet sind die TPM Werte der Genotype-Tissue Expression (GTEx) Datenbank für 13 verschiedene Gehirngewebe. TPM: Transkripte pro 1 Million (Median). BA: Brodmann Areal.

Tabelle 10 Differenzielle Genexpression bei Patienten mit SCZ und Kontrollen

Gen	GSE 12654	GSE 17612	GSE 12649	GSE 53987 HIPPO	GSE 53987 PFC	GSE 62191	PMID_ 22869038	GSE 35978	GSE 46509	GSE 21138	GSE 37981	PMID_ 25113377	PMID_ 24744400	PMID_ 24594777	PMID_ 24169640	PMID_ 22558445
ACOT2	-	-	1.17	-	-	-	-	-	-	-	-	-	-	-	-	-
BCMO1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CGRRF1	-	-	-	0.85	-	-	-	-	-	-	-	-	-	-	-	-
CLCNKA	-	-	-	1.1	-	-	-	-	-	-	-	-	-	-	-	-
CLCNKB	-	-	-	1.14	-	-	-	-	-	-	-	-	-	-	-	-
GSN	-	-	-	-	-	-	-	-	-	-	0.33	-	-	-	-	-
KIAA1586	-	-	-	0.84	-	-	-	-	-	-	-	-	-	-	-	-
PKD1L2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RNASE2	-	-	-	-	-	-	-	-	-	-	0.26	-	-	-	-	-
RNASE3	-	-	-	1.17	-	-	-	-	-	-	-	-	-	-	-	-
RSRC2	-	-	-	0.85	-	-	-	-	-	-	-	-	-	-	-	-
ZCCHC8	-	-	1.22	1.22	-	-	-	-	-	-	-	-	-	-	-	-
ZNF451	-	1.42	-	1.2	0.9	-	-	-	-	-	-	-	-	-	-	-

Die Tabelle zeigt die differenzielle Genexpression (*fold change*) bei Patienten mit Schizophrenie (SCZ) und gesunden Kontrollen in 15 verschiedenen Studien mit insgesamt 746 Probanden (Catts et al., 2014; Chen et al., 2013; de Baumont et al., 2015; Fillman et al., 2013; Hwang et al., 2013; Iwamoto et al., 2004, 2005; Kohen et al., 2014; Lanz et al., 2015; Maycox et al., 2009; Narayan et al., 2008; Pietersen, Mauney, Kim, Lim, et al., 2014; Pietersen, Mauney, Kim, Passeri, et al., 2014; Wu et al., 2012; Zhao et al., 2015). In einer Studie (GSE53987; Lanz et al., 2015) wurden zwei verschiedene Hirnregionen (HIPPO; Hippocampus und frontaler Cortex; PFC) analysiert. *Fold change* > 1: Die Expression war bei den Patienten signifikant höher als bei den Kontrollprobanden. *Fold change* < 1: Die Expression war bei den Patienten signifikant niedriger als bei den Kontrollprobanden. – keine signifikant differenzielle Genexpression. Weitere Informationen zu den 15 Studien sind in Tabelle A7 aufgeführt.

Protein-Protein Interaktionsanalysen

Basierend auf den 13 neu identifizierten Kandidatengenomen wurde zusätzlich zu den Netzwerken Network_Literature und Network_HIPPIE_minus_Literature noch ein drittes Netzwerk erstellt. Dieses Netzwerk (Network_Study) enthielt alle *high confidence* PPIs der Kandidatenproteine. Insgesamt wurden sieben der Kandidatenproteine mit 31 Partnerproteinen in 32 PPIs vernetzt. Drei der 32 PPIs vernetzten eines der Kandidatenproteine mit einem Protein aus der Literatur (Tabelle 11).

Anschließend wurde untersucht, wie viele der PPIs in Network_Study ebenfalls in den Netzwerken Network_Literature und Network_HIPPIE_minus_Literature vorkamen (Tabelle 12). Weder der Chi-Quadrat-Test noch der Exakte Fisher-Test ($p_{\text{Chi}}=0,57$, $p_{\text{Fisher}}=0,61$) zeigten eine signifikante Anreicherung von PPIs der neu identifizierten Kandidatenproteine im Netzwerk Network_Literature. Bei der Netzwerk-basierten Priorisierung der Kandidatenproteine mittels ToppNet (Chen et al., 2007, 2009) erhielten die Proteine GSN, ZCCHC8, und CGRRF1 die höchsten Scores (Tabelle 13).

Tabelle 11 Protein-Protein Interaktionen in Network_Study

Gen Name 1	Gen Name 2
GSN	CASP4
GSN	PTK2B
ZCCHC8	NCBP2

Gelistet sind die drei Protein-Protein Interaktionen der neu identifizierten Schizophrenie Kandidatenproteine (links) mit Interaktionspartnern aus Network_Literature (rechts).

Tabelle 12 Anzahl der Protein-Protein Interaktionen von der HIPPIE Datenbank in den untersuchten Netzwerken

	Network_Literature	Network_HIPPIE_minus_Literature	
Network_Study	3	29	32
PPIs in HIPPIE ohne Network_Study	5.051	29.889	34.940
	5.054	29.918	Alle <i>high confidence</i> PPIs in HIPPIE (n = 34.972)

Die Tabelle zeigt die Anzahl der *high confidence* Protein-Protein Interaktionen (PPIs) in den untersuchten Netzwerken Network_Literature (PPIs der n = 1.271 Schizophrenie Kandidatenproteine aus der Literatur), Network_HIPPIE_minus_Literature (PPIs in HIPPIE Datenbank ohne die PPIs in Network_Literature) und Network_Study (PPIs der n = 13 neu identifizierten Kandidatenproteine) sowie aller PPIs in HIPPIE ohne die PPIs von Network_Study.

Tabelle 13 Priorisierung der Kandidatengene mittels ToppNet

Rang	Gen ID	Gen	Anzahl PPIs	Score
1	2934	GSN	105	1.35E-04
2	55596	ZCCHC8	71	6.76E-05
3	10668	CGRRF1	36	6.18E-05
4	26036	ZNF451	35	3.99E-05
5	65117	RSRC2	26	3.00E-05
6	10965	ACOT2	17	1.59E-05
7	6037	RNASE3	13	1.35E-05
8	57691	KIAA1586	4	5.31E-06
9	6036	RNASE2	4	4.00E-06
10	114780	PKD1L2	2	2.64E-06
11	1187	CLCNKA	1	1.57E-06
12	53630	BCO1	1	9.20E-07
13	1188	CLCNKB	0	

Die Tabelle zeigt die 13 neu identifizierten Kandidatengene priorisiert mittels ToppNet (Chen et al., 2007, 2009) nach ihrer Konnektivität zu bekannten Schizophrenie Kandidatengenen aus der Literatur. ID: Entrez Gene ID (<http://www.ncbi.nlm.nih.gov/gene>).

5 Diskussion

Die bisher in der Literatur beschriebenen genetischen Risikofaktoren für SCZ erklären nur etwa 30% der geschätzten Heritabilität. Folglich wurde ein großer Teil der krankheitsrelevanten Varianten noch nicht identifiziert. Da der Fokus früherer Studien vor allem auf häufigen Varianten und großen CNVs lag, ist anzunehmen, dass auch seltene, hoch penetrante Einzelnukleotid-Varianten und kleinere CNVs, die mit den bisher verwendeten Genotypisierungsmethoden nicht nachgewiesen werden konnten, einen Teil der Pathogenese der SCZ erklären könnten. Ziel dieser Arbeit war daher die Identifizierung seltener pathogener Varianten mittels neuerer *Next Generation Sequencing* Methoden. Die Arbeit gliedert sich in zwei Hauptbestandteile:

(i) Familien-Pilotstudie: In dieser Studie wurden drei multiplex Familien mit jeweils drei Familienmitgliedern mit SCZ exomweit sequenziert. Der Fokus lag sowohl auf der Identifizierung von Einzelnukleotid-Varianten und kleinen Indels (Insertionen und Deletionen) als auch auf der Identifizierung von CNVs.

(ii) CNV-Studie: In dieser Studie wurde bei 47 weiteren SCZ multiplex Familien eine CNV-Analyse basierend auf Exom-Sequenzierungsdaten durchgeführt.

Im Folgenden werden die Hauptidegebnisse der vorliegenden Arbeit im Kontext der aktuellen Literatur diskutiert.

5.1 Pilotstudie – Evidenz für die Relevanz der Chromatinmodifikation in der Pathogenese der Schizophrenie

Im Rahmen der Familien-basierten Pilotstudie konnten mehrere SCZ Kandidatengene identifiziert werden, welche anschließend mittels PPI Analysen hinsichtlich ihrer möglichen Krankheitsrelevanz priorisiert wurden. Dies basierte auf der Annahme, dass Proteine, die relevant für dieselbe Erkrankung sind, auch häufig miteinander interagieren, in den gleichen biologischen Stoffwechselwegen eine Rolle spielen und dass pathogene Veränderungen in interagierenden Proteinen häufig zu ähnlichen Phänotypen führen (Barabási et al., 2011). In der Pilotstudie wurden insgesamt 39 Gene mit co-segregierenden seltenen Varianten identifiziert. Folgende Gene scheinen jedoch am ehesten phänotyprelevant zu sein: *MECP2* (*methyl CpG binding protein 2*), *HDAC5* (*histone deacetylase 5*), *SRRM2* (*serine/arginine repetitive matrix 2*), *SMARCC1* (*SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 1*) und *FOXA1* (*forkhead box A1*). Mit Ausnahme von *FOXA1* sind alle diese Gene im Gehirn exprimiert (basierend auf der GTEx Datenbank). Im hausinternen Datensatz sowie in einer publizierten Studie war *FOXA1* jedoch zwischen Patienten mit SCZ und Kontrollen differenziell exprimiert. Vier der Kandidatengene (*MECP2*,

HDAC5, *SMARCC1* und *FOXA1*) scheinen eine Rolle bei der Chromatinmodifikation zu spielen. Als Chromatin bezeichnet man die Kombination aus DNA und Proteinen (u.a. Histonen) aus denen sich ein Chromosom zusammensetzt. Durch die Modifikation des Chromatins kann die Expression von Genen beeinflusst werden, ohne dass die Nukleotidsequenz verändert wird. Zu den wichtigsten bekannten Mechanismen der Chromatinmodifikation zählen die DNA-Methylierung und die Histonmodifikation (Schaaf & Zschocke, 2018).

Bereits in früheren Exom-Sequenzierungsstudien wurde ein Zusammenhang der Chromatinmodifikation mit SCZ beschrieben. In einer Studie, in der 42 Patienten mit SCZ exomweit auf *de novo* Mutationen untersucht wurden, wurde eine Anreicherung von pathogenen Varianten in chromatinverändernden Genen nachgewiesen (McCarthy et al., 2014). In einer weiteren Studie mit 231 SCZ Trios wurden zwei *de novo loss-of-function* Varianten im Gen *SETD1A* identifiziert. Dieses Gen kodiert für eine Untereinheit der Histon-Methyltransferase (Takata et al., 2014). Zusätzliche Evidenz für das Gen *SETD1A* lieferte eine Studie von Singh et al. (2016). Durch die Analyse der Exom-Sequenzierungsdaten von 4.264 Patienten mit SCZ, 9.343 Kontrollen und 1.077 Trios wurde in dieser Studie eine genomweit signifikante Assoziation zwischen seltenen *loss-of-function* Varianten in dem Gen *SETD1A* und dem Risiko für eine Erkrankung mit SCZ identifiziert. *SETD1A* ist bisher eines von nur 10 Genen, bei dem eine statistisch eindeutige Assoziation ($p = 2 \times 10^{-6}$) von seltenen Varianten mit SCZ nachgewiesen werden konnte (Singh et al., 2022). Eine *de novo* Stopp-Mutation im Gen *CHD2* (*chromodomain helicase DNA binding protein 2*) wurde in einem Fallbericht bei einem Patienten mit frühmanifestester SCZ und Epilepsie identifiziert (Poisson et al., 2020). *CHD2* kodiert für ein Chromodomänenprotein, das an der Neurogenese, der Chromatinmodifikation und der Genexpression beteiligt ist (Nieto-Estevez & Hsieh, 2018).

Das Gen *MECP2* kodiert für ein Protein, welches an methylierte CpG-Stellen bindet und somit die Transkription beeinflussen kann (J. D. Lewis et al., 1992; Nan et al., 1997). Pathogene Varianten in *MECP2* sind vor allem als Ursache für das Rett-Syndrom, eine X-chromosomale neurologische Entwicklungsstörung, bekannt (Gonzales & LaSalle 2010). Das Rett-Syndrom zählt zu den häufigsten genetischen Ursachen geistiger Behinderung bei Mädchen. Nach einer unauffälligen Entwicklung in den ersten 6-18 Lebensmonaten kommt es zur Stagnation und anschließend zum Verlust bereits erworbener Funktionen. Patienten mit Rett-Syndrom zeigen folgende charakteristische Auffälligkeiten: Teilweiser oder vollständiger Verlust des erlernten, sinnvollen Gebrauchs der Hände, teilweiser oder vollständiger Verlust der Sprache, Gangstörung oder Gangunfähigkeit sowie stereotype Handbewegungen (Schaaf & Zschocke, 2018).

