

Identifikation von Kopienzahlvarianten und deren Assoziation mit neuropsychiatrischen Erkrankungen

Dissertation

zur Erlangung des Doktorgrades (Doctor rerum naturalium)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Lutz Priebe

aus

Troisdorf

Bonn, Februar 2014

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

1. Gutachter: Prof. Dr. rer. nat. Sven Cichon

2. Gutachter: Prof. Dr. med. Joachim Schultze

Tag der Promotion: 31.10.2014

Erscheinungsjahr: 2014

Abkürzungsverzeichnis

95-KI	95%-Konfidenzintervall
°C	Grad Celsius
A	Adenin
Abb.	Abbildung
aCGH	<i>array comparative genomic hybridization</i>
BAF	<i>B allele frequency</i>
bp	Basenpaare
bzw.	beziehungsweise
C	Cytosin
ca.	circa
CGH	<i>comparative genomic hybridization</i>
Chr	Chromosom
CHOP	<i>Children's Hospital of Philadelphia</i>
CNV	<i>copy number variant</i> , Kopienzahlvariante
Ct	<i>cycle threshold</i>
CV	<i>confidence value</i>
Δ	Delta
ddATP	Didesoxyadenosintriphosphat
ddCTP	Didesoxycytosintriphosphat
ddGTP	Didesoxyguanintriphosphat
ddNTP	Didesoxynukleosidtriphosphat
ddUTP	Didesoxyuraciltriphosphat
DECIPHER	<i>Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources</i>
DGV	<i>Database of Genomic Variants</i>
d.h.	das heißt
dNTP	Desoxynukleosidtriphosphat
DNA	<i>deoxyribonucleic acid</i> , Desoxyribonukleinsäure
DNP	2,4-dinitrophenol
DSM-IV	<i>Diagnostic and Statistical Manual of Mental Disorders fourth edition</i>
et al.	<i>et alii</i> , und andere
FoSTeS	<i>Fork Stalling and Template Switching</i>
G	Guanin
g	Gramm
GAD	<i>Genetic Association Database</i>
H610Q	Human610-Quadv1 Chip
H660W	Human660-Quadv1 Chip
HCl	Salzsäure
HH550	HumanHap550v3 Chip
HMM	Hidden Markov Model
Kap.	Kapitel
kb	Kilobasenpaare
IBF	<i>log Bayes factor</i>
log	Logarithmus
LRR	<i>log R ratio</i>
Mb	Megabasenpaare
MgCl ₂	Magnesiumchlorid
MH	Mikrohologie

Min.	Minute
µl	Mikroliter
ml	Milliliter
MLPA	<i>multiplex ligation-dependent probe amplification</i>
NAD	Nicotinamidadeninukleotid
NAHR	<i>nonallelic homologous recombination</i> , nicht allelische homologe Rekombination
NCBI	<i>National Center for Biotechnology Information</i>
ng	Nanogramm
NHEJ	<i>nonhomologous DNA end joining</i> , nicht homologe Reparatur von Doppelstrangbrüchen
nm	Nanometer
nmol	Nanomol
OR	<i>odds ratio</i>
OR-CI	95%-Konfidenzintervall des <i>odds ratios</i>
ORF	<i>open reading frame</i> , offener Leserahmen
PCR	<i>polymerase chain reaction</i> , Polymerasekettenreaktion
PGC	<i>Psychiatric Genomics Consortium</i>
pmol	Pikomol
qRT-PCR	<i>quantitative real-time PCR</i> , quantitative Echtzeit-PCR
RefSeq	<i>NCBI reference sequence</i>
R	normalisierter Intensitätswert
SA	Standardabweichung
SBE	<i>single base extension</i> , Einzelbasenerweiterung
SD	segmentale Duplikation
Sek.	Sekunde
SLEP	<i>Sullivan Lab Evidence Project</i>
SNP	<i>single nucleotide polymorphism</i> , Einzelnukleotidpolymorphismus
T	Tyrosin
Tab.	Tabelle
Tris	Tris(hydroxymethyl)-aminomethan
U	<i>active unit</i> , aktive Einheit
u.a.	unter anderem
UCSC	<i>University of California in Santa Cruz</i>
WTCCC	<i>Wellcome Trust Case Control Consortium</i>
z.B.	zum Beispiel

Inhaltsverzeichnis

Abkürzungsverzeichnis	I
Inhaltsverzeichnis	III
1 Einleitung	1
1.1 Genetische Variabilität	1
1.2 Strukturelle Variabilität	2
1.3 Kopienzahlvarianten	2
1.3.1 Mikrodeletions-/Mikroduplikationssyndrome.....	4
1.3.2 Genomweite, populationsbasierte CNV Analysen	4
1.3.3 Katalogisierung von CNVs in der <i>Database of Genomic Variants</i>	6
1.3.4 Entstehungsmechanismen von CNVs	6
1.3.5 Effekt von CNVs auf die Genexpression	10
1.3.6 Genomweite Detektionsmethoden von CNVs	12
1.4 Neuropsychiatrische Erkrankungen	13
1.4.1 CNVs bei neuropsychiatrischen Erkrankungen	15
2 Zielsetzung der Arbeit	17
3 Material und Methoden	18
3.1 Verwendete Materialien	18
3.1.1 Verwendete Geräte.....	18
3.1.2 Chemikalien und Lösungen.....	19
3.1.3 Kommerzielle Systeme (Kits)	20
3.1.4 Software und Datenbanken.....	21
3.2 Molekularbiologische Methoden	22
3.2.1 Probandenmaterial und DNA-Isolierung.....	22
3.2.2 Qualitäts- und Konzentrationsbestimmung der DNA	22
3.2.3 Genomweite Genotypisierung mit Illumina	22
3.2.4 Verifikation von <i>in silico</i> CNVs	27
3.2.5 <i>Multiplex ligation-dependent probe amplification</i>	27
3.2.6 <i>Quantitative real-time PCR</i> mit TaqMan-Sonden	29
3.2.7 CopyCaller	32
3.2.8 <i>Quantitative real-time PCR</i> mit SYBR-Green	33
3.2.9 Auswahl der Verifikationsmethoden	35
3.3 Bioinformatische und statistische Methoden	36
3.3.1 GenomeStudio.....	36
3.3.2 Detektion von CNVs in GenomeStudio	38

3.3.3 Genomweite <i>in silico</i> Detektion von CNVs	39
3.3.4 Qualitätskontrolle der untersuchten Personen und SNPs.....	41
3.3.5 Durchführung und Ausgabe von QuantiSNP/PennCNV	43
3.3.6 Qualitätskontrolle der detektierten CNVs	44
3.3.7 Zusammenfügen der Datensätze	45
3.3.8 Statistische Analysen.....	47
3.3.9 Einschätzung der biologischen Relevanz von CNVs	50
3.4 Verwendete Kollektive.....	52
3.4.1 Bipolare Störung	52
3.4.2 Majore Depression.....	53
3.3.3 Schizophrenie	54
3.5 Durchgeführte Analysen	55
3.5.1 Etablierung der Filterkriterien	55
3.5.2 Beteiligung von CNVs an der Erkrankungsentstehung.....	56
4 Ergebnisse	59
4.1 Etablierung von Filterkriterien	59
4.2 Beteiligung von CNVs bei der Bipolaren Störung	60
4.2.1 Qualitätskontrollierter Datensatz	60
4.2.2 Genomweite <i>burden</i> Tests	60
4.2.3 Spezifische Assoziationen	62
4.2.4 Frequenz von CNVs in vorbeschriebenen Regionen.....	63
4.3 Beteiligung von CNVs bei der Majoren Depression	65
4.3.1 Qualitätskontrollierter Datensatz	65
4.3.2 Genomweite <i>burden</i> Tests	65
4.3.3 Spezifische Assoziationen	65
4.3.4 Phänotypische Charakterisierung der 16p11.2-CNV-Träger	69
4.3.5 Frequenz von CNVs in vorbeschriebenen Regionen.....	70
4.4 Beteiligung von CNVs der Schizophrenie.....	72
4.4.1 Qualitätskontrollierter Datensatz	72
4.4.2 Genomweite <i>burden</i> Tests	72
4.4.3 Spezifische Assoziationen	72
4.4.4 Frequenz von CNVs in vorbeschriebenen Regionen.....	72
5 Diskussion	75
5.1 Etablierung der Auswertestrategie.....	75
5.2 Bipolare Störung	76
5.2.1 Genomweiter <i>burden</i>	76
5.2.2 Assoziation spezifischer Regionen.....	79