In den untersuchten Familien der Pilotstudie war ID ein Ausschlusskriterium, daher sind unsere Daten ein weiterer Hinweis auf eine Erweiterung des phänotypischen Spektrums der pathogenen Varianten in *MECP2*. Zunehmend werden auch in der Literatur pathogene Varianten in *MECP2* bei Patienten mit SCZ, einem im Vergleich zum Vollbild des Rett-Syndroms mit schwerer ID eher milden Phänotyp, berichtet. In der bereits oben beschriebenen Studie von McCarthy et al. (2014) wurde die *de novo missense* Mutation c.568C>T (p.R190C) bei einem Patienten mit SCZ identifiziert und in einer weiteren SCZ *de novo* Studie die *missense* Mutation c.419C>T (p. A140V) (Howrigan et al., 2020). In einer Exom-Sequenzierungsstudie mit 192 Familien mit nicht-syndromatischer ID wurde die *missense* Mutation c.569G>A (p.R190H) in einer konsanguinen Familie mit SCZ, Depression und milder ID detektiert (Harripaul et al., 2018). Cohen et al. beschrieben in einem Fallbericht die *missense* Mutation c.419C>T (p.A140V) bei einem Jungen mit Sprachentwicklungsstörung und frühmanifesten SCZ (Cohen et al., 2002). Dieselbe pathogene Variante wurde auch innerhalb einer Familie bei mehreren Familienmitgliedern mit ID identifiziert. Bei einigen der Familienmitglieder wurde zusätzlich eine Psychose diagnostiziert (Klauck et al., 2002). Außerdem wurde in einem weiteren Fallbericht die *missense* Mutation c.491G>T (p.S164I) bei dizygoten Zwillingen mit Asperger Syndrom, einer Autismus-Spektrum-Störung, identifiziert, von denen einer zusätzlich an frühmanifesten SCZ erkrankt war (Curie et al., 2017). Zwei der Familienmitglieder, bei denen die Mutation im *MECP2*-Gen in der vorliegenden Studie identifiziert wurde, erkrankten vor dem 16. Lebensjahr. Zusammenfassend deuten die Studienergebnisse darauf hin, dass pathogene Varianten im *MECP2*-Gen zu einem breiten Spektrum an psychiatrischen Symptomen sowie zu einem früheren Erkrankungszeitpunkt und somit einem vergleichsweise schwereren Phänotyp bei Patienten mit SCZ führen können.

Interessanterweise wurden in der Familien-Pilotstudie sowie den oben beschriebenen Studien aus der Literatur sowohl bei männlichen als auch weiblichen Patienten mit SCZ Mutationen in *MECP2* beschrieben (McCarthy et al., 2014; Harripaul et al., 2018; Cohen et al., 2002; Klauck et al., 2002; Curie et al., 2017; Howrigan et al., 2020). Das klassische Rett-Syndrom wird ausschließlich bei weiblichen Patienten diagnostiziert (Kaur & Christodoulou, 1993). In seltenen Fällen werden bei lebend geborenen Jungen pathogene Varianten in *MECP2* nachgewiesen. Diese führen zu einem breiten Spektrum an klinischen Phänotypen, von milder ID bis hin zu schwerer neonataler Enzephalopathie und frühzeitigem Tod. Rett-Mutationen können bei männlichen Patienten zu einem Rett-Phänotyp führen, wenn der Patient das Klinefelter-Syndrom (Karyotyp 47,XXY) hat oder die Variante als Mosaik vorliegt (Ravn et al., 2003; Pascual-Alonso et al., 2021). In den Familien der Pilotstudie gab es basierend auf den Sequenzierungsdaten keinen Hinweis auf ein Mosaik oder ein Klinefelter-Syndrom.

Weiterhin handelt es sich bei *MECP2* um ein dosissensitives Gen. Während *loss-of-function* Mutationen zum Rett-Syndrom führen, führt eine Duplikation des gesamten Gens zum MECP2-Duplikations-Syndrom, einer schweren, syndromalen Form von ID, von der hauptsächlich Jungen betroffen sind (Pascual-Alonso et al., 2021). Das breite Spektrum der phänotypischen Variabilität scheint also mit der Art der Mutation in *MECP2* zu korrelieren und gegebenenfalls spielt auch die Inaktivierung des X-Chromosoms eine Rolle (Cheadle et al., 2000; Amir et al., 2000). Dies wurde jedoch in den untersuchten Familien nicht getestet.

Das Phänomen, dass pathogene Varianten in einem Gen zu unterschiedlichen Phänotypen führen können, ist im Zusammenhang mit SCZ bereits aus CNV-Studien bekannt. Es konnte gezeigt werden, dass die mit SCZ assoziierten CNVs auch das Risiko für andere neuropsychiatrische Erkrankungen wie ID und ASD erhöhen (Rees et al., 2014, 2016). In einer SCZ Exom-Sequenzierungsstudie konnte eine Überschneidung von *de novo loss-of-function* Mutationen bei Patienten mit SCZ und Patienten mit ASD und ID nachgewiesen werden (Fromer et al., 2014). Ebenso sind auch *loss-of-function* Mutationen im oben genannten Gen *SETD1A* neben SCZ auch mit schwerer ID und Entwicklungsverzögerung assoziiert (Singh et al., 2016). Übergeordnet können all diese Diagnosen (SCZ, ID, ASD) als Entwicklungsstörungen des Gehirns beschrieben werden. Warum es zu unterschiedlichen phänotypischen Manifestationen der seltenen Varianten kommt, ist noch nicht vollständig verstanden. Jedoch wird angenommen, dass zum Beispiel häufige Varianten dabei eine Phänotyp-modifizierende Rolle spielen und dass z.B. CNV-Träger mit SCZ zusätzlich eine hohe Last an häufigen SCZ Risikoallelen besitzen (Tansey et al., 2016). Bei dem angenommenen pleiotropen Effekt einer seltenen Variante könnte es sich also um ein Zusammenspiel von verschiedenen genetischen und auch umweltbedingten Faktoren handeln, welche Auswirkungen auf verschiedene Bereiche der Hirnfunktion haben (O'Donovan & Owen, 2016).

Das Gen *HDAC5* kodiert für ein Protein aus der Familie der Histon-Deacetylasen, welche Acetylgruppen von Histonen entfernen und dadurch die Chromatinstruktur verändern und die Bindung von Transkriptionsfaktoren an die DNA beeinflussen können (Shahbazian & Grunstein, 2007). In einer *de novo* Studie von Howrigan et al. (2020) wurde bei einem Patienten mit SCZ die *missense* Mutation c.3275T>C (p.L1092S) nachgewiesen. Eine Dysregulierung der Histon-Acetylierung wurde in der Literatur außerdem in Zusammenhang mit dem Rubinstein-Taybi-Syndrom, einer neurologischen Entwicklungsstörung, die durch kognitive Beeinträchtigungen gekennzeichnet ist, beschrieben (Peixoto & Abel, 2013). In einer Studie mit *HDAC5 knockout* Mäusen zeigten die Mäuse depressionsähnliche Verhaltensweisen als Reaktion auf chronischen Stress (Renthal et al., 2007).

Ein weiteres potentielles SCZ Kandidatengen ist *SMARCC1*. Dieses Gen kodiert für ein Protein aus der SWI/SNF (*SWItch/Sucrose Non-Fermentable*) Familie. SWI/SNF Komplexe verändern die Chromatinstruktur durch mittels Adenosintriphosphat (ATP)-Hydrolyse gewonnene Energie und erhöhen dadurch die Zugänglichkeit für Transkriptionsfaktoren (Sokpor et al., 2017). Studien konnten zeigen, dass SWI/SNF-Komplexe an der Aufrechterhaltung der neuronalen Entwicklung beteiligt sind und dass pathogene Varianten in diesen an neurologischen Entwicklungsstörungen wie z.B. ASD oder ID beteiligt sind (Ronan et al., 2013; Valencia et al., 2023). Weitere Gene, die für Proteine aus dieser Familie kodieren sind unter anderem *ARID1B*, *ARID2* und *SMARCC2*. In diesen Genen wurden bereits Mutationen bei Patienten mit SCZ (Fromer et al., 2014; Rees et al., 2020; Howrigan et al., 2020; Toyama et al., 2022), ID und ASD beschrieben (Backx et al., 2011; D’Gama et al., 2015; Halgren et al., 2012; Shang et al., 2015; O’Roak et al., 2014; Neale et al., 2012; De Rubeis et al., 2014; O’Roak et al., 2012). Außerdem konnten in der HIPPIE Datenbank auch Interaktionen zwischen diesen drei Genen und *SMARCC1* nachgewiesen werden. In der Literatur wurden bisher noch keine Duplikationen in *SMARCC1* bei Patienten mit SCZ beschrieben. Dies könnte unter anderem daran liegen, dass die Duplikationen nicht die Einschlusskriterien (z.B. Größe der CNVs oder Anzahl der betroffenen Sonden auf dem verwendeten Microarray) der bisher publizierten genomweiten CNV-Analysen erfüllten.

Das in dieser Studie identifizierte Kandidatengen *FOXA1* spielt ebenfalls eine Rolle bei der Chromatinmodifikation. *FOXA1* gehört zur FOXA-Familie der Transkriptionsfaktoren. Diese dienen als Pionier-Faktoren und können kondensiertes Chromatin öffnen und dadurch die Bindung anderer Transkriptionsfaktoren, Chromatin- und Nukleosom-Modifikatoren ermöglichen (Zaret & Carroll, 2011). Bisher sind noch keine pathogenen Varianten in *FOXA1* im Zusammenhang mit SCZ beschrieben. Jedoch spielt *Foxa1* neben weiteren Transkriptionsfaktoren eine Rolle bei der Entwicklung von mesencephalen dopaminergen Neuronen (mDA) (Ferri et al., 2007; Tian et al., 2022). Diese steuern verschiedene Gehirnfunktionen, wie zum Beispiel willkürliche Bewegungen und kognitive Prozesse. Die Degeneration von mDA spielt bei der Parkinson-Krankheit eine Rolle, während eine veränderte Dopaminübertragung mit neuropsychiatrischen Störungen wie SCZ, Depression oder Autismus assoziiert ist (Grace, 2016; Marotta et al., 2020; Sonnenschein et al., 2020; Islam et al., 2021).

Neben den vier Genen, die in die Chromatinmodifikation involviert sind, wurde auch das Gen *SRRM2*, welches am Spleißen der prä-messenger RNA beteiligt ist, in dieser Studie als Kandidatengen identifiziert. In zwei großen SCZ Fall-Kontroll Exom-Sequenzierungsstudien waren sehr seltene pathogene Varianten in *SRRM2* bei Patienten im Vergleich zu Kontrollen angereichert (Singh et al., 2022; Liu et al., 2023). In der Studie von Liu et al. (2023) erreichte

das Gen eine exomweite Signifikanz. Weiterhin wurde die *missense* Mutation c.6233G>A (p.R2078H) in *SRRM2* bei einer SCZ multiplex Familie identifiziert (Toyama et al., 2022). Bei Patienten mit ASD und Entwicklungsverzögerung wurden *de novo* Mutationen in *SRRM2* beschrieben (Turner et al., 2019).

Zusammenfassend konnten durch diese Familien-basierte Exom-Sequenzierungsstudie fünf SCZ Kandidatengene – *HDAC5*, *MECP2*, *SRRM2*, *SMARCC1* und *FOXA1* – identifiziert werden. Vier dieser Gene spielen eine Rolle bei der Chromatinmodifikation, was im Einklang mit früheren Studien steht und weitere Hinweise für einen möglichen Zusammenhang von chromatinregulierenden Genen und der SCZ Pathogenese liefert.

5.2 CNV-Studie – Identifizierung exonischer Kopienzahlvarianten in einer großen Kohorte von SCZ multiplex Familien

Nach aktuellem Wissensstand ist die vorliegende CNV-Studie eine der größten bisher durchgeführten NGS-basierten Analysen von CNVs in Familien mit mehreren Patienten mit SCZ und stellt auch aufgrund der stringenten Qualitätskontrolle eine einzigartige Datenressource dar. In einer 2018 publizierten Studie wurden 91 SCZ multiplex Familien genomweit sequenziert und anschließend auf CNVs untersucht (Khan et al., 2018). Jedoch waren in diesen Familien durchschnittlich 1,2 Familienmitglieder an SCZ erkrankt, wohingegen in der vorliegenden Studie durchschnittlich 2,8 Familienmitglieder betroffen waren. Die in der vorliegenden Arbeit identifizierten CNVs wurden in der Studie von Khan et al. nicht berichtet.

Insgesamt wurden in der CNV-Studie 13 Gene identifiziert, die von co-segregierenden seltenen Deletionen oder Duplikationen betroffen waren. Durch zahlreiche multimodale Ansätze (weitere genetische Evidenz, Genexpression, PPI Analysen) wurden die Kandidatengene hinsichtlich ihrer Krankheitsrelevanz priorisiert. Für zwei Kandidatengene (*RSRC2* und *ZCCHC8*) wurde eine signifikante Assoziation zwischen häufigen Varianten und einem Risiko für SCZ nachgewiesen. Interessanterweise gehörte das Gen *ZCCHC8* auch zu den mittels PPI Analysen am höchsten priorisierten Genen und ist im Gehirn exprimiert.

ZCCHC8 (*zinc finger CCHC-type containing 8*) kodiert für ein Protein aus der Zinkfinger CCHC-Typ Familie, welche an den verschiedenen Schritten des RNA-Stoffwechsels beteiligt ist. *ZCCHC8* bildet mit zwei weiteren Proteinen einen Komplex (NEXT-Komplex) und lenkt nicht-kodierende RNAs zum Exosom, wo diese abgebaut werden (Wang et al., 2021). Bisher sind in der Literatur noch keine pathogenen Varianten in *ZCCHC8* im Zusammenhang mit SCZ beschrieben. Jedoch wurde eine Stopp-Mutation in *ZCCHC8* bei einer Familie mit nicht-syndromaler ID (Najmabadi et al., 2011) sowie eine stille *de novo* Mutation bei einem Patienten mit ASD nachgewiesen (Zhou et al., 2022). Beide Diagnosen zählen wie SCZ ebenfalls zu den Entwicklungsstörungen des Gehirns.

Eine eindeutige Krankheitsrelevanz konnte für kein Kandidatengen der CNV-Studie nachgewiesen werden. Dies ist aufgrund der bekannten genetischen Heterogenität und Komplexität der Erkrankung nicht ungewöhnlich. Lediglich 10 Gene erreichten in der bisher größten SCZ Fall-Kontroll Studie exomweite Signifikanz für eine Anreicherung von sehr seltenen pathogenen Varianten bei Patienten im Vergleich zu Kontrollen (Singh et al., 2022). Bei ASD zum Beispiel sind bisher klar pathogene Varianten (bereits in der klinischen Routine Diagnostik etabliert) in ~ 800 Genen bekannt (Genovese & Butler, 2020). Daher ist anzunehmen, dass bei SCZ noch deutlich mehr krankheitsrelevante Gene gefunden werden.

Die Anwendung einer sehr stringenten Filterstrategie mit Fokussierung auf seltene und hoch penetrante CNVs ist ein zweischneidiges Schwert. Einerseits wird dadurch die Anzahl falsch positiver Ergebnisse reduziert, andererseits wird aber auch die Wahrscheinlichkeit erhöht, dass potentiell krankheitsursächliche Deletionen und Duplikationen in den untersuchten Familien übersehen wurden. Aus technischen Gründen wurden im Rahmen der vorliegenden Studie außerdem ausschließlich CNVs auf den Autosomen untersucht, CNVs auf den Geschlechtschromosomen wurden nicht analysiert. In anderen Studien (Marshall et al., 2017; Bache & DeLisi, 2018) konnte bereits gezeigt werden, dass diese eine Rolle bei der Pathogenese der SCZ spielen. Basierend auf den vorliegenden Stammbauminformationen wäre in den untersuchten Familien eine geschlechtschromosomale Vererbung denkbar. Daher werden diese in einer zukünftigen Studie erneut auf CNVs auf den Geschlechtschromosomen untersucht werden.

Weiterhin sind sowohl die Generierung als auch die Interpretation von CNV-Daten basierend auf Exom-Sequenzierungen eine große Herausforderung. In allen in der vorliegenden Studie verwendeten CNV-Algorithmen wurde die Abdeckung der Reads (*read depth*) zur Detektion der CNVs verwendet. Andere zum Zeitpunkt der Studiendurchführung verfügbare Methoden (z.B. *paired-end mapping*) sind weniger geeignet zur Analyse von Exom-Daten. Diese untersuchen die Bruchpunkte der CNVs, welche häufig in den intronischen Bereichen eines Gens liegen. Bei der Exom-Sequenzierung werden jedoch nur die Exone und flankierende Bereiche abgedeckt (Tan et al., 2014). Aufgrund des rasanten Fortschritts bei der Entwicklung bioinformatischer Methoden ist eine zukünftige erneute Auswertung der Daten mit verbesserten Methoden sinnvoll. Dabei sollten auch weitere genetische Informationen wie zum Beispiel seltene, pathogene SNVs oder polygene Risikoscores in die Analysen integriert werden.

Bisher gibt es noch keine großen Datensätze von CNVs aus NGS-basierten Studien von SCZ multiplex Familien, die zur Replikation der vorliegenden Ergebnisse geeignet wären. Da diese jedoch durch internationale Kooperationen zunehmend verfügbar werden, werden unsere Ergebnisse in Zukunft in unabhängigen Datensätzen weiterverfolgt werden können.

5.3 Ausblick

Durch die Untersuchung von über 50 SCZ multiplex Familien mittels Exom-Sequenzierung konnten zahlreiche potentielle SCZ Suszeptibilitätsgene identifiziert werden. Die neuen Erkenntnisse sind relevant für ein besseres Verständnis der biologischen Grundlagen der SCZ. Jedoch sind diese noch nicht ausreichend für eine Translation in die klinische Diagnostik.

Eine Möglichkeit, zusätzliche genetische Evidenz für die detektierten Gene zu erlangen, ist der Nachweis einer Anreicherung pathogener Varianten in einem unabhängigen Kollektiv von Patienten mit SCZ im Vergleich zu einer Kontrollkohorte. Daher wurden in einem Folgeprojekt, an dem ich wesentlich beteiligt war, drei der in den hier vorgestellten Studien implizierten Gene (*HDAC5*, *MECP2* und *SMARCC1*) in einem großen Kollektiv von 1.947 Patienten mit SCZ und 1.921 Kontrollen mittels *single molecule molecular inversion probes* (smMIPs) sequenziert. Eine schematische Darstellung der smMIP Methode ist in Abbildung A1 dargestellt. Das Design der smMIPs sowie die Sequenzierung wurden von mir selbst durchgeführt. Da diese Methode im Labor neu etabliert wurde, nahm dies erhebliche Zeit in Anspruch. Die bioinformatische Auswertung fand erst nach Beendigung der Laborarbeit statt und war bei der Erstellung dieser Arbeit noch nicht abgeschlossen. Die Analyse der Daten stellt eine große Herausforderung dar und wird in Zusammenarbeit mit Sugirthan Sivalingam und Andreas Buness (Core Unit for Bioinformatics Data Analysis, Universitätsklinikum Bonn) durchgeführt.

Neben den drei oben genannten Genen wurden in diesem Projekt auch weitere 33 SCZ Kandidatengene aus der Literatur sequenziert (Tabelle A8). Insgesamt wurden für die Sequenzierung der 36 ausgewählten Kandidatengene 1.458 smMIPs designt. In einer anschließenden *Burden* Analyse wird untersucht, ob es in einem der Gene eine Häufung seltener Varianten bei den Patienten im Vergleich zu den Kontrollen gibt. Ein erhöhter *Burden* an seltenen Varianten im Patienten-Kollektiv wäre weitere genetische Evidenz für das implizierte Gen.

Zusätzlich zu einer erneuten bioinformatischen Auswertung der vorliegenden Daten wird zukünftig auch die Generierung von genomweiten Sequenzierungsdaten wichtig sein. Diese Methode ist derzeit noch mit hohen Kosten verbunden und stellt aufgrund der enormen Komplexität der Datenauswertung eine große Herausforderung dar. Dennoch wäre eine genomweite Sequenzierung sinnvoll, da gezeigt werden konnte, dass diese im Vergleich zur exomweiten Sequenzierung zu einem qualitativ hochwertigeren Exom führt (Belkadi et al., 2015). Weiterhin bietet die Genom-Sequenzierung den Vorteil, auch Varianten in nicht-kodierenden Bereichen detektieren zu können, welche zum Beispiel die Genexpression oder das Spleißen beeinflussen.

Außerdem wird auch die Identifizierung weiterer struktureller Varianten eine wichtige Rolle spielen. Studien konnten zeigen, dass mittels neuerer *long read* Technologien strukturelle Varianten wie CNVs und Inversionen effektiver detektiert werden können als mit *short read* Sequenzierungen, wie der hier verwendeten Exom-Sequenzierung (Pauper et al., 2021). Inversionen waren in der vorliegenden CNV-Studie nicht nachweisbar, da es sich hierbei um balancierte chromosomale Veränderungen handelt. Nicht auszuschließen wäre ein klinischer Effekt durch einen Bruch innerhalb eines Gens. Aber auch durch die erneute Auswertung der vorliegenden Daten mit verbesserten bioinformatischen Methoden könnten weitere strukturelle Varianten identifiziert werden. Durch die zusätzlichen genetischen Analysen der SCZ multiplex Familien sowie die Untersuchung der identifizierten Varianten in größeren Kohorten wird eine bessere Einschätzung der Kandidatengene hinsichtlich ihrer Krankheitsrelevanz möglich sein.

Da es bereits Hinweise auf interaktive Effekte von polygenen Risikoscores, einem Maß für die kumulativen Auswirkungen häufiger krankheitsassoziiierter Varianten, und bereits bekannten CNVs auf das Risiko für SCZ gibt (Bergen et al., 2019), wird auch die Untersuchung der Kombination von seltenen und häufigen Varianten zur Aufklärung der fehlenden Heritabilität der SCZ beitragen.

Ebenso rücken auch epigenetische Analysen in den Fokus der SCZ Forschung. Es wird angenommen, dass zusätzlich zu den genetischen Risikofaktoren auch Veränderungen im Epigenom an der Pathogenese verschiedener Entwicklungsstörungen des Gehirns beteiligt sind und dass epigenetische Mechanismen wie DNA-Methylierung und Histon-Modifikation für die normale Entwicklung des Gehirns eine wichtige Rolle spielen (Reichard & Zimmer-Bensch 2021). Für die hier untersuchten Familien ist dies nicht zu klären, da zu Beginn meiner Doktorarbeit keine epigenetischen Daten der Familien vorlagen.

Da durch den alleinigen Nachweis pathogener Varianten bei Patienten mit SCZ nicht der funktionelle Effekt einer Variante erklärt werden kann, können funktionelle Studien helfen, die biologischen Grundlagen der Krankheitsentstehung zu verstehen. Gut geeignet hierfür sind Patienten-spezifische induzierte pluripotente Stammzellen (*induced pluripotent stem cells*; iPSCs). Diese können aus peripherem Blut von Patienten gewonnen und anschließend in neuronale Zellen differenziert werden. In Studien an iPSCs von Patienten mit SCZ konnten unter anderem eine verringerte neuronale Konnektivität und Aktivität von Nervenzellen, eine verringerte spontane Freisetzung von Neurotransmittern sowie Defekte in der synaptischen Dichte und Arborisierung von kortikalen Interneuronen beobachtet werden (Brennand et al., 2011; Yu et al., 2014; Shao et al., 2019). Bei den hier untersuchten Familien war eine Generierung von iPSCs nicht möglich, da die Patienten vor langer Zeit rekrutiert wurden. Daher waren keine für diese Untersuchung benötigten Blutproben und Einverständniserklärungen vorhanden. Die Rekrutierung von multiplex Familien läuft an den unterschiedlichen

Standorten jedoch weiter. Die Entnahme von entsprechendem Material für die Generierung von iPSCs gehört mittlerweile zum Studienprotokoll.

Eine große Herausforderung der genetischen Grundlagenforschung ist die Translation ihrer Ergebnisse in die klinische Praxis. In Deutschland ist die genetische Diagnostik zur Ursachenabklärung bei Menschen mit ID bereits in die entsprechende klinische Leitlinie verankert (<https://www.awmf.org/leitlinien/detail/II/028-042.html>). Weiterhin wird bei Menschen mit ASD und zusätzlichen Symptomen wie Epilepsie oder Organfehlbildungen eine genetische Untersuchung empfohlen (<https://www.awmf.org/leitlinien/detail/II/028-018.html>). Diese ist seit mehreren Jahren Leistung der gesetzlichen Krankenkassen.

Bisher gibt es keine einheitlichen Empfehlungen für eine genetische Diagnostik bei Patienten mit SCZ. Allerdings gibt es bereits nationale und internationale Bestrebungen, genetische Diagnostik bei ausgewählten Patienten mit SCZ (z.B. bei frühmanifesten Psychose) zu veranlassen (<https://www.cost.eu/actions/CA17130/>, Leitung PD Dr. med. Franziska Degenhardt). Klinisch wäre zum Beispiel der Nachweis einer 22q11.2 Deletion relevant. Die Deletion führt zu einem im Vergleich zu Kontrollen ca. 60-fach erhöhten Risiko an SCZ zu erkranken (Marshall et al., 2017) und ist eine bekannte Ursache für ASD und ID (McDonald-McGinn et al., 1993). Insbesondere bei jüngeren Patienten mit SCZ kann der Nachweis einer genetischen Ursache für ein breites Spektrum an Symptomen das Ende einer diagnostischen Odyssee bedeuten, da genetische Syndrome in der Regel mehrere Organe betreffen. Die Patienten können dadurch auch Zugang zu syndromspezifischen Selbsthilfegruppen sowie einer besseren somatischen und psychiatrischen Gesundheitsversorgung erhalten. Patienten mit einer diagnostizierten 22q11.2 Deletion haben zum Beispiel ein erhöhtes Risiko für angeborene Herzerkrankungen und sollten dahingehend von einem Kardiologen untersucht werden (McDonald-McGinn et al., 1993). Außerdem sprechen Patienten mit einer 22q11.2 Deletion und einer Psychose gut auf Clozapin, einem atypischen Antipsychotikum, an (Butcher et al., 2015). Da jedoch sowohl Clozapin als auch die Deletion die Schwelle für epileptische Anfälle senken, sollte die Einnahme in Kombination mit einem Antiepileptikum erwogen werden (Wither et al., 2017). Weiterhin kann das Verständnis der Ätiologie der eigenen Erkrankung einen psychologischen Nutzen für die Patienten haben und zur Entstigmatisierung der Erkrankung beitragen und auch für Eltern kann die Diagnose eines genetischen Syndroms als Ursache für die psychiatrischen Symptome ihres Kindes eine Erleichterung darstellen (Degenhardt et al., 2021).

6 Zusammenfassung

Schizophrenie (SCZ) ist eine schwere neuropsychiatrische Erkrankung mit einem Lebenszeitrisko von ca. 1% (Perälä et al., 2007) und einer geschätzten Heritabilität von etwa 60-80% (Hilker et al., 2018; Wray & Gottesman, 2012). Die genetische Architektur der SCZ ist komplex und an der Entstehung der Erkrankung ist ein großes frequenzüberspannendes Spektrum an genetischen Varianten beteiligt. Dieses beinhaltet sowohl häufigere Varianten mit einer niedrigeren Penetranz als auch seltenere Varianten mit einer höheren Penetranz (McCarthy et al., 2008; Sullivan et al., 2012). Häufige Varianten werden in der Regel mittels genomweiter Assoziationsstudien (GWAS) identifiziert und der Anteil an der phänotypischen Varianz, welcher durch häufige Varianten erklärt werden kann, wird derzeit auf etwa 20-30% geschätzt (Trubetskoy et al., 2022; Pardiñas et al., 2018). Eine der großen Herausforderungen der GWAS ist, dass die identifizierten häufigen Varianten nicht direkt einem Gen zugeordnet werden können.