5.3 Majore Depression	81
5.3.1 Genomweiter <i>burden</i>	81
5.3.2 Assoziation spezifischer Regionen.....	82
5.4 Schizophrenie	84
5.5 Schlussfolgerungen	85
6 Zusammenfassung	87
7 Ausblick	89
8 Abbildungs- und Tabellenverzeichnis	92
8.1 Abbildungen.....	92
8.2 Tabellen.....	95
9 Literaturverzeichnis	97
10 Eigene Publikationen	110
11 Anhang	113
Danksagung	118
Lebenslauf	119
Erklärung	120

1 Einleitung

1.1 Genetische Variabilität

Das diploide menschliche Genom besteht aus 46 Chromosomen mit etwa sechs Milliarden Basenpaaren und knapp 30.000 protein-kodierenden Transkripten (Pruitt et al., 2012). Zwei zufällig ausgewählte, nicht miteinander verwandte Personen gleichen Geschlechts sind auf Ebene der DNA-Sequenz (*deoxyribonucleic acid*, Desoxyribonukleinsäure) zu etwa 99,8 % identisch. Ein Großteil dieser interindividuellen Variabilität basiert auf Einzelbasenunterschieden und strukturellen chromosomalen Veränderungen (siehe Kap. 1.2). Weitere Elemente, die zur genetischen Variabilität beitragen, sind zum Beispiel Trinukleotidwiederholungen, Mikrosatelliten (Wiederholungen von 2-6 Basenpaaren) und Minisatelliten (Wiederholungen von 10-100 Basenpaaren).

Unterschiede auf Einzelbasenniveau machen etwa 40 Prozent der interindividuellen Variabilität aus (Malhotra und Sebat, 2012). Die meisten dieser Unterschiede sind Einzelnukleotidpolymorphismen (SNPs, *single nucleotide polymorphisms*). SNPs sind Einzelbasenaustausche, bei denen das in der Bevölkerung seltenere Allel eine Frequenz größer oder gleich ein Prozent hat. SNPs werden in der dbSNP Datenbank des *National Center for Biotechnology Information* (NCBI) gesammelt (<http://www.ncbi.nlm.nih.gov/projects/SNP/>, Sherry et al., 2001). In der aktuellen Version der Datenbank (build 139, Oktober 2013) sind etwa 73 Millionen SNPs gelistet. In einer weiteren Datenbank sind die Ergebnisse genomweiter Assoziationsstudien mit SNPs aufgelistet. In dieser Datenbank, dem *GWAS Catalog*, sind derzeit 12.359 SNPs enthalten, die in 1.798 Publikationen als genetische Risikofaktoren für unterschiedliche Erkrankungen identifiziert wurden (<http://www.genome.gov/gwastudies/>; Hindorff et al., 2009).

Fokus der vorliegenden Arbeit waren die Kopienzahlvarianten (*copy number variants*, CNVs, siehe Kap. 1.3), eine weitere Art der genetischen Variabilität. Die Rolle dieser potentiellen, genetischen Risikofaktoren wurde bei den neuropsychiatrischen Erkrankungen Bipolare Störung, Majore Depression und Schizophrenie (siehe Kap. 1.4) untersucht.

1.2 Strukturelle Variabilität

Strukturelle Veränderungen der DNA können in mikroskopisch detektierbare (größer als drei Megabasenpaare (Mb)) und nicht detektierbare - also submikroskopische (kleiner als drei Mb) - Veränderungen unterteilt werden (Feuk et al., 2006). Strukturelle Veränderungen umfassen Inversionen, Translokationen, Deletionen sowie Duplikationen und betreffen die Orientierung (Inversionen), Position (Translokationen) oder Anzahl (Deletionen, Duplikationen) der betroffenen DNA-Abschnitte. Bei einer Inversion wird ein DNA-Abschnitt, im Vergleich zum Referenzgenom, um 180 Grad gedreht und wieder in die DNA-Sequenz integriert. Eine Translokation ist der Transfer eines DNA-Segments von einer genomischen Region in eine andere Region. Translokationen können sowohl interchromosomal als auch intrachromosomal stattfinden. Translokationen und Inversionen haben keinen Einfluss auf die Anzahl der Kopien eines DNA-Abschnittes, man bezeichnet sie daher als balancierte strukturelle Veränderungen. Im Gegensatz dazu verändern Deletionen und Duplikationen die Kopienzahl des betroffenen DNA-Segmentes. Daher bezeichnet man diese unbalancierten Veränderungen als Kopienzahlvarianten, kurz CNVs (*copy number variants*, siehe Kap. 1.3).

Im Gegensatz zu SNPs ist die Frequenz von strukturellen Veränderungen im menschlichen Genom geringer. Allerdings sind die einzelnen Ereignisse deutlich länger und betreffen insgesamt mehr Basenpaare als SNPs (Sebat, 2007).

1.3 Kopienzahlvarianten

Wie in Kapitel 1.2 beschrieben, handelt es sich bei CNVs um unbalancierte, submikroskopische strukturelle Veränderungen. In Bezug auf ihre Größe sind CNVs zwischen kleinen Varianten, die einzelne bis einige hundert Basenpaare betreffen (z.B. SNPs, Trinukleotidwiederholungen, Mikro- oder Minisatelliten, InDels) und im Lichtmikroskop sichtbaren chromosomalen Veränderungen (z.B. Trisomien oder Translokationen) einzuordnen. Die Kopienzahl des betroffenen DNA-Abschnittes weicht von der Referenzsequenz mit zwei Kopien ab. Bei Deletionen ist die Kopienzahl verringert; bei heterozygoten Deletionen fehlt eine der beiden Kopien, bei homozygoten Deletionen sind beide Kopien

deletiert (siehe Abb. 1.1). Ist die Zahl der Kopien eines DNA-Abschnittes im Vergleich zur Referenzsequenz erhöht, spricht man von einer Amplifikation. Die einfachste und häufigste Art der Amplifikation ist eine Duplikation einer der beiden Kopien, der entsprechende DNA-Abschnitt liegt also dreimal vor. Komplexere Amplifikationen mit vier und mehr Kopien desselben DNA-Abschnittes sind weniger häufig.

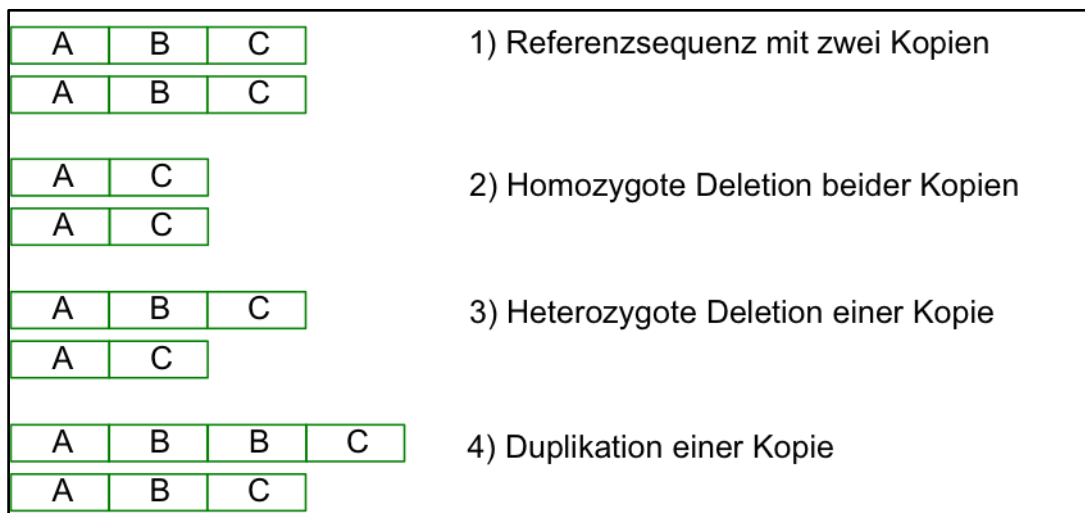


Abb.1.1: Schematische Darstellung von CNVs. In der Referenzsequenz liegt ein DNA-Segment (grüne Kästen) in zwei Kopien vor (1). Bei einer homozygoten Deletion (2) fehlen beide, bei einer heterozygoten Deletion (3) fehlt eine Kopie des betroffenen DNA-Segments (In diesen Beispielen Segment B). Bei Amplifikationen in die Kopienzahl des DNA-Segments erhöht, z.B. liegen drei Kopien vor (4).

Die Untersuchung von CNVs hat in den letzten Jahren viel Aufmerksamkeit erhalten. Durch die Verbesserung bestehender (z.B. *comparative genomic hybridization*, CGH) und Entwicklung neuer Detektionsmethoden (z.B. *next generation sequencing*) wurde die Auflösung, mit der potentielle CNVs entdeckt werden können, deutlich erhöht. Die exakte Bruchpunktbestimmung und Detektion von kleinen (d.h. kleiner als ein Kilobasenpaare (kb)) CNVs ist mittlerweile möglich. Die Definition eines CNVs ändert sich deswegen stetig. Derzeit gibt es keine einheitliche, standardisierte CNV Definition (Malhotra und Sebat, 2012). In der vorliegenden Arbeit werden alle Deletionen und Duplikationen, die mindestens eine Kilobase umfassen, als CNVs bezeichnet, da dies derzeit die gängigste Definition in der Literatur ist.

Im Folgenden wird eine Übersicht über die bislang durchgeführten Arbeiten auf dem Gebiet der CNVs gegeben; von Beobachtungen in einzelnen, spezifischen

Erkrankungen bis hin zu genomweiten Untersuchungen in großen populationsbasierten Kollektiven (siehe Kap. 1.3.2).

1.3.1 Mikrodeletions-/Mikroduplikationssyndrome

Zunächst wurden nur einzelne, spezifische CNVs und deren Rolle bei monogenen Erkrankungen untersucht. In diesen Fällen spricht man dann von Mikrodeletions-/Mikroduplikationssyndromen. Mittlerweile hat man für zahlreiche dieser Syndrome die ursächlichen Deletionen und Duplikationen identifizieren können.

Eine der ersten dieser Arbeiten zeigte, dass Duplikationen in der chromosomalen Subbande 17p11.2 ursächlich für die Charcot-Marie-Tooth Erkrankung sind (Lupski et al., 1991). Die reziproken Deletionen wurden wenig später als ursächlich für erbliche Neuropathie mit Neigung zu Druckpareesen (*hereditary neuropathy with liability to pressure palsies*; Chance et al., 1993) identifiziert. Bei beiden Erkrankungen handelt es sich um motorisch-sensorische Neuropathien.

Als einer der ersten genetischen Risikofaktoren für die Schizophrenie wurden bereits 1995 Deletionen auf Chromosom 22 identifiziert (Karayiorgou et al., 1995). Patienten mit einer Deletion in der chromosomalen Bande 22q11.21 haben im Vergleich zur Allgemeinbevölkerung ein um das 21-fache erhöhtes Risiko eine Schizophrenie zu entwickeln (The International Schizophrenia Consortium, 2008). Diese Deletionen sind der derzeit stärkste bekannte genetische Risikofaktor für die Schizophrenie. Neben dem erhöhten Risiko für eine Schizophrenie können Patienten mit einer Deletion in 22q11.21 unter anderem Herzfehler, Gaumenspalten und charakteristische Gesichtszüge aufweisen, daher bezeichnet man dieses Mikrodeletionssyndrom auch als velo-cardio-faziales Syndrom.

1.3.2 Genomweite, populationsbasierte CNV Analysen

Die Entwicklung neuer Techniken (siehe Kap. 1.3.6) mit einer genomweiten Markerabdeckung und geringeren Kosten ermöglichte die Untersuchung von großen populationsbasierten Kollektiven hinsichtlich CNVs. Die Ergebnisse dieser Studien zeigten, dass zuvor die Frequenz der CNVs in der Allgemeinbevölkerung deutlich unterschätzt worden war und dass CNVs einen

deutlich größeren Anteil an der genetischen Variabilität hatten als zuvor angenommen. Die Frequenz und Größe der detektierten CNVs in genomweiten Studien war abhängig von der verwendeten Detektionsmethode. Außerdem konnten CNVs mit einer geringen Frequenz in der Allgemeinbevölkerung nur in Studien mit einer entsprechend hohen Kollektivgröße entdeckt werden. Daher können die Ergebnisse der einzelnen Studien nicht direkt miteinander verglichen werden. Die Ergebnisse der einzelnen Studien deuten aber in die gleiche Richtung.

Erstmals zeigten zwei Studien 2004 unabhängig voneinander, dass Deletionen und Duplikationen häufig im Genom gesunder Personen vorkommen und über das gesamte Genom verteilt auftreten (Iafrate et al., 2004; Sebat et al., 2004). Diese Ergebnisse konnten von verschiedenen Gruppen (u.a. Sharp et al., 2005; Tuzun et al., 2005; Conrad, et al., 2006; Hinds, et al., 2006; Redon et al., 2006) repliziert werden.

Die umfangreichste der frühen genomweiten Studien wies bei 270 untersuchten Personen insgesamt 1.447 Regionen nach, in denen CNVs auftraten. Pro Person wurden durchschnittlich 5 CNVs mit einer Gesamtlänge von 14,3 Mb detektiert (Redon et al., 2006).

Im Jahr 2010 wurde die bis dato mit der höchsten Markerabdeckung durchgeführte populationsbasierte, genomweite CNV-Studie veröffentlicht (Conrad et al., 2010). In 450 untersuchten Personen wurden 8.599 CNVs identifiziert. Im Durchschnitt wurden 19 CNVs pro Person mit einer Gesamtlänge von 112,7 Mb, was in etwa 3,7 % des Genoms entspricht detektiert (Conrad et al., 2010).

Im Jahr 2011 wurde zum ersten Mal die Methode der genomweite Sequenzierung für die Detektion von CNVs in einem populationsbasierten Kollektiv verwendet (Mills et al., 2011). Bei 185 Personen wurden 28.025 CNVs detektiert, was eine deutlich größere Anzahl im Vergleich zu den bis dato publizierten Studien war. Grund hierfür war die wesentlich höhere Abdeckung der genomweite Sequenzierung. Dadurch wurden in erster Linie Varianten identifiziert, die deutlich kleiner als 1 kb waren und damit nicht der Definition

von CNVs in dieser Arbeit entsprachen. Diese Klasse der strukturellen Varianten ist jedoch interessant für zukünftige Analysen.