Durch die Einführung von Next Generation Sequencing (NGS) Technologien rückten seltenere, höher penetrante Varianten, welche mit früheren Methoden nicht nachweisbar waren, in den Fokus. Diese Technologien ermöglichen die Sequenzierung ganzer Exome und somit auch die Identifizierung seltener proteinkodierender Einzelnukleotid-Varianten sowie kleinerer Kopienzahlvarianten (CNVs), welche einzelne Gene direkt betreffen. Dies erleichtert die Identifizierung neuer krankheitsrelevanter Gene. Aufgrund der großen Anzahl von Varianten pro menschlichem Exom ist die Unterscheidung zwischen phänotypneutralen und phänotyprelevanten Varianten eine große Herausforderung. Ein vielversprechender Ansatz zur Identifizierung von krankheitsrelevanten Varianten ist die Analyse von sogenannten multiplex Familien, in welchen über mehrere Generationen hinweg Familienmitglieder erkrankt sind. Es ist anzunehmen, dass in diesen Familien mindestens ein genetischer Risikofaktor mit großem Effekt zur Krankheitsentstehung beiträgt und innerhalb der Familie mit der Erkrankung co-segregiert.

Daher wurden in einer Pilotstudie zunächst drei multiplex Familien mit jeweils mindestens drei Familienmitgliedern mit SCZ mittels Exom-Sequenzierung auf co-segregierende, potentiell pathogene seltene CNVs und kleinere Sequenzvarianten untersucht. Insgesamt wurden 35 Einzelnukleotid-Varianten und Indels sowie zwei Duplikationen in den chromosomalen Regionen 3p21.31 und 8q21.13 detektiert. Zur besseren Einschätzung der Krankheitsrelevanz der identifizierten Gene wurden diese anschließend mittels unterschiedlicher Strategien priorisiert. Diese beinhalteten den Abgleich der eigenen Daten mit bereits publizierten Studien zu Entwicklungsstörungen des Gehirns, Gen-basierte Tests, Protein-Protein Interaktionsanalysen sowie Genexpressionsanalysen.

Die fünf hinsichtlich ihrer Krankheitsrelevanz am höchsten priorisierten Gene waren *MECP2* (*methyl CpG binding protein 2*), *HDAC5* (*histone deacetylase 5*), *SRRM2* (*serine/arginine repetitive matrix 2*), *SMARCC1* (*SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 1*) und *FOXA1* (*forkhead box A1*). Vier dieser Gene (*MECP2*, *HDAC5*, *SMARCC1* und *FOXA1*) scheinen eine Rolle bei der Chromatinmodifikation zu spielen. Dies steht in Einklang mit früheren Studien, in denen bereits eine Assoziation von chromatinregulierenden Genen mit SCZ beschrieben wurde und liefert weitere Hinweise für einen möglichen Zusammenhang zwischen der Chromatinmodifikation und der SCZ Pathogenese.

In einer anschließenden Studie wurden 47 weitere exomweit sequenzierte SCZ multiplex Familien mit jeweils mindestens drei betroffenen Familienmitgliedern auf seltene, potentiell pathogene CNVs untersucht. In dieser Studie wurden fünf Deletionen in den chromosomalen Regionen 1p36.13, 14q11.2, 14q22.2, 14q24.3 und 16q23.2, sowie drei Duplikationen in den Regionen 6p12.1, 9q33.2 und 12q24.31 detektiert. Zur Priorisierung der identifizierten Gene wurde auch in dieser Studie eine multimodale Analysestrategie angewandt (genetische und Genexpressionsdaten, Protein-Protein Interaktionsanalysen).

Das höchstpriorisierte Gen war *ZCCHC8* (zinc finger CCHC-type containing 8), ein im Gehirn exprimiertes Gen, welches am Abbau von nukleären RNAs beteiligt ist. Eine eindeutige Krankheitsrelevanz konnte jedoch für kein Kandidatengen nachgewiesen werden, was aufgrund der bekannten genetischen Heterogenität und Komplexität der Erkrankung nicht ungewöhnlich ist. Trotzdem stellt diese Studie eine einzigartige Datenressource dar. Unser Datensatz deutet auf neue Gene in der Pathogenese der SCZ hin und wird zukünftig mit neueren, verbesserten Methoden erneut ausgewertet werden.

In einem Folgeprojekt, welches bei Erstellung dieser Arbeit noch nicht abgeschlossen war, wurden neben 33 SCZ Kandidatengenen aus der Literatur drei der in den hier vorgestellten Studien implizierten Gene (*HDAC5*, *MECP2* und *SMARCC1*) in einem großen Kollektiv von 1.947 Patienten mit SCZ und 1.921 Kontrollen sequenziert, um zu untersuchen, ob es in einem der Gene eine Häufung von seltenen Varianten bei den Patienten im Vergleich zu den Kontrollen gibt.

Zusammenfassend konnten im Rahmen dieser Arbeit durch die Auswertung der Exom-Sequenzierungsdaten von 50 SCZ multiplex Familien – dem aktuell weltweit größten Kollektiv – und anschließender Genexpressions- und Protein-Protein Interaktionsanalysen sowie Literatur- und Datenbankrecherchen mehrere potentielle neue SCZ Kandidatengene identifiziert werden. Diese Ergebnisse sind relevant für ein besseres Verständnis der biologischen Grundlagen der SCZ. Für eine Translation in die klinische Diagnostik sind diese

jedoch noch nicht ausreichend. Um weitere Evidenz für die implizierten Gene zu erhalten, sollten zukünftig neben der Untersuchung von großen Patienten- und Kontrollkollektiven auch weitere Analysen, wie zum Beispiel funktionelle Studien mittels iPSCs durchgeführt werden.

7 Literaturverzeichnis

- 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., et al. (2015) 'A global reference for human genetic variation', *Nature*, 526(7571), pp. 68–74.
- Ahn, K., Gotay, N., Andersen, T.M., Anvari, A.A., Gochman, P., et al. (2014) 'High rate of disease-related copy number variations in childhood onset schizophrenia', *Molecular Psychiatry*, 19(5), pp. 568–572.
- Aleman, A., Kahn, R.S. & Selten, J.-P. (2003) 'Sex differences in the risk of schizophrenia: evidence from meta-analysis', *Archives of General Psychiatry*, 60(6), pp. 565–571.
- Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester, H.E., et al. (2017) 'The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers', *Nature Communications*, 8p. 14648.
- Ambalavanan, A., Girard, S.L., Ahn, K., Zhou, S., Dionne-Laporte, A., et al. (2016) 'De novo variants in sporadic cases of childhood onset schizophrenia', *European journal of human genetics: EJHG*, 24(6), pp. 944–948.
- American Psychiatric Association (1980) *Diagnostic and statistical manual of mental disorders: DSM-III*. Washington (DC): American Psychiatric Association.
- American Psychiatric Association (1987) *Diagnostic and statistical manual of mental disorders: DSM-III-R*. Washington (DC): American Psychiatric Association.
- American Psychiatric Association (1994) *Diagnostic and statistical manual of mental disorders: DSM-IV, 4th ed. Washington (DC): American Psychiatric Association.*, p. 886.
- Amir, R.E., Van den Veyver, I.B., Schultz, R., Malicki, D.M., Tran, C.Q., et al. (2000) 'Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes', *Annals of Neurology*, 47(5), pp. 670–679.
- Bache, W.K. & DeLisi, L.E. (2018) 'The Sex Chromosome Hypothesis of Schizophrenia: Alive, Dead, or Forgotten? A Commentary and Review', *Molecular Neuropsychiatry*, 4(2), pp. 83–89.
- Backx, L., Seuntjens, E., Devriendt, K., Vermeesch, J. & Van Esch, H. (2011) 'A balanced translocation t(6;14)(q25.3;q13.2) leading to reciprocal fusion transcripts in a patient with intellectual disability and agenesis of corpus callosum', *Cytogenetic and Genome Research*, 132(3), pp. 135–143.
- Balakrishna, T. & Curtis, D. (2020) 'Assessment of Potential Clinical Role for Exome Sequencing in Schizophrenia', *Schizophrenia Bulletin*, 46(2), pp. 328–335.
- Barabási, A.-L., Gulbahce, N. & Loscalzo, J. (2011) 'Network Medicine: A Network-based Approach to Human Disease', *Nature reviews. Genetics*, 12(1), pp. 56–68.
- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., et al. (2013) 'NCBI GEO: archive for functional genomics data sets—update', *Nucleic Acids Research*, 41(D1), pp. D991–D995.
- Barsh, G.S., Copenhaver, G.P., Gibson, G. & Williams, S.M. (2012) 'Guidelines for Genome-Wide Association Studies', *PLoS Genetics*, 8(7).

- de Baumont, A., Maschietto, M., Lima, L., Carraro, D.M., Olivieri, E.H., et al. (2015) 'Innate immune response is differentially dysregulated between bipolar disease and schizophrenia', *Schizophrenia Research*, 161(2–3), pp. 215–221.
- Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q.B., et al. (2015) 'Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants', *Proceedings of the National Academy of Sciences*, 112(17), pp. 5473–5478.
- Bergen, S.E., Ploner, A., Howrigan, D., CNV Analysis Group and the Schizophrenia Working Group of the Psychiatric Genomics Consortium, O'Donovan, M.C., et al. (2019) 'Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk for Schizophrenia', *The American Journal of Psychiatry*, 176(1), pp. 29–35.
- Brennand, K., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., et al. (2011) 'Modeling schizophrenia using hiPSC neurons', *Nature*, 473(7346), pp. 221–225.
- Brown, A.S., Schaefer, C.A., Quesenberry, C.P., Liu, L., Babulas, V.P., et al. (2005) 'Maternal exposure to toxoplasmosis and risk of schizophrenia in adult offspring', *The American Journal of Psychiatry*, 162(4), pp. 767–773.
- Butcher, N.J., Fung, W.L.A., Fitzpatrick, L., Guna, A., Andrade, D.M., et al. (2015) 'Response to clozapine in a clinically identifiable subtype of schizophrenia', *The British Journal of Psychiatry: The Journal of Mental Science*, 206(6), pp. 484–491.
- Catts, V.S., Wong, J., Fillman, S.G., Fung, S.J. & Shannon Weickert, C. (2014) 'Increased expression of astrocyte markers in schizophrenia: Association with neuroinflammation', *The Australian and New Zealand Journal of Psychiatry*, 48(8), pp. 722–734.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., et al. (2015) 'Second-generation PLINK: rising to the challenge of larger and richer datasets', *GigaScience*, 4p. 7.
- Cheadle, J.P., Gill, H., Fleming, N., Maynard, J., Kerr, A., et al. (2000) 'Long-read sequence analysis of the MECP2 gene in Rett syndrome patients: correlation of disease severity with mutation type and location', *Human Molecular Genetics*, 9(7), pp. 1119–1129.
- Chen, C., Cheng, L., Grennan, K., Pibiri, F., Zhang, C., et al. (2013) 'Two gene co-expression modules differentiate psychotics and controls', *Molecular Psychiatry*, 18(12), pp. 1308–1314.
- Chen, C.-H., Huang, A., Huang, Y.-S. & Fang, T.-H. (2021) 'Identification of a Rare Novel KMT2C Mutation That Presents with Schizophrenia in a Multiplex Family', *Journal of Personalized Medicine*, 11(12), p. 1254.
- Chen, J., Aronow, B.J. & Jegga, A.G. (2009) 'Disease candidate gene identification and prioritization using protein interaction networks', *BMC bioinformatics*, 10p. 73.
- Chen, J., Xu, H., Aronow, B.J. & Jegga, A.G. (2007) 'Improved human disease candidate gene prioritization using mouse phenotype', *BMC bioinformatics*, 8p. 392.
- Cohen, D., Lazar, G., Couvert, P., Desportes, V., Lippe, D., et al. (2002) 'MECP2 mutation in a boy with language disorder and schizophrenia', *The American Journal of Psychiatry*, 159(1), pp. 148–149.
- Collins, R.L., Glessner, J.T., Porcu, E., Lepamets, M., Brandon, R., et al. (2022) 'A cross-disorder dosage sensitivity map of the human genome', *Cell*, 185(16), pp. 3041–3055.e25.