1.3.3 Katalogisierung von CNVs in der *Database of Genomic Variants*

Die *Database of Genomic Variants* (DGV, <http://dgv.tcag.ca/dgv/app/home/>; lafrate et al., 2004; Zhang, et al., 2006) ist die größte öffentlich zugängliche Datenbank für CNVs beim Menschen. Insgesamt 55 genomweite, populationsbasierte Studien sind in der DGV zusammengefasst. Die Einträge in der Datenbank wurden durch unterschiedliche technische CNV-Identifizierungsmethoden gewonnen (z.B. *array comparative genomic hybridization*, SNP-Genotypisierungschips, genomweite Sequenzierung; siehe Kap. 1.3.6) und unterscheiden sich deshalb in ihrer Markerabdeckung und Aussagekraft. In der aktuellsten Version der Datenbank (Juli 2013) sind 109.863 CNVs gelistet.

1.3.4 Entstehungsmechanismen von CNVs

CNVs können während der Mitose oder Meiose entstehen. Derzeit sind vier mögliche Mechanismen bekannt, die zu einer Veränderung der Kopienzahl führen können: nicht allelische homologe Rekombination (NAHR, *nonallelic homologous recombination*), nicht homologe Reparatur von Doppelstrangbrüchen (NHEJ, *nonhomologous DNA end joining*), Springen der Replikationsgabel (FoSTeS, *Fork Stalling and Template Switching*) und Retrotransposition.

Anhand von genomweiten Sequenzierungsdaten untersuchten Mills und Kollegen, in welchem Ausmaß die einzelnen Mechanismen zur Entstehung von CNVs beitragen (Mills et al., 2011). Duplikationen und Deletionen, die länger als 1 kb waren, waren in mehr als 95 % der Fälle von segmentalen Duplikationen flankiert. Dies deutet darauf hin, dass größere, rezidivierende CNVs hauptsächlich durch NAHR verursacht wurden.

Nonallelic homologous recombination

Während der Metaphasen der Meiose und Mitoselagern sich die homologen Chromosomenpaare aneinander an und das so genannte *crossing over* kann stattfinden. Dabei werden Doppelstrangbrüche der beteiligten DNA-Moleküle zusammengefügt. Die Doppelstrangbrüche entstehen in homologen

Sequenzen, die auf beiden Chromosomen vorhanden sind (Alberts et al., 2002). Diese Art der DNA-Rekombination wird auch als allelische homologe Rekombination bezeichnet, die homologen Sequenzen sind dabei die Allele.

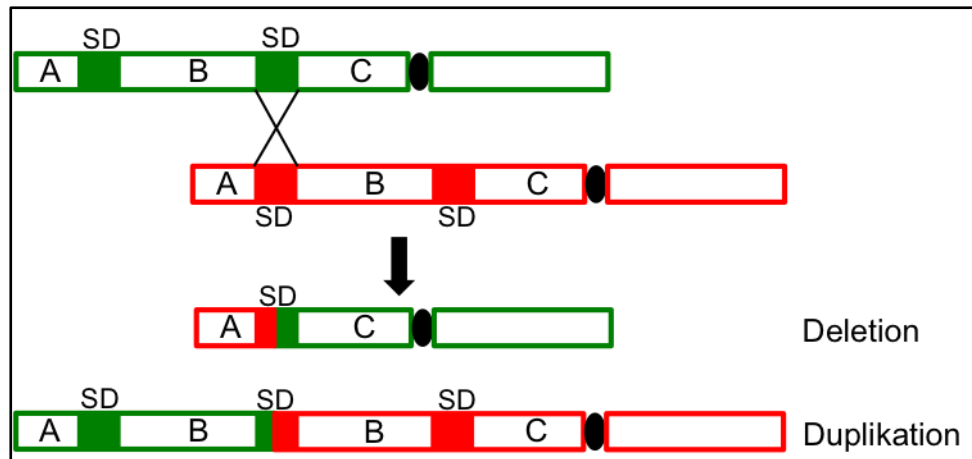


Abb. 1.2: Schematische Darstellung der *nonallelic homologous recombination*. Zwischen DNA-Segmenten mit hoher Homologie, den segmentalen Duplikationen (SD) kann es zu einer Rekombination kommen. Sind diese segmentalen Duplikationen in die gleiche Richtung ausgerichtet, entsteht eine Deletion und eine Duplikation. Die DNA-Abschnitte sind mit A, B und C bezeichnet. Hier ist der Abschnitt B von NAHR betroffen. Die homologen Chromosomen sind grün bzw. rot markiert. (modifiziert nach Malhotra und Sebat, 2012).

Die NAHR findet durch ungleiches *crossing over* zwischen segmentalen Duplikationen statt (Pentao et al., 1992). Segmentale Duplikationen sind DNA-Segmente, die in zwei oder mehr Kopien im Genom auftreten und untereinander eine Sequenzidentität von mehr als 90 % aufweisen (Scherer et al., 2007). Segmentale Duplikationen können dazu führen, dass sich Chromosomen oder Chromatide während der Meiose (Turner et al., 2008) und seltener während der Mitose (Lam und Jeffreys, 2006) versetzt aneinander anlagern (siehe Abb. 1.2) und so eine Rekombination zwischen verschiedenen DNA-Abschnitten bedingen. Wenn die segmentalen Duplikationen zwischen denen die NAHR stattfindet in die gleiche Richtung ausgerichtet sind, entsteht auf einem der beteiligten DNA-Moleküle eine Deletion und auf dem anderen eine Duplikation des dazwischen liegenden DNA-Abschnittes (A, B und C in Abb. 1.2). Im Gegensatz dazu entstehen bei gegensätzlich orientierten segmentalen Duplikationen keine CNVs, sondern balancierte Inversionen. NAHR kann in drei Weisen erfolgen: zwischen segmentalen Duplikationen desselben Chromatids, von Schwesterchromatiden oder von homologen Chromosomen (Turner et al., 2008). Im Genom sind Regionen bekannt in denen die Rate für NAHR stark erhöht ist. In diesen *hotspots* ist die

Rekombinationsrate um das bis zu 237-fache im Vergleich zum restlichen Genom erhöht (Hurles und Lupski, 2006).

Nonhomologous DNA end joining

Doppelstrangbrüche der DNA, die in jeder Zelle auftreten können, werden durch enzymatische Reaktionen repariert (Lieber, 2008). Ein Mechanismus für die Reparatur solcher Doppelstrangbrüche ist das NHEJ. Beim NHEJ werden die Doppelstrangbrüche durch DNA-Ligation zusammengefügt (siehe Abb. 1.3A; Alberts et al., 2002). Bei dieser schnellen Art der Reparatur können Fehler auftreten. Dabei handelt es sich häufig um kleine Deletionen oder Insertionen von einigen Basen (Malhotra und Sebat, 2012). Durch kurze, homologe Sequenzen, sogenannte Mikrohomologien (MH in Abb. 1.3), kann es zu größeren Deletionen kommen. Ähnlich wie bei der NAHR kommt es zwischen diesen Mikrohomologien zu einer Rekombination und damit einer Deletion des dazwischen liegenden DNA-Segments (siehe Abb. 1.3B; Malhotra und Sebat, 2012). Diese Deletion kann nur wenig Basen oder einige Kilobasen lang sein, abhängig von der Lage der Mikrohomologien.