- Curie, A., Lesca, G., Bussy, G., Manificat, S., Arnaud, V., et al. (2017) 'Asperger syndrome and early-onset schizophrenia associated with a novel MECP2 deleterious missense variant', *Psychiatric Genetics*, 27(3), pp. 105–109.
- De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., et al. (2014) 'Synaptic, transcriptional, and chromatin genes disrupted in autism', *Nature*, 515(7526), pp. 209–215.
- Degenhardt, F., Gradl-Dietsch, G. & Hebebrand, J. (2021) 'Need for psychiatric phenotyping in patients with rare genetic disorders', *European Child & Adolescent Psychiatry*, 30(3), pp. 327–329.
- Degenhardt, F., Priebe, L., Meier, S., Lennertz, L., Streit, F., et al. (2013) 'Duplications in RB1CC1 are associated with schizophrenia; identification in large European sample sets', *Translational Psychiatry*, 3(11), p. e326.
- D’Gama, A.M., Pochareddy, S., Li, M., Jamuar, S.S., Reiff, R.E., et al. (2015) 'Targeted DNA Sequencing from Autism Spectrum Disorder Brains Implicates Multiple Genetic Mechanisms', *Neuron*, 88(5), pp. 910–917.
- Dilling, H., Mombour, W. & Schmidt, M.H. (2015) Google-Books-ID: QoTYsgEACAAJ. *Internationale Klassifikation psychischer Störungen: ICD-10 Kapitel V (F) - Klinisch-diagnostische Leitlinien*. Hogrefe.
- Edgar, R., Domrachev, M. & Lash, A.E. (2002) 'Gene Expression Omnibus: NCBI gene expression and hybridization array data repository', *Nucleic Acids Research*, 30(1), pp. 207–210.
- Egawa, J., Hoya, S., Watanabe, Y., Nunokawa, A., Shibuya, M., et al. (2016) 'Rare UNC13B variations and risk of schizophrenia: Whole-exome sequencing in a multiplex family and follow-up resequencing and a case-control study', *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*,
- Faraone, S.V., Tsuang, M.T. & Tsuang, D.W. (1999) *Genetics of Mental Disorders: What Practitioners and Students Need to Know*. New York, London: The Guilford Press.
- Ferri, A.L.M., Lin, W., Mavromataki, Y.E., Wang, J.C., Sasaki, H., et al. (2007) 'Foxa1 and Foxa2 regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner', *Development (Cambridge, England)*, 134(15), pp. 2761–2769.
- Feuk, L., Carson, A.R. & Scherer, S.W. (2006) 'Structural variation in the human genome', *Nature Reviews. Genetics*, 7(2), pp. 85–97.
- Fillman, S.G., Cloonan, N., Catts, V.S., Miller, L.C., Wong, J., et al. (2013) 'Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia', *Molecular Psychiatry*, 18(2), pp. 206–214.
- Fromer, M., Moran, J.L., Chambert, K., Banks, E., Bergen, S.E., et al. (2012) 'Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth', *American Journal of Human Genetics*, 91(4), pp. 597–607.
- Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., et al. (2014) 'De novo mutations in schizophrenia implicate synaptic networks', *Nature*, 506(7487), pp. 179–184.
- Genovese, A. & Butler, M.G. (2020) 'Clinical Assessment, Genetics, and Treatment Approaches in Autism Spectrum Disorder (ASD)', *International Journal of Molecular Sciences*, 21(13), p. 4726.

- Genovese, G., Fromer, M., Stahl, E.A., Ruderfer, D.M., Chambert, K., et al. (2016) 'Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia', *Nature Neuroscience*, 19(11), pp. 1433–1441.
- Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., et al. (2011) 'Increased exonic de novo mutation rate in individuals with schizophrenia', *Nature Genetics*, 43(9), pp. 860–863.
- Gonzales, M.L. & LaSalle, J.M. (2010) 'The role of MeCP2 in brain development and neurodevelopmental disorders', *Current Psychiatry Reports*, 12(2), pp. 127–134.
- Grace, A.A. (2016) 'Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression', *Nature reviews. Neuroscience*, 17(8), pp. 524–532.
- GTEx Consortium (2017) 'Genetic effects on gene expression across human tissues', *Nature*, 550(7675), pp. 204–213.
- GTEx Consortium (2013) 'The Genotype-Tissue Expression (GTEx) project', *Nature Genetics*, 45(6), pp. 580–585.
- Guipponi, M., Santoni, F.A., Setola, V., Gehrig, C., Rotharmel, M., et al. (2014) 'Exome sequencing in 53 sporadic cases of schizophrenia identifies 18 putative candidate genes', *PloS One*, 9(11), p. e112745.
- Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., et al. (2013) 'Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network', *Cell*, 154(3), pp. 518–529.
- Halgren, C., Kjaergaard, S., Bak, M., Hansen, C., El-Schich, Z., et al. (2012) 'Corpus callosum abnormalities, intellectual disability, speech impairment, and autism in patients with haploinsufficiency of ARID1B', *Clinical Genetics*, 82(3), pp. 248–255.
- Harper, P.S. (2004) *Genetic Counselling: specific organ systems - disorders of mental function*. In: Arnold E, Hrsg. *Practical Genetic Counselling*. London: Hodder Arnold: .
- Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M.A., Mittal, K., et al. (2018) 'Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families', *Molecular Psychiatry*, 23(4), pp. 973–984.
- Heins, M., Simons, C., Lataster, T., Pfeifer, S., Versmissen, D., et al. (2011) 'Childhood trauma and psychosis: a case-control and case-sibling comparison across different levels of genetic liability, psychopathology, and type of trauma', *The American Journal of Psychiatry*, 168(12), pp. 1286–1294.
- Hiatt, J.B., Pritchard, C.C., Salipante, S.J., O'Roak, B.J. & Shendure, J. (2013) 'Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation', *Genome Research*, 23(5), pp. 843–854.
- Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., et al. (2018) 'Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register', *Biological Psychiatry*, 83(6), pp. 492–498.
- Hjorthøj, C., Stürup, A.E., McGrath, J.J. & Nordentoft, M. (2017) 'Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis', *The Lancet. Psychiatry*, 4(4), pp. 295–301.

- Homann, O.R., Misura, K., Lamas, E., Sandrock, R.W., Nelson, P., et al. (2016) 'Whole-genome sequencing in multiplex families with psychoses reveals mutations in the SHANK2 and SMARCA1 genes segregating with illness', *Molecular Psychiatry*,
- Hornig, T., Grüning, B., Kundu, K., Houwaart, T., Backofen, R., et al. (2017) 'GRIN3B missense mutation as an inherited risk factor for schizophrenia: whole-exome sequencing in a family with a familiar history of psychotic disorders', *Genetics Research*, 99p. e1.
- Howrigan, D.P., Rose, S.A., Samocha, K.E., Fromer, M., Cerrato, F., et al. (2020) 'Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations', *Nature Neuroscience*, 23(2), pp. 185–193.
- Hwang, Y., Kim, J., Shin, J.Y., Kim, J.I., Seo, J.S., et al. (2013) 'Gene expression profiling by mRNA sequencing reveals increased expression of immune/inflammation-related genes in the hippocampus of individuals with schizophrenia', *Translational Psychiatry*, 3p. e321.
- Islam, K.U.S., Meli, N. & Blaess, S. (2021) 'The Development of the Mesoprefrontal Dopaminergic System in Health and Disease', *Frontiers in Neural Circuits*, 15p. 746582.
- Iwamoto, K., Bundo, M. & Kato, T. (2005) 'Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis', *Human Molecular Genetics*, 14(2), pp. 241–253.
- Iwamoto, K., Kakiuchi, C., Bundo, M., Ikeda, K. & Kato, T. (2004) 'Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders', *Molecular Psychiatry*, 9(4), pp. 406–416.
- Jablensky, A., Sartorius, N., Ernberg, G., Anker, M., Korten, A., et al. (1992) 'Schizophrenia: manifestations, incidence and course in different cultures A World Health Organization Ten-Country Study', *Psychological Medicine Monograph Supplement*, 20pp. 1–97.
- John, J., Kukshal, P., Bhatia, T., Chowdari, K.V., Nimgaonkar, V.L., et al. (2017) 'Possible role of rare variants in Trace amine associated receptor 1 in schizophrenia', *Schizophrenia Research*,
- John, J., Kukshal, P., Sharma, A., Bhatia, T., Nimgaonkar, V.L., et al. (2018) 'Rare variants in Protein tyrosine phosphatase, receptor type A (PTPRA) in schizophrenia: Evidence from a family based study', *Schizophrenia Research*,
- John, J., Sharma, A., Kukshal, P., Bhatia, T., Nimgaonkar, V.L., et al. (2019) 'Rare Variants in Tissue Inhibitor of Metalloproteinase 2 as a Risk Factor for Schizophrenia: Evidence From Familial and Cohort Analysis', *Schizophrenia Bulletin*, 45(1), pp. 256–263.
- Joober, R., Rouleau, G.A., Lal, S., Bloom, D., Lalonde, P., et al. (2005) 'Increased prevalence of schizophrenia spectrum disorders in relatives of neuroleptic-nonresponsive schizophrenic patients', *Schizophrenia Research*, 77(1), pp. 35–41.
- Kaur, S. & Christodoulou, J. (1993) 'MECP2 Disorders', in Margaret P. Adam, Holly H. Ardinger, Roberta A. Pagon, Stephanie E. Wallace, Lora JH Bean, et al. (eds.) *GeneReviews*®. [Online]. Seattle (WA): University of Washington, Seattle.
- Khan, F.F., Melton, P.E., McCarthy, N.S., Morar, B., Blangero, J., et al. (2018) 'Whole genome sequencing of 91 multiplex schizophrenia families reveals increased burden of rare, exonic copy number variation in schizophrenia probands and genetic heterogeneity', *Schizophrenia Research*, 197pp. 337–345.

- Kircher, M., Witten, D.M., Jain, P., O’Roak, B.J., Cooper, G.M., et al. (2014) ‘A general framework for estimating the relative pathogenicity of human genetic variants’, *Nature Genetics*, 46(3), pp. 310–315.
- Kirkbride, J.B., Errazuriz, A., Croudace, T.J., Morgan, C., Jackson, D., et al. (2012) ‘Incidence of Schizophrenia and Other Psychoses in England, 1950–2009: A Systematic Review and Meta-Analyses’, *PLoS ONE*, 7(3).
- Klauck, S.M., Lindsay, S., Beyer, K.S., Splitt, M., Burn, J., et al. (2002) ‘A Mutation Hot Spot for Nonspecific X-Linked Mental Retardation in the MECP2 Gene Causes the PPM-X Syndrome’, *American Journal of Human Genetics*, 70(4), pp. 1034–1037.
- Kohen, R., Dobra, A., Tracy, J.H. & Haugen, E. (2014) ‘Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness’, *Translational Psychiatry*, 4p. e366.
- Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., et al. (2012) ‘Rate of de novo mutations, father’s age, and disease risk’, *Nature*, 488(7412), pp. 471–475.
- Kos, M.Z., Carless, M.A., Peralta, J., Curran, J.E., Quillen, E.E., et al. (2017) ‘Exome sequences of multiplex, multigenerational families reveal schizophrenia risk loci with potential implications for neurocognitive performance’, *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 174(8), pp. 817–827.
- Kranz, T.M., Goetz, R.R., Walsh-Messinger, J., Goetz, D., Antonius, D., et al. (2015) ‘Rare variants in the neurotrophin signaling pathway implicated in schizophrenia risk’, *Schizophrenia Research*, 168(1–2), pp. 421–428.
- Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P.J.P., El Mokhtari, N.E., et al. (2006) ‘PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships’, *Community Genetics*, 9(1), pp. 55–61.
- Krumm, N., Sudmant, P.H., Ko, A., O’Roak, B.J., Malig, M., et al. (2012) ‘Copy number variation detection and genotyping from exome sequence data’, *Genome Research*, 22(8), pp. 1525–1532.
- Kummeling, J., Stremmelaar, D.E., Raun, N., Reijnders, M.R.F., Willemsen, M.H., et al. (2021) ‘Characterization of SETD1A haploinsufficiency in humans and Drosophila defines a novel neurodevelopmental syndrome’, *Molecular Psychiatry*, 26(6), pp. 2013–2024.
- Lanz, T.A., Joshi, J.J., Reinhart, V., Johnson, K., Grantham II, L.E., et al. (2015) ‘STEP Levels Are Unchanged in Pre-Frontal Cortex and Associative Striatum in Post-Mortem Human Brain Samples from Subjects with Schizophrenia, Bipolar Disorder and Major Depressive Disorder’, *PLoS ONE*, 10(3).
- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., et al. (2016) ‘Analysis of protein-coding genetic variation in 60,706 humans’, *Nature*, 536(7616), pp. 285–291.
- Leonhard, K. (1999) *Classification of Endogenous Psychoses and their Differentiated Etiology*. Springer Science & Business Media.
- Lewis, G., David, A., Andréasson, S. & Allebeck, P. (1992) ‘Schizophrenia and city life’, *The Lancet*, 340(8812), pp. 137–140.

- Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., et al. (1992) 'Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA', *Cell*, 69(6), pp. 905–914.
- Li, S., DeLisi, L.E. & McDonough, S.I. (2021) 'Rare germline variants in individuals diagnosed with schizophrenia within multiplex families', *Psychiatry Research*, 303p. 114038.
- Liu, D., Meyer, D., Fennessy, B., Feng, C., Cheng, E., et al. (2023) 'Schizophrenia risk conferred by rare protein-truncating variants is conserved across diverse human populations', *Nature Genetics*, 55(3), pp. 369–376.
- Livak, K.J. & Schmittgen, T.D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods (San Diego, Calif.)*, 25(4), pp. 402–408.
- Löhrs, L. & Hasan, A. (2019) '[Risk factors for the development of schizophrenia]', *Fortschritte Der Neurologie-Psychiatrie*, 87(2), pp. 133–143.
- Ludwig, K.U., Degenhardt, F. & Nöthen, M.M. (2019) 'Die Rolle seltener Varianten bei häufigen Krankheiten', *medizinische genetik*, 31(2), pp. 212–221.
- Malaspina D, Harlap S, Fennig S, & et al (2001) 'ADvancing paternal age and the risk of schizophrenia', *Archives of General Psychiatry*, 58(4), pp. 361–367.
- Marotta, R., Risoleo, M.C., Messina, G., Parisi, L., Carotenuto, M., et al. (2020) 'The Neurochemistry of Autism', *Brain Sciences*, 10(3), p. 163.
- Marshall, C.R., Howrigan, D.P., Merico, D., Thiruvahindrapuram, B., Wu, W., et al. (2017) 'Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects', *Nature Genetics*, 49(1), pp. 27–35.
- Maycox, P.R., Kelly, F., Taylor, A., Bates, S., Reid, J., et al. (2009) 'Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function', *Molecular Psychiatry*, 14(12), pp. 1083–1094.
- McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., et al. (2008) 'Genome-wide association studies for complex traits: consensus, uncertainty and challenges', *Nature Reviews. Genetics*, 9(5), pp. 356–369.
- McCarthy, S.E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., et al. (2014) 'De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability', *Molecular Psychiatry*, 19(6), pp. 652–658.
- McDonald-McGinn, D.M., Hain, H.S., Emanuel, B.S. & Zackai, E.H. (1993) '22q11.2 Deletion Syndrome', in Margaret P. Adam, Holly H. Ardinger, Roberta A. Pagon, Stephanie E. Wallace, Lora JH Bean, et al. (eds.) *GeneReviews*®. [Online]. Seattle (WA): University of Washington, Seattle.
- McGrath, J., El-Saadi, O., Cardy, S., Chapple, B., Chant, D., et al. (2001) 'Urban birth and migrant status as risk factors for psychosis: an Australian case-control study', *Social Psychiatry and Psychiatric Epidemiology*, 36(11), pp. 533–536.
- McHugh, M.L. (2013) 'The chi-square test of independence', *Biochemia Medica*, 23(2), pp. 143–149.