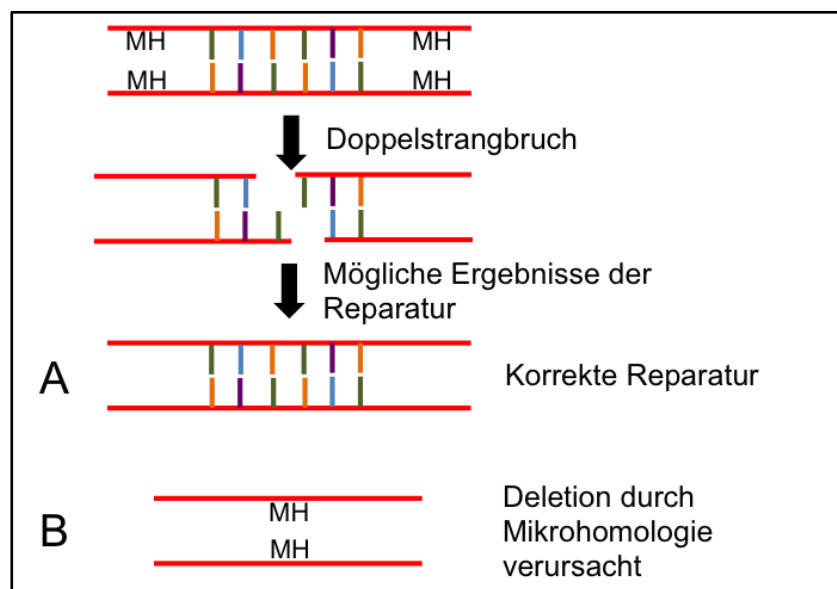


Abb. 1.3: Schematische Darstellung des *nonhomologous DNA end joinings*. Bei der Reparatur von Doppelstrangbrüchen durch NHEJ kann es zu Fehlern mit verschiedenen Resultaten kommen: A: korrekte Reparatur; B: Deletion durch Rekombination zwischen Mikrohomologien (MH). Farbige vertikale Striche repräsentieren Basenpaare (modifiziert nach Malhotra und Sebat, 2012).

Fork Stalling and Template Switching

FoSTeS ist ein Mechanismus, der bei der Replikation der DNA auftritt (Lee et al., 2007). Wenn die Replikationsgabel auf einen Bruch im DNA-Einzelstrang trifft, bricht die Replikation an dieser Stelle ab. Durch Anlagerung an kurze homologe Sequenzen, den Mikrohomologien, kann der Rückwärtsstrang eine neue Replikationsgabel bilden (siehe Abb. 1.4). Je nach Abstand und Orientierung zwischen der neuen und der alten Replikationsgabel entsteht eine Deletion oder Duplikation von unterschiedlicher Größe. Die so entstandenen CNVs können DNA-Segmente aus unterschiedlichen Regionen beinhalten (Malhotra und Sebat, 2012).

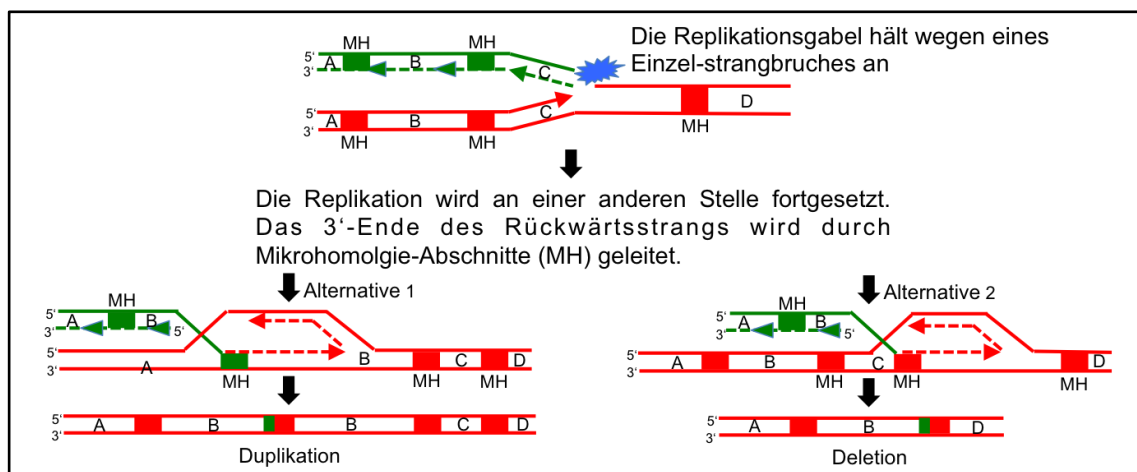


Abb. 1.4: Schematische Darstellung des Fork Stalling and Template Switching. Stößt die Replikationsgabel auf einen Einzelstrangbruch (blauer Stern) dann kommt es zu einem Stopp der Replikation. Die aktive Replikationsgabel kann durch Anlagern an Mikrohomologien (MH) an anderer Stelle die Replikation fortsetzen. Die DNA-Abschnitte zwischen den Mikrohomologien sind mit A, B, C und D gekennzeichnet. Je nach Lokalisation der Mikrohomologien entsteht eine Duplikation oder Deletion (modifiziert nach Malhotra und Sebat, 2012).

Retrotransposition

Retrotransposons sind DNA-Abschnitte, die sich selbst vervielfältigen und durch reverse Transkription an anderer Stelle in die DNA wieder integrieren können. Aktive Retrotransposons kodieren für die RNA-Polymerase II, die für die Transkription des Retrotransposons verantwortlich ist. Außerdem beinhaltet einer der beiden *open reading frames* (ORF, siehe Abb. 1.5) des Retrotransposons, der *ORF2*, eine reverse Transkriptase und Endonuklease. Beide Enzyme sind für die Funktion des Retrotransposons nötig. Die DNA eines aktiven Retrotransposons wird zunächst in RNA transkribiert. Bei der Translation der RNA wird die reverse Transkriptase gebildet. Diese ist

spezifisch für die RNA des Retrotransposons und wandelt diese in DNA um. Die Endonuklease fügt die so neu entstandene Retrotransposon-DNA wieder in die DNA ein. Dabei hat sie spezifische Zielsequenzen für die Insertion (siehe Abb. 1.5). Die ursprüngliche DNA des Retrotransposons verbleibt unverändert in der DNA. Durch die Retrotransposition wird die Kopienzahl des aktiven Retrotransposons erhöht.

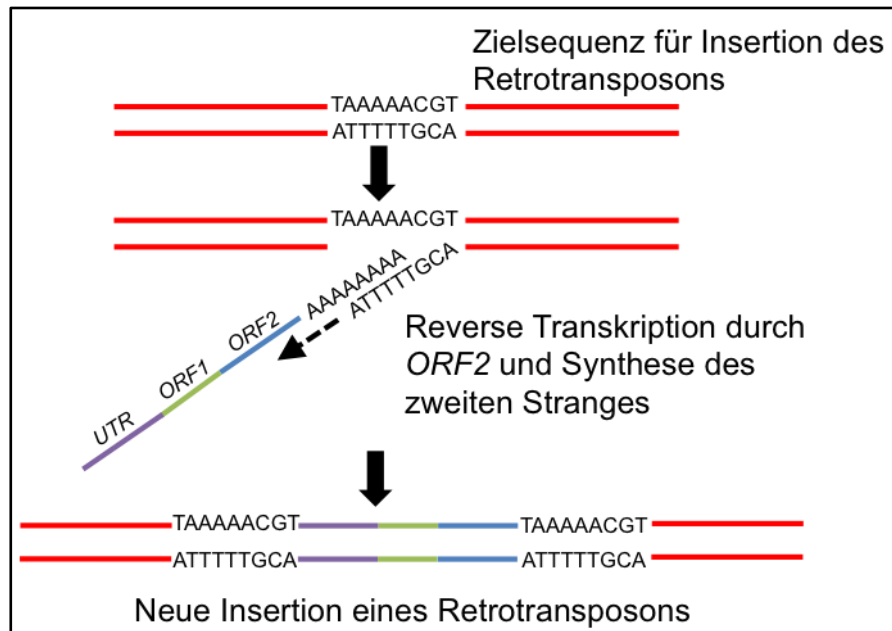


Abb. 1.5: Schematische Darstellung der Funktion eines Retrotransposons. Ein aktives Retrotransposon wird zunächst translatiert und transkribiert. An einer geeigneten Zielsequenz wird die, durch reverse Transkription entstandene Retrotransposon-DNA eingefügt. Die dafür nötigen Enzyme (eine reverse Transkriptase und eine Endonuklease) sind auf dem Retrotransposon selbst kodiert. Das ursprüngliche Retrotransposon verbleibt unverändert in der DNA, durch die Retrotransposition wird seine Kopienzahl erhöht (modifiziert nach Malhotra und Sebat, 2012).