- Miller, B., Messias, E., Miettunen, J., Alaräisänen, A., Järvelin, M.-R., et al. (2011) 'Meta-analysis of Paternal Age and Schizophrenia Risk in Male Versus Female Offspring', *Schizophrenia Bulletin*, 37(5), pp. 1039–1047.
- Mishra, A. & Macgregor, S. (2015) 'VEGAS2: Software for More Flexible Gene-Based Testing', *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, 18(1), pp. 86–91.
- Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., et al. (2011) 'Deep sequencing reveals 50 novel genes for recessive cognitive disorders', *Nature*, 478(7367), pp. 57–63.
- Nan, X., Campoy, F.J. & Bird, A. (1997) 'MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin', *Cell*, 88(4), pp. 471–481.
- Narayan, S., Tang, B., Head, S.R., Gilmartin, T.J., Sutcliffe, J.G., et al. (2008) 'Molecular profiles of schizophrenia in the CNS at different stages of illness', *Brain Research*, 1239pp. 235–248.
- Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., et al. (2012) 'Patterns and rates of exonic de novo mutations in autism spectrum disorders', *Nature*, 485(7397), pp. 242–245.
- Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., et al. (2009) 'Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes', *Nature*, 461(7261), pp. 272–276.
- Nicolson, R., Brookner, F.B., Lenane, M., Gochman, P., Ingraham, L.J., et al. (2003) 'Parental schizophrenia spectrum disorders in childhood-onset and adult-onset schizophrenia', *The American Journal of Psychiatry*, 160(3), pp. 490–495.
- Nieto-Estevez, V. & Hsieh, J. (2018) 'CHD2: One Gene, Many Roles', *Neuron*, 100(5), pp. 1014–1016.
- Nordentoft, M., Laursen, T.M., Agerbo, E., Qin, P., Høyer, E.H., et al. (2004) 'Change in suicide rates for patients with schizophrenia in Denmark, 1981-97: nested case-control study', *BMJ : British Medical Journal*, 329(7460), p. 261.
- Nordentoft, M., Wahlbeck, K., Hällgren, J., Westman, J., Ösby, U., et al. (2013) 'Excess Mortality, Causes of Death and Life Expectancy in 270,770 Patients with Recent Onset of Mental Disorders in Denmark, Finland and Sweden', *PLOS ONE*, 8(1), p. e55176.
- Nöthlings, U. & Krawczak, M. (2012) '[PopGen. A population-based biobank with prospective follow-up of a control group]', *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 55(6–7), pp. 831–835.
- O'Brien, N.L., Fiorentino, A., Curtis, D., Rayner, C., Petrosellini, C., et al. (2018) 'Rare variant analysis in multiply affected families, association studies and functional analysis suggest a role for the ITGB4 gene in schizophrenia and bipolar disorder', *Schizophrenia Research*, 199pp. 181–188.
- O'Donovan, M.C. & Owen, M.J. (2016) 'The implications of the shared genetics of psychiatric disorders', *Nature Medicine*, 22(11), pp. 1214–1219.
- O'Roak, B.J., Stessman, H.A., Boyle, E.A., Witherspoon, K.T., Martin, B., et al. (2014) 'Recurrent de novo mutations implicate novel genes underlying simplex autism risk', *Nature Communications*, 5p. 5595.

- O’Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., et al. (2012) ‘Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations’, *Nature*, 485(7397), pp. 246–250.
- Os, J. van, Hanssen, M., Bijl, R.V. & Vollebergh, W. (2001) ‘Prevalence of Psychotic Disorder and Community Level of Psychotic Symptoms: An Urban-Rural Comparison’, *Archives of General Psychiatry*, 58(7), pp. 663–668.
- van Os, J., Kenis, G. & Rutten, B.P.F. (2010) ‘The environment and schizophrenia’, *Nature*, 468(7321), pp. 203–212.
- Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., et al. (2018) ‘Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection’, *Nature Genetics*, 50(3), pp. 381–389.
- Pascual-Alonso, A., Martínez-Monseny, A.F., Xiol, C. & Armstrong, J. (2021) ‘MECP2-Related Disorders in Males’, *International Journal of Molecular Sciences*, 22(17), p. 9610.
- Pauper, M., Kucuk, E., Wenger, A.M., Chakraborty, S., Baybayan, P., et al. (2021) ‘Long-read trio sequencing of individuals with unsolved intellectual disability’, *European Journal of Human Genetics*, 29(4), pp. 637–648.
- Peixoto, L. & Abel, T. (2013) ‘The Role of Histone Acetylation in Memory Formation and Cognitive Impairments’, *Neuropsychopharmacology*, 38(1), pp. 62–76.
- Perälä, J., Suvisaari, J., Saarni, S.I., Kuoppasalmi, K., Isometsä, E., et al. (2007) ‘Lifetime prevalence of psychotic and bipolar I disorders in a general population’, *Archives of General Psychiatry*, 64(1), pp. 19–28.
- Pérez-Piñar, M., Mathur, R., Foguet, Q., Ayis, S., Robson, J., et al. (2016) ‘Cardiovascular risk factors among patients with schizophrenia, bipolar, depressive, anxiety, and personality disorders’, *European Psychiatry*, 35pp. 8–15.
- Pfuhlmann, B., Franzek, E. & Stöber, G. (1998) ‘[Cataphasia: a by formal thought disorders and speech characteristics distinguished psychosis of the schizophrenic domain]’, *Der Nervenarzt*, 69(3), pp. 257–263.
- Pietersen, C.Y., Mauney, S.A., Kim, S.S., Lim, M.P., Rooney, R.J., et al. (2014) ‘Molecular profiles of pyramidal neurons in the superior temporal cortex in schizophrenia’, *Journal of Neurogenetics*, 28(1–2), pp. 53–69.
- Pietersen, C.Y., Mauney, S.A., Kim, S.S., Passeri, E., Lim, M.P., et al. (2014) ‘Molecular Profiles of Parvalbumin-Immunoreactive Neurons in the Superior Temporal Cortex in Schizophrenia’, *Journal of neurogenetics*, 28(0), pp. 70–85.
- Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., et al. (2012) ‘A robust model for read count data in exome sequencing experiments and implications for copy number variant calling’, *Bioinformatics (Oxford, England)*, 28(21), pp. 2747–2754.
- Poisson, A., Chatron, N., Labalme, A., Fournieret, P., Ville, D., et al. (2020) ‘Chromatin remodeling dysfunction extends the etiological spectrum of schizophrenia: a case report’, *BMC medical genetics*, 21(1), p. 10.

- Power, R.A., Kyaga, S., Uher, R., MacCabe, J.H., Långström, N., et al. (2013) 'Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings', *JAMA psychiatry*, 70(1), pp. 22–30.
- Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., et al. (2014) 'A polygenic burden of rare disruptive mutations in schizophrenia', *Nature*, 506(7487), pp. 185–190.
- Ravn, K., Nielsen, J.B., Uldall, P., Hansen, F.J. & Schwartz, M. (2003) 'No correlation between phenotype and genotype in boys with a truncating MECP2 mutation', *Journal of Medical Genetics*, 40(1), pp. e5–e5.
- Raychaudhuri, S. (2011) 'Mapping rare and common causal alleles for complex human diseases', *Cell*, 147(1), pp. 57–69.
- Rees, E., Han, J., Morgan, J., Carrera, N., Escott-Price, V., et al. (2020) 'De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia', *Nature Neuroscience*, 23(2), pp. 179–184.
- Rees, E., Kendall, K., Pardiñas, A.F., Legge, S.E., Pocklington, A., et al. (2016) 'Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia', *JAMA psychiatry*, 73(9), pp. 963–969.
- Rees, E., Walters, J.T.R., Georgieva, L., Isles, A.R., Chambert, K.D., et al. (2014) 'Analysis of copy number variations at 15 schizophrenia-associated loci', *The British Journal of Psychiatry*, 204(2), pp. 108–114.
- Renthal, W., Maze, I., Krishnan, V., Covington, H.E., Xiao, G., et al. (2007) 'Histone deacetylase 5 epigenetically controls behavioral adaptations to chronic emotional stimuli', *Neuron*, 56(3), pp. 517–529.
- Rey, E.-R. (2011) 'Psychotische Störungen und Schizophrenie', in *Klinische Psychologie & Psychotherapie*. 2nd edition [Online]. Wittchen, H.-U., Hoyer, J. (Hrsg.). Berlin Heidelberg New York: Springer.
- Ronan, J.L., Wu, W. & Crabtree, G.R. (2013) 'From neural development to cognition: unexpected roles for chromatin', *Nature reviews. Genetics*, 14(5), pp. 347–359.
- Rössler, W. (2011) 'Epidemiologie der Schizophrenie', *Swiss Medical Forum*, 11(48), pp. 885–888.
- Schaaf, C. & Zschocke, J. (2018) Google-Books-ID: ZYpWDwAAQBAJ. *Basiswissen Humangenetik*. Springer-Verlag.
- Schaefer, M.H., Fontaine, J.-F., Vinayagam, A., Porras, P., Wanker, E.E., et al. (2012) 'HIPPIE: Integrating Protein Interaction Networks with Experiment Based Quality Scores', *PLOS ONE*, 7(2), p. e31826.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) 'Biological insights from 108 schizophrenia-associated genetic loci', *Nature*, 511(7510), p. 421.
- Senormanci, O., Karakas Celik, S., Valipour, E., Dogan, V. & Senormanci, G. (2018) 'Determination of candidate genes involved in schizophrenia using the whole-exome sequencing', *Bratislavske Lekarske Listy*, 119(9), pp. 572–576.

- Shahbazian, M.D. & Grunstein, M. (2007) 'Functions of site-specific histone acetylation and deacetylation', *Annual Review of Biochemistry*, 76pp. 75–100.
- Shang, B., Yang, R., Lian, K., Dong, L., Liu, H., et al. (2024) 'Family-based genetic analysis in schizophrenia by whole-exome sequence to identify rare pathogenic variants', *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 195(5), p. e32968.
- Shang, L., Cho, M.T., Retterer, K., Folk, L., Humberson, J., et al. (2015) 'Mutations in ARID2 are associated with intellectual disabilities', *Neurogenetics*, 16(4), pp. 307–314.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., et al. (2003) 'Cytoscape: a software environment for integrated models of biomolecular interaction networks', *Genome Research*, 13(11), pp. 2498–2504.
- Shao, Z., Noh, H., Kim, W.B., Ni, P., Nguyen, C., et al. (2019) 'Dysregulated protocadherin-pathway activity as an intrinsic defect in iPSC-derived cortical interneurons from patients with schizophrenia', *Nature neuroscience*, 22(2), pp. 229–242.
- Singh, T., Kurki, M.I., Curtis, D., Purcell, S.M., Crooks, L., et al. (2016) 'Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders', *Nature Neuroscience*, 19(4), pp. 571–577.
- Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., et al. (2022) 'Rare coding variants in ten genes confer substantial risk for schizophrenia', *Nature*, 604(7906), pp. 509–516.
- Smedemark-Margulies, N., Brownstein, C.A., Vargas, S., Tembulkar, S.K., Towne, M.C., et al. (2016) 'A novel de novo mutation in ATP1A3 and childhood-onset schizophrenia', *Cold Spring Harbor Molecular Case Studies*, 2(5), p. a001008.
- Sokpor, G., Xie, Y., Rosenbusch, J. & Tuoc, T. (2017) 'Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders', *Frontiers in Molecular Neuroscience*, 10.
- Sonnenschein, S.F., Gomes, F.V. & Grace, A.A. (2020) 'Dysregulation of Midbrain Dopamine System and the Pathophysiology of Schizophrenia', *Frontiers in Psychiatry*, 11p. 613.
- Stöber, G., Saar, K., Rüschemdorf, F., Meyer, J., Nürnberg, G., et al. (2000) 'Splitting Schizophrenia: Periodic Catatonia–Susceptibility Locus on Chromosome 15q15', *American Journal of Human Genetics*, 67(5), pp. 1201–1207.
- Suddaby, J.S., Silver, J. & So, J. (2019) 'Understanding the schizophrenia phenotype in the first patient with the full SCN2A phenotypic spectrum', *Psychiatric Genetics*,
- Sullivan, P.F., Daly, M.J. & O'Donovan, M. (2012) 'Genetic Architectures of Psychiatric Disorders: The Emerging Picture and Its Implications', *Nature reviews. Genetics*, 13(8), pp. 537–551.
- Szatkiewicz, J.P., Fromer, M., Nonneman, R.J., Ancalade, N., Johnson, J.S., et al. (2020) 'Characterization of Single Gene Copy Number Variants in Schizophrenia', *Biological psychiatry*, 87(8).
- Takata, A., Xu, B., Ionita-Laza, I., Roos, J.L., Gogos, J.A., et al. (2014) 'Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene', *Neuron*, 82(4), pp. 773–780.