1.3.5 Effekt von CNVs auf die Genexpression

In Abhängigkeit von der genomischen Position und Größe eines CNVs kann dieser auf unterschiedliche Weise einen Effekt auf die Expression eines oder mehrerer Gene haben (siehe Abb. 1.6). Die so veränderte Genexpression kann ursächlich für unterschiedliche Phänotypen sein. CNVs, die direkt ein Gen betreffen verändern die Anzahl der funktionell aktiven Genkopien (siehe Punkte 1-3, Abb. 1.6). Auch Gene, die nicht direkt von einem CNV betroffen sind, können durch diesen in ihrer Expression verändert werden (siehe Punkte 4-6, Abb. 1.6).

1) Komplettes Gen betroffen:

Überspannt ein CNV die komplette kodierende Sequenz eines Gens, so wird die Zahl der funktionellen Genkopien entweder erhöht (bei einer Duplikation) oder verringert (bei einer Deletion, siehe Abb. 1.6).

2) Ein Gen partiell betroffen:

Ist ein Gen nur zum Teil von einem CNV betroffen, wird die Anzahl funktioneller Genkopien verringert, unabhängig davon ob es sich um eine Deletion oder Duplikation handelt (siehe Abb. 1.6).

3) Zwei Gene partiell betroffen:

Wenn beide Bruchpunkte eines CNVs innerhalb der kodierenden Sequenz zweier Gene liegen, kommt es zur Genfusion. Die Zahl der funktionellen Kopien der beiden betroffenen Gene ist reduziert (siehe Abb. 1.6).

4) Positionseffekt

Betrifft ein CNV nur ein regulatorisches Element, jedoch kein Gen direkt kann dies einen Effekt auf die Expression eines oder mehrerer Gene haben. Je nach Art des regulatorischen Elements (*Enhancer* oder *Silencer*) kann die Expression erhöht oder verringert werden (siehe Abb. 1.6). Ist zum Beispiel ein *Enhancer* von einer Duplikation betroffen so kann die Expression der durch diesen *Enhancer* regulierten Gene weiter erhöht werden.

5) Rezessive Mutationen und funktionelle Polymorphismen

Deletionen können die Demaskierung eines rezessiven Allels oder eines funktionell relevanten Polymorphismuses verursachen (siehe Abb. 1.6). Der Effekt der Variante kommt durch die Deletion des kompensierenden Allels zum Tragen.

6) Veränderung der Chromatinstruktur

Insbesondere große CNVs (d.h. mehrere hundert Kilobasen) können dazu führen, dass die Chromatinstruktur der betroffenen Region verändert wird. Dies kann die 3D-Struktur der DNA verändern und so die Genexpression beeinflussen. So kann es beispielsweise sein, dass durch die veränderte 3D-Struktur Teile der DNA nicht mehr in RNA transkribiert werden können.

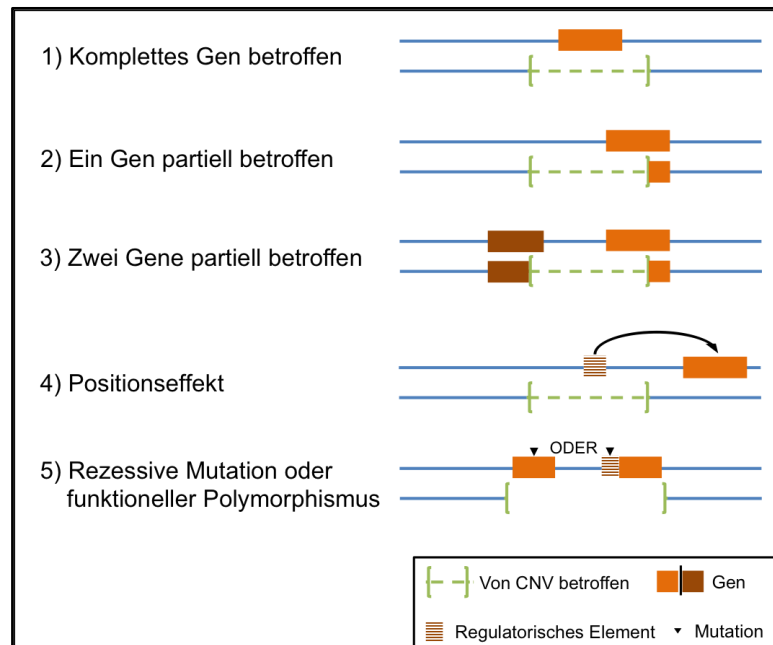


Abb. 1.6: Mögliche Auswirkungen von CNVs auf die Genexpression 1) Ist ein komplettes Gen (Oranges/braunes Rechteck) von einem CNV (grüne, gestrichelte Linie) überspannt, ist die Zahl der funktionellen Genkopien erhöht (bei einer Duplikation) oder verringert (bei einer Deletion). 2) Ist ein Gen nur teilweise betroffen, verringert sich die Zahl der aktiven Genkopien. 3) Liegen die Bruchpunkte eines CNVs in zwei Genen kann ein Fusionsgen entstehen. 4) Wenn ein regulatorisches Element (braun-weißes Rechteck) von einem CNV überspannt wird, kann dies einen Einfluss auf die Expression eines oder mehrerer Gene haben. 5) Eine Deletion kann dazu führen, dass eine rezessive Mutation oder ein funktioneller Polymorphismus demaskiert wird, wenn das kompensierende Allel deletiert wird. Abbildung modifiziert nach: Lupski und Stankiewicz, 2005.

Insbesondere bei größeren CNVs ist eine Kombination der verschiedenen Mechanismen denkbar. Dadurch ist die Bestimmung der dem jeweiligen Phänotypen zugrunde liegenden Gene nicht unmittelbar möglich. Mehrere Gene könnten von einem größeren CNV betroffen sein und außerdem noch regulatorische Elemente, die wiederum mehrere Gene regulieren. Der genaue Wirkungsmechanismus eines CNVs lässt sich also nicht immer direkt bestimmen.

1.3.6 Genomweite Detektionsmethoden von CNVs

Unterschiedliche Herangehensweisen können für die genomweite Identifikation von CNVs verwendet werden: *array comparative genomic hybridization*, SNP-Genotypisierungschips und Genomweite Sequenzierung.

Array comparative genomic hybridization

Die erste Methode ist die Array basierte vergleichende Genomhybridisierung (aCGH, *array comparative genomic hybridization*). Dabei verwendet man Fluoreszenz-markierte Fragmente des zu untersuchenden Genoms und

Fragmente eines Referenz-Genoms. Die Fragmente sind jeweils andersfarbig markiert und werden mit Sonden hybridisiert, die auf Mikroarrays immobilisiert sind. Anhand der gemessenen Intensitäten der hybridisierten Fragmente kann man rückschließen, ob an der entsprechenden Stelle ein CNV vorliegt oder nicht (Solinas-Toldo et al., 1997). Die Auswahl und Verteilung der verwendeten Hybridisierungssonden ist dabei entscheidend für die Anwendung und Auflösung des aCGH.