- Tan, R., Wang, Y., Kleinstein, S.E., Liu, Y., Zhu, X., et al. (2014) 'An evaluation of copy number variation detection tools from whole-exome sequencing data', *Human Mutation*, 35(7), pp. 899–907.
- Tang, J., Fan, Y., Li, H., Xiang, Q., Zhang, D.-F., et al. (2017) 'Whole-genome sequencing of monozygotic twins discordant for schizophrenia indicates multiple genetic risk factors for schizophrenia', *Journal of Genetics and Genomics*, 44(6), pp. 295–306.
- Tansey, K.E., Rees, E., Linden, D.E., Ripke, S., Chambert, K.D., et al. (2016) 'Common alleles contribute to schizophrenia in CNV carriers', *Molecular Psychiatry*, 21(8), p. 1153.
- The International Schizophrenia Consortium (2008) 'Rare chromosomal deletions and duplications increase risk of schizophrenia', *Nature*, 455(7210), pp. 237–241.
- Tian, L., Al-Nusaif, M., Chen, X., Li, S. & Le, W. (2022) 'Roles of Transcription Factors in the Development and Reprogramming of the Dopaminergic Neurons', *International Journal of Molecular Sciences*, 23(2), p. 845.
- Tiihonen, J., Lönngqvist, J., Wahlbeck, K., Klaukka, T., Niskanen, L., et al. (2009) '11-year follow-up of mortality in patients with schizophrenia: a population-based cohort study (FIN11 study)', *Lancet (London, England)*, 374(9690), pp. 620–627.
- Timms, A.E., Dorschner, M.O., Wechsler, J., Choi, K.Y., Kirkwood, R., et al. (2013) 'Support for the N-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families', *JAMA psychiatry*, 70(6), pp. 582–590.
- Toyama, M., Takasaki, Y., Branko, A., Kimura, H., Kato, H., et al. (2022) 'Exome sequencing of Japanese schizophrenia multiplex families supports the involvement of calcium ion channels', *PLOS ONE*, 17(5), p. e0268321.
- Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., et al. (2022) 'Mapping genomic loci implicates genes and synaptic biology in schizophrenia', *Nature*, 604(7906), pp. 502–508.
- Turner, T.N., Wilfert, A.B., Bakken, T.E., Bernier, R.A., Pepper, M.R., et al. (2019) 'Sex-Based Analysis of De Novo Variants in Neurodevelopmental Disorders', *American Journal of Human Genetics*, 105(6), pp. 1274–1285.
- Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H.-H., et al. (2011) 'Duplications of the Neuropeptide Receptor VIPR2 Confer Significant Risk for Schizophrenia', *Nature*, 471(7339), pp. 499–503.
- Valencia, A.M., Sankar, A., van der Sluijs, P.J., Satterstrom, F.K., Fu, J., et al. (2023) 'Landscape of mSWI/SNF chromatin remodeling complex perturbations in neurodevelopmental disorders', *Nature Genetics*, 55(8), pp. 1400–1412.
- Vaucher, J., Keating, B.J., Lasserre, A.M., Gan, W., Lyall, D.M., et al. (2018) 'Cannabis use and risk of schizophrenia: a Mendelian randomization study', *Molecular Psychiatry*, 23(5), pp. 1287–1292.
- Wang, Q., Li, Miaoxin, Yang, Z., Hu, X., Wu, H.-M., et al. (2015) 'Increased co-expression of genes harboring the damaging de novo mutations in Chinese schizophrenic patients during prenatal development', *Scientific Reports*, 5p. 18209.

- Wang, Y., Yu, Y., Pang, Y., Yu, H., Zhang, W., et al. (2021) 'The distinct roles of zinc finger CCHC-type (ZCCHC) superfamily proteins in the regulation of RNA metabolism', *RNA biology*, 18(12), pp. 2107–2126.
- Watanabe, Y., Nunokawa, A., Shibuya, M., Ikeda, M., Hishimoto, A., et al. (2016) 'Rare truncating variations and risk of schizophrenia: Whole-exome sequencing in three families with affected siblings and a three-stage follow-up study in a Japanese population', *Psychiatry Research*, 235pp. 13–18.
- van der Werf, M., Hanssen, M., Köhler, S., Verkaaik, M., Verhey, F.R., et al. (2014) 'Systematic review and collaborative recalculation of 133,693 incident cases of schizophrenia', *Psychological Medicine*, 44(1), pp. 9–16.
- Wither, R.G., Borlot, F., MacDonald, A., Butcher, N.J., Chow, E.W.C., et al. (2017) '22q11.2 deletion syndrome lowers seizure threshold in adult patients without epilepsy', *Epilepsia*, 58(6), pp. 1095–1101.
- World Medical Association (2013) 'World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects', *JAMA*, 310pp. 2191–2194.
- Wray, N.R. & Gottesman, I.I. (2012) 'Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder', *Frontiers in Genetics*, 3p. 118.
- Wu, J.Q., Wang, X., Beveridge, N.J., Tooney, P.A., Scott, R.J., et al. (2012) 'Transcriptome sequencing revealed significant alteration of cortical promoter usage and splicing in schizophrenia', *PloS One*, 7(4), p. e36351.
- Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., et al. (2012) 'De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia', *Nature Genetics*, 44(12), pp. 1365–1369.
- Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., et al. (2011) 'Exome sequencing supports a de novo mutational paradigm for schizophrenia', *Nature Genetics*, 43(9), pp. 864–868.
- Xue, C.-B., Xu, Z.-H., Zhu, J., Wu, Y., Zhuang, X.-H., et al. (2018) 'Exome Sequencing Identifies TENM4 as a Novel Candidate Gene for Schizophrenia in the SCZD2 Locus at 11q14-21', *Frontiers in Genetics*, 9p. 725.
- Yu, D.X., Di Giorgio, F.P., Yao, J., Marchetto, M.C., Brennand, K., et al. (2014) 'Modeling Hippocampal Neurogenesis Using Human Pluripotent Stem Cells', *Stem Cell Reports*, 2(3), pp. 295–310.
- Zaret, K.S. & Carroll, J.S. (2011) 'Pioneer transcription factors: establishing competence for gene expression', *Genes & Development*, 25(21), pp. 2227–2241.
- Zhao, Z., Xu, J., Chen, J., Kim, S., Reimers, M., et al. (2015) 'Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder', *Molecular Psychiatry*, 20(5), pp. 563–572.
- Zhou, X., Feliciano, P., Shu, C., Wang, T., Astrovskaya, I., et al. (2022) 'Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes', *Nature Genetics*, 54(9), pp. 1305–1319.

Zhou, Z., Hu, Zhengmao, Zhang, L., Hu, Zhaoting, Liu, H., et al. (2016) 'Identification of RELN variation p.Thr3192Ser in a Chinese family with schizophrenia', *Scientific Reports*, 6p. 24327.

8 Anhang

Tabelle A1 Verwendete Einstellungen und Filter zur Qualitätskontrolle in Varbank

Filterkriterium	Einstellung
Allele read frequency range	25-100%
Minimal read coverage	6
Minimal variation quality	10
GATK best practice (minimal VQSLOD)	-8
Maximal number observed in inhouse database (CCG-inhouse database)	2
Maximal target distance	100
Transcript biotype	protein coding
Variation location	CDS; UTR; intron; spanning genes; promoter; composed
Variation type	SNP; insertion; deletion; indel
Variation consequences	protein structure affected; strong 5'SS/3'SS effect; medium 5'SS/3'SS effect

Dargestellt sind die verwendeten Einstellungen und Filter zur Qualitätskontrolle in der Varbank Pipeline. GATK: *Genome Analysis Toolkit*; VQSLOD: *variant quality score log-odds*; CDS: *coding sequence*; UTR: *untranslated region*; SNP: *single nucleotide polymorphism*; SS: *splice site*.

Tabelle A2 Verwendete Einstellungen und Filter zur Qualitätskontrolle in der CLC bio Biomedical Genomics Workbench

Filterkriterium	Einstellung
“Prepare raw data”	
Ambiguous limit	2
Quality limit	0.01
Maximum number of nucleotides in reads	1 000
Minimum number of nucleotides in reads	15
“Identify and Annotate Variants (WES)”	
Significance (%)	1.0
Minimum coverage	10
Minimum count per strand	5
Minimum frequency (%)	25
Base quality filter	on

Dargestellt sind die verwendeten Einstellungen und Filter zur Qualitätskontrolle in der CLC bio Biomedical Genomics Workbench. WES: *Whole Exome Sequencing*.

Tabelle A3 Allelfrequenzen der 35 identifizierten SNVs und Indels der Pilotstudie

Gen	Mutation	MAF ExAC (%)	145 deutsche Kontroll-Exome	MAF deutsche Kontrollen in % (Genotypisierungsdaten)
<i>ADAMTS19</i>	c.2270C>T	-	-	-
<i>ANKLE1</i>	c.1244C>T	-	-	x
<i>ATG9B</i>	c.2490G>A	-	-	-
<i>BAIAP2L1</i>	c.1435G>A	-	-	-
<i>C17orf80</i>	c.424G>C	0.019	-	x
<i>C9orf64</i>	c.158A>C	-	-	-
<i>CYP27C1</i>	c.349C>T	0.007	-	-
<i>FAM185A</i>	c.442G>A	-	-	x
<i>FOXA1</i>	c.890C>T	-	-	x
<i>FRG2C</i>	c.463_464delinsA	*	-	x
<i>GBAS</i>	c.267C>A	0.002	-	-
<i>GZF1</i>	c.1880A>G	-	-	-
<i>HDAC5</i>	c.1225C>T	0.003	-	x
<i>KIAA0226</i>	c.1916G>A	0.010	-	-
<i>KRBA1</i>	c.1924G>T	-	-	x
<i>KRT79</i>	c.386T>C	-	-	0.046
<i>L3MBTL1</i>	c.824C>T	0.014	-	-
<i>LRP5L</i>	c.380C>T	0.007	-	-
<i>MECP2</i>	c.250C>T	-	-	x
<i>NOS3</i>	c.1465G>A	0.002	-	x
<i>OR2B11</i>	c.640C>T	0.033	-	x
<i>PARP6</i>	c.244G>A	-	-	-
<i>PASD1</i>	c.2209_2209delC	-	-	x
<i>PDE7B</i>	c.52G>A	0.005	-	-
<i>RGS17</i>	c.259A>G	0.078	-	x
<i>RMI1</i>	c.980_997delAGAAAGAACAGATGGAAA	0.012	-	-
<i>RYR2</i>	c.3380A>G	0.088	-	x
<i>SRRM2</i>	c.2951C>T	0.007	-	-
<i>TET1</i>	c.1022C>T	0.067	-	x
<i>TEX11</i>	c.2659G>A	0.024	-	-
<i>TGDS</i>	c.229_232del(AT)2insAT	-	-	-
<i>TMEM74B</i>	c.286C>T	0.010	-	x
<i>TXNDC16</i>	c.373A>G	-	-	-
<i>ZIC5</i>	c.1306G>C	-	-	-
<i>ZNF426</i>	c.1498G>A	0.005	-	x

Dargestellt sind die Allelfrequenzen der identifizierten Varianten in folgenden Kohorten: (i) Nicht-finnische Europäer in der nicht-psychiatrischen Kohorte des Exome Aggregation Consortium [Spalte 3; <http://exac.broadinstitute.org>, ExAC.r0.3.nonpsych.sites.vcf.gz (~ 45.000 Individuen)](Lek et al., 2016); (ii) Exome von 145 nicht-psychiatrischen deutschen Kontrollen (Spalte 4); und (iii) Genotypisierungsdaten von ~ 1.000 deutschen nicht-psychiatrischen Kontrollen (Blutspender; Spalte 5); - nicht gelistet; * *FRG2C* war in ExAC nicht abgedeckt; x aus technischen Gründen keine Daten vorhanden. Die kodierende Position basiert auf dem längsten proteinkodierenden Ensembl-Transkript des jeweiligen Gens (GRCh37).

Tabelle A4 *In silico* Vorhersagen der 35 identifizierten SNVs und Indels der Pilotstudie

Gen	Mutation	CADD score
GBAS	c.267C>A	40.0
TMEM74B	c.286C>T	36.0
TGDS	c.229_232del(AT)2insAT	35.0
KIAA0226	c.1916G>A	35.0
MECP2	c.250C>T	34.0
HDAC5	c.1225C>T	34.0
ATG9B	c.2490G>A	34.0
L3MBTL1	c.824C>T	33.0
RYR2	c.3380A>G	31.0
FAM185A	c.442G>A	31.0
CYP27C1	c.349C>T	29.0
BAIAP2L1	c.1435G>A	28.3
TET1	c.1022C>T	26.7
C9orf64	c.158A>C	26.0
ZNF426	c.1498G>A	25.8
LRP5L	c.380C>T	25.8
PDE7B	c.52G>A	25.5
GZF1	c.1880A>G	25.1
ANKLE1	c.1244C>T	24.5
ZIC5	c.1306G>C	24.4
PASD1	c.2209_2209delC	24.3
C17orf80	c.424G>C	23.4
KRT79	c.386T>C	23.3
OR2B11	c.640C>T	23.2
ADAMTS19	c.2270C>T	22.8
RGS17	c.259A>G	22.8
TEX11	c.2659G>A	22.4
TXNDC16	c.373A>G	21.6
SRRM2	c.2951C>T	21.2
NOS3	c.1465G>A	19.3
FRG2C	c.463_464delinsA	18.2
PARP6	c.244G>A	16.8
RMI1	c.980_997delAGAAAGAACAGATGGAAA	16.2
FOXA1	c.890C>T	15.2
KRBA1	c.1924G>T	15.0

Zur Einschätzung der Pathogenität der Varianten wurde der Combined Annotation Dependent Depletion Score (CADD Score; ≥ 15) (Kircher et al., 2014) v1.3 verwendet.