SNP-Genotypisierungschips

Die zweite Array basierte Methode verwendet die Signalintensitäten von SNP-Chips. Diese Chips wurden ursprünglich für die genomweite Genotypisierung von SNPs entwickelt. Sie verwenden Fluoreszenz-markierte Marker für die Bestimmung der Genotypen der untersuchten SNPs (siehe Kap. 3.2.3). Die Fluoreszenzintensität kann auch verwendet werden, um mit Hilfe von bioinformatischen Programmen (siehe Kap. 3.3.3) CNVs zu detektieren. Die beiden Hauptproduzenten von kommerziellen SNP-Chips sind Affymetrix (Santa Clara, Kalifornien, USA) und Illumina (San Diego, Kalifornien, USA). In der vorliegenden Arbeit wurden SNP-Chips der Firma Illumina für die Identifizierung von CNVs verwendet (siehe Kap. 3.2.3).

Genomweite Sequenzierung

Die dritte und neueste Technologie ist die genomweite Sequenzierung. Mit dieser Technologie ist die Basenpaar-genaue Bestimmung der Bruchpunkte eines CNVs möglich (Korbel et al., 2007). Die Kosten für diese Methode waren anfangs noch zu hoch, um in großen Kollektiven verwendet werden zu können. Inzwischen wurden aber die ersten Studien veröffentlicht, bei denen größere Kollektive genomweit sequenziert wurden (Gudmundsson et al., 2012; The 1000 Genomes Project Consortium, 2012).

1.4 Neuropsychiatrische Erkrankungen

Zu den neuropsychiatrischen Erkrankungen gehören die Schizophrenie, die Bipolare Störung und die Major Depression. Die Lebensqualität von Patienten, die unter einer dieser Erkrankungen leiden, ist deutlich eingeschränkt. Derzeit können lediglich die Symptome der drei Erkrankungen behandelt werden. Im Vergleich zur Allgemeinbevölkerung ist zudem die Sterblichkeit der Patienten

aufgrund eines erhöhten Suizidrisikos und Komorbiditäten wie zum Beispiel Erkrankungen des Herz-Kreislauf-Systems erhöht (Eaton et al., 2008).

Die drei Erkrankungen gehören zu den genetisch komplexen, multifaktoriellen Erkrankungen. Das bedeutete, dass mehrere genetische Veränderungen im Zusammenspiel mit Umweltfaktoren und epigenetischen Faktoren zur Krankheitsentstehung beitragen.

Bipolare Störung

Patienten mit einer Bipolaren Störung leiden unter extremen Schwankungen der Stimmung. Sie erleben Phasen von Depression im Wechsel mit manischen Phasen, in denen die Stimmung stark überhöht ist. Entsprechend dem Diagnostischen und Statistischen Handbuch Psychischer Störungen (DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders fourth edition*; American Psychiatric Association, 2007) wird zwischen zwei Typen der Bipolaren Störung unterschieden. Patienten mit einer Bipolar I Störung haben starke depressive und manische Episoden. Bei Patienten mit Bipolar II Störung sind die manischen Episoden weniger stark ausgeprägt, sie werden auch als hypomane Episoden bezeichnet. Die Episoden können durch Phasen unterbrochen sein, in denen die Patienten euthym sind, also keine Hoch- oder Tiefphasen durchleben. Die Lebenszeitprävalenz (also die Anzahl der bestehenden Erkrankungsfälle über den Lebensverlauf im Verhältnis zur Anzahl der Personen in der Bevölkerung in der gleichen Zeiteinheit) der Bipolaren Störung in Deutschland wird auf etwa 3 Prozent geschätzt und ist bei Männern und Frauen gleich (DGPPN, 2012). Basierend auf Zwillings- und Familienstudien wird davon ausgegangen, dass die Bipolare Störung zu 60 bis 80 Prozent durch erbliche Faktoren bedingt ist (McGuffin et al., 2003; Lichtenstein et al., 2009; Wray and Gottesman, 2012).

Majore Depression

Patienten mit einer Depression leiden unter einer gedrückten Stimmung und starker Interessenlosigkeit (DSM-IV, American Psychiatric Association, 2007). Häufig zeigen sie außerdem ein gestörtes Schlaf- und Essverhalten, sowie Konzentrationsstörungen (DSM-IV, American Psychiatric Association, 2007). Oft ziehen sich Patienten mit einer Majoren Depression sozial zurück. Das Risiko im Laufe des Lebens an einer Depression zu erkranken liegt bei 12 bis

26 Prozent wobei Frauen häufiger erkranken als Männer (Hautzinger, 2006). Die Heritabilitätsschätzungen für die Major Depression variieren stark, sie liegen zwischen 30 und 75 % (Kendler, 1993; McGuffin et al., 1996; Sullivan et al., 2000; Wray and Gottesman, 2012).

Schizophrenie

Bei Patienten mit einer Schizophrenie sind die Wahrnehmung, das Denken und die Empfindung erheblich gestört. Dies kann sich in folgenden Symptomen äußern: Wahn, Halluzinationen, desorganisierte Sprechweise, grob desorganisiertes Verhalten oder flacher Affekt (DSM-IV, American Psychiatric Association, 2007). Die Patienten können unter anderem unter optischen, taktilen oder akustischen Sinnestäuschungen leiden. Ein Großteil der Patienten zeigt außerdem einen Verfolgungswahn. Die Erkrankten weisen häufig kognitive Defizite auf, dies kann sich unter anderem in der Unfähigkeit komplexe Zusammenhänge zu erfassen oder in einem verarmten sprachlichen Ausdruck äußern. Bei vielen Patienten ist eine reduzierte Mimik und Gestik zu beobachten. Die Lebenszeitprävalenz für eine Schizophrenie liegt zwischen 0,5 und 1,6 Prozent, Männer und Frauen sind gleich häufig betroffen, Männer erkranken jedoch etwas früher (DGPPN, 2006). Der Anteil genetischen Faktoren wird auf 65 bis 80 Prozent geschätzt (Sullivan et al., 2003; Lichtenstein et al., 2009; Wray and Gottesman, 2012).

1.4.1 CNVs bei neuropsychiatrischen Erkrankungen

Zu Beginn der vorliegenden Arbeit war nur eine genomweite Studie zur Beteiligung von CNVs bei der Entstehung der Bipolaren Störung publiziert (Zhang et al., 2009). Für die Major Depression war noch keine genomweite CNV-Studie durchgeführt worden. Die Relevanz von CNVs bei der Entstehung der Schizophrenie war bereits besser verstanden (Karayiorgou et al., 1995; Walsh et al. 2008; Xu et al., 2008; Stefansson et al., 2008; The International Schizophrenia Consortium, 2008).

Bei der Bipolaren Störung hatte die Studie von Zhang et al. zum ersten Mal genomweit die Verteilung von CNVs bei Patienten untersucht. Diese Studie zeigte, dass große (> 100 kb), seltene CNVs (mit einer Frequenz von weniger

als einem Prozent in der Allgemeinbevölkerung) signifikant häufiger bei Patienten als bei Kontrollen zu finden waren (Zhang et al., 2009).

Eine Beteiligung von seltenen CNVs bei der Entwicklung einer Schizophrenie wurde in zwei unabhängigen, genomweiten Studien nachgewiesen (Walsh et al. 2008; Xu et al., 2008). Kurz darauf wurden, durch genomweite Assoziationsstudien, spezifische Regionen identifiziert, in denen CNVs mit der Schizophrenie assoziiert waren. (1q21.1, 15q11.2, 15q13.3 und 22q11.21 Stefansson et al., 2008; The International Schizophrenia Consortium, 2008). Wie in Kapitel 1.2 beschrieben, war eine Beteiligung von Deletionen in der chromosomalen Bande 22q11.21 bei der Entstehung der Schizophrenie bereits bekannt (Karayiorgou et al., 1995).

2 Zielsetzung der Arbeit

Die Beteiligung von genetischen Faktoren bei der Entstehung von neuropsychiatrischen Erkrankungen ist durch formalgenetische Studien wie z.B. Zwillings- und Familienstudien belegt. Die bislang identifizierten, häufigen genetischen Risikofaktoren (in der Regel SNPs) erklären jedoch nur einen Teil der Erblichkeit dieser Erkrankungen.