Tabelle A5 Genexpression der 39 Kandidatengene der Pilotstudie in verschiedenen Gehirngeweben

Gen	Mutation	FPKM in allen Hirnregionen (Durchschnitt) GTEx	FPKM Frontaler Cortex GTEx	FPKM Hippocampus GTEx
ADAMTS19	c.2270C>T	0.5	0.5	0.3
ANKLE1	c.1244C>T	1.3	0.6	0.5
ATG9B	c.2490G>A	1.0	0.8	0.8
BAIAP2L1	c.1435G>A	0.4	0.4	0.5
C17orf80	c.424G>C	5.1	4.0	4.0
C9orf64	c.158A>C	2.1	2.0	2.0
CYP27C1	c.349C>T	0.1	0.2	0.1
FAM185A	c.442G>A	1.6	2.0	1.0
FOXA1	c.890C>T	0.1	0.0	0.0
FRG2C	c.463_464delinsA	0.1	0.0	0.0
GBAS	c.267C>A	11.6	11.0	10.0
GZF1	c.1880A>G	5.4	6.0	4.0
HDAC5	c.1225C>T	17.9	17.0	15.0
KIAA0226	c.1916G>A	8.6	8.0	7.0
KRBA1	c.1924G>T	3.9	4.0	3.0
KRT79	c.386T>C	-	-	-
L3MBTL1	c.824C>T	4.8	3.0	3.0
LRP5L	c.380C>T	0.5	0.5	0.6
MECP2	c.250C>T	8.3	7.0	6.0
NOS3	c.1465G>A	1.3	1.0	1.0
OR2B11	c.640C>T	-	-	-
PARP6	c.244G>A	12.0	11.0	10.0
PASD1	c.2209_2209delC	-	-	-
PDE7B	c.52G>A	2.7	4.0	2.0
RGS17	c.259A>G	1.5	1.0	1.0
RMI1	c.980_997delAGAAAGAACAGATGGAAA	2.1	2.0	1.0
RYR2	c.3380A>G	5.0	5.0	2.0
SRRM2	c.2951C>T	48.9	31.0	28.0
TET1	c.1022C>T	0.3	0.2	0.2
TEX11	c.2659G>A	-	-	-
TGDS	c.229_232del(AT)2insAT	2.3	2.0	2.0
TMEM74B	c.286C>T	5.9	7.0	4.0
TXNDC16	c.373A>G	4.0	4.0	3.0
ZIC5	c.1306G>C	5.4	0.8	1.0
ZNF426	c.1498G>A	1.2	0.8	0.8
DHX30	Duplikation	11.8	10.0	8.0
MRPS28	Duplikation	4.1	5.0	4.0
SMARCC1	Duplikation	4.9	4.0	5.0
TPD52	Duplikation	12.6	19.0	13.0

Gelistet sind die FPKM Werte der Genotype-Tissue Expression (GTEx) Datenbank (GTEx Consortium, 2013) für (i) alle Hirnregionen (Durchschnitt); (ii) Frontaler Cortex; und (iii) Hippocampus. FPKM Fragmente pro Kilobase des Transkripts pro 1 Million Reads. – keine Daten vorhanden.

Tabelle A6 Differenzielle Expression der 39 Kandidatengene der Pilotstudie bei Patienten mit SCZ und Kontrollen

Gen	GSE 12654	GSE 17612	GSE 12649	GSE 53987_HIPPO	GSE 53987_PFC	GSE 62191	PMID_ 22869038	GSE 35978	GSE 46509	GSE 21138	GSE 37981	PMID_ 25113377	PMID_ 24744400	PMID_ 24594777	PMID_ 24169640	PMID_ 22558445	Hausinterner Datensatz
ADAMTS19	-	1.65	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ANKLE1	-	-	-	1.10	-	-	-	-	-	-	-	-	-	-	-	-	-
ATG9B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BAIAP2L1	-	-	-	1.16	-	-	-	-	-	-	-	-	-	-	-	-	-
C17orf80	-	-	-	-	-	-	-	-	0.42	-	-	-	-	-	-	-	-
C9orf64	-	-	-	1.33	1.14	-	-	-	-	1.08	-	-	-	-	-	-	-
CYP27C1	-	-	-	-	-	-	-	-	0.22	-	-	-	-	-	-	-	-
FAM185A	-	-	-	1.20	1.12	-	-	-	-	-	-	-	-	-	-	-	-
FOXA1	-	-	-	-	-	-	-	-	4.67	-	-	-	-	-	-	-	1.82
FRG2C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GBAS	-	-	-	0.76	-	-	-	-	-	-	-	-	-	-	-	-	-
GZF1	-	1.07	-	-	-	-	-	-	-	-	-	-	-	-	-	0.87	-
HDAC5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KIAA0226	-	-	-	-	-	-	-	-	-	0.94	-	-	-	-	-	-	-
KRBA1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KRT79	-	-	-	1.08	-	-	-	-	-	-	-	-	-	-	-	-	-
L3MBTL1	-	-	1.23	1.11	1.12	-	-	-	0.20	-	-	-	-	-	-	-	1.9
LRP5L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MECP2	-	-	-	1.08	-	-	-	-	-	-	-	-	-	-	-	-	-
NOS3	-	-	-	-	-	0.89	-	-	-	-	-	-	-	-	-	-	-
OR2B11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PARP6	-	-	-	0.88	0.92	-	-	-	-	-	-	-	-	-	-	-	0.63
PASD1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PDE7B	-	0.91	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RGS17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RMI1	-	-	1.14	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RYR2	-	-	1.36	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SRRM2	-	-	-	1.16	1.09	-	-	-	-	-	-	-	-	-	-	-	-
TET1	-	-	-	1.23	1.26	1.10	-	-	-	-	-	-	-	-	-	-	-
TEX11	-	-	-	1.11	-	-	-	-	-	-	-	-	-	-	-	-	-
TGDS	2.53	-	-	-	0.93	-	-	-	4.39	-	0.27	-	-	-	-	-	-
TMEM74B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TXNDC16	-	-	-	-	-	1.04	-	-	-	-	-	-	-	-	-	-	-
ZIC5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ZNF426	-	-	-	1.12	-	-	-	-	0.33	-	-	-	-	-	-	-	-
SMARCC1	-	-	-	1.20	1.07	-	-	-	-	-	-	-	-	-	-	-	-
DHX30	-	-	-	1.14	0.94	-	-	-	-	-	-	-	-	-	-	-	-
MRPS28	-	-	-	0.76	-	-	-	-	-	-	-	-	-	-	-	-	-
TPD52	-	-	-	0.73	1.08	-	-	-	-	-	-	-	-	-	-	-	-

Die Tabelle zeigt die differenzielle Genexpression (*fold change*) bei Patienten mit SCZ und gesunden Kontrollen in 15 verschiedenen Studien mit insgesamt 746 Probanden (Catts et al., 2014; Chen et al., 2013; de Baumont et al., 2015; Fillman et al., 2013; Hwang et al., 2013; Iwamoto et al., 2004, 2005; Kohen et al., 2014; Lanz et al., 2015; Maycox et al., 2009; Narayan et al., 2008; Pietersen, Mauney, Kim, Lim, et al., 2014; Pietersen, Mauney, Kim, Passeri, et al., 2014; Wu et al., 2012; Zhao et al., 2015) und einem hausinternen Datensatz von neun Patienten mit SCZ und neun Kontrollen. In einer Studie (GSE53987; Lanz et al., 2015) wurden zwei verschiedene Hirnregionen (HIPPO; Hippocampus und frontaler Cortex; PFC) analysiert. *Fold change* > 1: Die Expression war bei den Patienten signifikant höher als bei den Kontrollprobanden. *Fold change* < 1: Die Expression war bei den Patienten signifikant niedriger als bei den Kontrollprobanden. – keine signifikant differenzielle Genexpression. Weitere Informationen zu den 15 Studien sind in Tabelle A7 dargestellt.

Tabelle A7 Informationen über die Studien zur differentiellen Genexpression

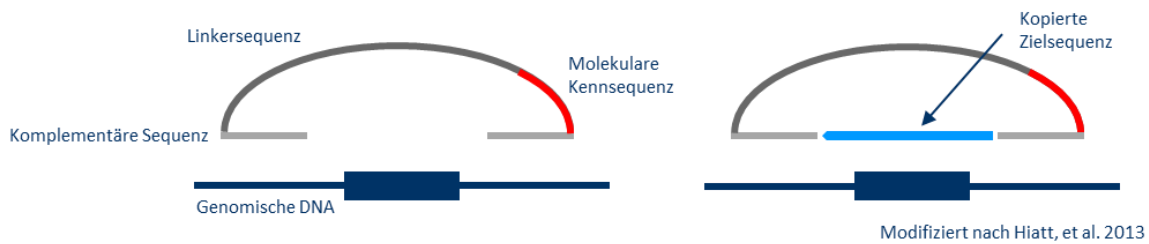
NCBI GEO/PubMed ID	Untersuchtes Gewebe	Kontrollen	Patienten	Methode	Studie
GSE12654	Gehirn (Präfrontaler Cortex)	15	13	Microarray	Iwamoto et al., 2004
GSE17612	Gehirn (BA 10)	23	28	Microarray	Maycox et al., 2009
GSE12649	Gehirn (BA 46)	34	35	Microarray	Iwamoto et al., 2005
GSE53987_HIPPO	Gehirn (Hippocampus)	19	19	Microarray	Lanz et al., 2015
GSE53987_PFC	Gehirn (Präfrontaler Cortex)	19	19	Microarray	Lanz et al., 2015
GSE62191	Gehirn (Frontaler Cortex)	30	29	Microarray	de Baumont et al., 2015
PMID_22869038	Gehirn (dorso lateral PFC)	20	20	RNA-Seq	Fillman et al., 2013
GSE35978	Kleinhirn	50	50	Microarray	Chen et al., 2013
GSE46509	Gehirn (BA 42)	8	8	Microarray	Pietersen et al., 2014a
GSE21138	Gehirn (BA 46)	29	30	Microarray	Narayan et al., 2008
GSE37981	Pyramidenzellen	9	9	Microarray	Pietersen et al., 2014b
PMID_25113377	Gehirn	35	35	RNA-Seq	Zhao et al., 2015
PMID_24744400	Gehirn	20	20	RNA-Seq	Catts et al., 2014
PMID_24594777	Gehirn	15	15	RNA-Seq	Kohen et al., 2014
PMID_24169640	Gehirn	15	14	RNA-Seq	Hwang et al., 2013
PMID_22558445	Gehirn	9	9	RNA-Seq	Wu et al., 2012
	Gehirn (BA 9)	9	9	qPCR	Hausinterner Datensatz

In der Tabelle sind Informationen über die Studien zur differentiellen Genexpression bei Patienten mit Schizophrenie und gesunden Kontrollen gelistet. RNA-Seq: RNA-Sequenzierung; BA: Brodmann Areal; PFC: Präfrontaler Cortex. qPCR: quantitative Polymerase-Kettenreaktion.

Tabelle A8 Ausgewählte Gene für die smMIP Sequenzierung

Gen	Studie
<i>AKAP6</i>	Pardiñas et al., 2018
<i>ATP2A2</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>CACNA1C</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>CACNA1I</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>CLCN3</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>CNNM2</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>CSMD1</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>CUL3</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>DGKI</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>DOPEY1</i>	Pardiñas et al., 2018
<i>FOXP1</i>	Pardiñas et al., 2018
<i>GABBR2</i>	Pardiñas et al., 2018
<i>GRIN2A</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>GRM3</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>HCN1</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>HDAC5</i>	Vorliegende Pilotstudie
<i>IGSF9B</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>KDM3B</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>MAN2A1</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>MECP2</i>	Vorliegende Pilotstudie
<i>MIR137</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>MIR548AJ2</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>NLGN4X</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>PDE4B</i>	Pardiñas et al., 2018
<i>PLCL2</i>	Pardiñas et al., 2018
<i>PTPRK</i>	Pardiñas et al., 2018
<i>RBFOX1</i>	Pardiñas et al., 2018
<i>REER</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>RPTOR</i>	Pardiñas et al., 2018
<i>SATB2</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>SEMA6D</i>	Pardiñas et al., 2018
<i>SMARCC1</i>	Vorliegende Pilotstudie
<i>STAG1</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>TCF20</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014
<i>WHSC1L1</i>	Pardiñas et al., 2018
<i>ZNF536</i>	Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014

Gelistet sind die 36 Kandidatengene aus der Literatur sowie der vorliegenden Familien-Pilotstudie, welche für die Sequenzierung mittels smMIPs in 1.947 Patienten mit Schizophrenie und 1.921 Kontrollen ausgewählt wurden.

**Abbildung A1 Schematische Darstellung der smMIP Methode**

smMIPs sind ~ 80 Basen lange einzelsträngige DNA Moleküle. Diese setzen sich zusammen aus zwei zur Zielsequenz komplementären Sequenzen („Arme“), welche durch eine 30 Basen lange Linkersequenz und eine aus fünf zufälligen Basen zusammengesetzte molekulare Kennsequenz verbunden sind. Durch die molekulare Kennsequenz werden Sequenzen, die von ein und demselben DNA Molekül stammen, markiert. Nach der Sequenzierung wird aus diesen eine Konsensussequenz erstellt, wodurch die Identifizierung falsch positiver Varianten aufgrund technischer Artefakte vermieden werden kann (Hiatt et al., 2013).