Ziel der vorliegenden Arbeit war die systematische, genomweite Untersuchung der Beteiligung von Kopienzahlvarianten (*copy number variants*, CNVs) als weitere potentielle genetische Risikofaktoren bei den neuropsychiatrischen Erkrankungen Bipolare Störung, Majore Depression und Schizophrenie. Zu Beginn dieser Arbeit waren nur wenige solcher genomweiten Untersuchungen veröffentlicht. In diesen konnte gezeigt werden, dass CNVs mögliche Risikofaktoren mit großer Effektstärke für die neuropsychiatrischen Erkrankungen Autismus und Schizophrenie sind. Bei der Bipolaren Störung und der Majoren Depression gab es zu Beginn der Arbeit noch kaum Untersuchungen zum Einfluss von CNVs auf die Krankheitsentstehung.

Zunächst musste eine Analysestrategie für die *in silico* Identifikation von CNVs und nachfolgenden Tests auf Assoziation der identifizierten CNVs etabliert werden. Im Gegensatz zu genomweiten SNP-Assoziationsuntersuchungen war und ist die Qualitätskontrolle von genomweiten CNV-Daten in der bislang publizierten Literatur nicht einheitlich. Die Etablierung dieser Qualitätskontrollparameter war daher ein Kernpunkt dieser Arbeit.

Die entwickelte Analysestrategie wurde dann auf große, deutsche Kollektive mit den neuropsychiatrischen Erkrankungen Bipolare Störung, Majore Depression und Schizophrenie angewendet. Das untersuchte Kollektiv für Bipolare Störung bestand aus 882 Patienten und 872 Kontrollen. Für Majore Depression wurden 575 Patienten und 1.618 Kontrollen analysiert. Der Datensatz für die Schizophrenie bestand aus 1.637 Patienten und 1.627 Kontrollen. In diesen Kollektiven sollte die Beteiligung von CNVs bei der Entstehung der Erkrankungen untersucht werden.

3 Material und Methoden

3.1 Verwendete Materialien

3.1.1 Verwendete Geräte

Autoklav	Varioklav 135 S Dampfsterilisator, H+P Labortechnik, Oberschleißheim
DNA-Bank/Verwaltung	2D CYPHER Tubes, 2DCYPHER 1,2 ml Cluster Tube Racks und 2D CYPHER SmartScan Solo, und SmartScan 96, ABgene, Thermo Scientific, Waltham, Massachusetts, USA
DNA-Extraktion	Chemagic Magnetic Separation Module I, Chemagen PerkinElmer, Baesweiler
Eismaschine	AF 100, Scotsman, Milano, Italien
Genotypisierungssysteme	iScan System, Illumina, San Diego, Kalifornien, USA
Kapillarelektrophorese	3130xl Genetic Analyzer, Life Technologies, Carlsbad, Kalifornien, USA
Mikrowelle	MW 9625, SEVERIN Elektro, Sundern
Multipipetten	Multipette/Transferpipette, Brand, Wertheim
Photometer	NanoDrop ND-1000 UV-Vis Spektrophotometer, Peq Lab, Erlangen
Pippettierhilfen	Eppendorf Research Plus 2,5 µl, 10 µl, 20 µl, 100 µl, 200 µl und 1000 µl, Eppendorf AG, Hamburg Research Plus 8-Kanal 2,5 µl - 10 µl und 10 µl - 100 µl, Eppendorf AG, Hamburg
Pipettierroboter	Biomek NX-MC, Beckman-Coulter, Fullerton, Kalifornien, USA Biomek NX-S8, Beckman-Coulter, Fullerton, Kalifornien, USA

Pumpensysteme	PC 2004 Vario, Vacuubrand & Co. KG MZ 2C Membran Vacuum Pumpe, Vacuubrand & Co. KG, Wertheim
<i>Real-time</i> PCR Gerät	7900HT, Applied BioSystems Foster City, Kalifornien, USA
Reinstwasser System	Milli-Q A10 Synthesis, Millipore, Molsheim Frankreich
Schüttler	Unimax 1010/Polymax 1040, Heidolph, Schwabach
Thermocycler	PTC 200, MJ Research, Watertown, Massachusetts, USA PTC 100, MJ Research, Watertown, Massachusetts, USA
Tischzentrifugen	Heraeus Megafuge 1.0, Thermo Scientific, Waltham, Massachusetts, USA Heraeus Megafuge 1.0 R, Thermo Scientific, Waltham, Massachusetts, USA Heraeus Biofuge fresco, Thermo Scientific, Waltham, Massachusetts, USA neoLab-Mini-Zentrifuge Spectrafuge, Heidelberg
Trockenschrank	Heraeus FunctionLine, Thermo Scientific, Waltham, Massachusetts, USA
Vortexer	Vortex Genie 2, Sientific Industries, Bohemia, New York, USA
Waage	Sartorius Talent, Sartorius AG, Göttingen

3.1.2 Chemikalien und Lösungen

Soweit möglich, wurden Chemikalien mit der Qualitätsbezeichnung „pro analysis“ verwendet.

- Agarose D1 Low EEO, Chromatrin, Dublin, Irland

- Dimethylsulfoxid, Sigma Aldrich, St. Louis, Missouri, USA
- Ethylendiamintetraessigsäure reinst, Merck, Darmstadt
- Essigsäure (Eisessig) 100 % Merck, Darmstadt
- Sephadex, G50 super fine, GE Healthcare, Little Chalfont, Großbritannien
- Tris $\geq 99,9$ %, Carl Roth, Karlsruhe
- Wasser PCR reagent, Sigma Aldrich, St. Louis, Missouri, USA

3.1.3 Kommerzielle Systeme (Kits)

- HumanHap550v3 DNA Analysis Kit, Illumina, San Diego, Kalifornien, USA
- Human610-Quadv1 DNA Analysis Kit, Illumina, San Diego, Kalifornien, USA
- Human660W-Quadv1 DNA Analysis Kit, Illumina, San Diego, Kalifornien, USA
- Power SYBR Green PCR Master Mix, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Reference Assay RNase P, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs02902865_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs04311265_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs03589363_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs03749977_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs03744366_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs03283591_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs04402646_cn“, Applied Biosystems, Foster City, Kalifornien, USA

- TaqMan Copy Number Assay „Hs03290371_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs02263567_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs06225510_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Copy Number Assay „Hs06195263_cn“, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Gene Expression Master Mix, Applied Biosystems, Foster City, Kalifornien, USA
- TaqMan Gene Genotyping Master Mix, Applied Biosystems, Foster City, Kalifornien, USA

3.1.4 Software und Datenbanken

- ABgene 2D CYPHER Pilot Databases, Thermo Scientific, Waltham, Massachusetts, USA
- Biomek Software 3.2, Beckman/Coulter, Fullerton, Kalifornien, USA
- CHOP <http://cnv.chop.edu/>
- DGV <http://dgv.tcag.ca/dgv/app/home>
- DECIPHER <http://decipher.sanger.ac.uk/>
- BeadScan, Illumina, San Diego, Kalifornien, USA
- GenomeStudio v2011.1, Illumina, San Diego, Kalifornien, USA
- Haploview <http://www.broad.mit.edu/mpg/haploview/index.php/>
- HapMap <http://www.hapmap.org/>
- HGNC <http://www.genenames.org/>
- Mac OSX 10.7.5 <http://www.apple.com/de/>
- NCBI <http://www.ncbi.nlm.nih.gov/>
- NanoDrop ND-100 v3.3.0, Peqlab, Erlangen
- PennCNV <http://www.openbioinformatics.org/penncnv/>
- Perl <http://www.perl.org/>
- PLINK <http://pngu.mgh.harvard.edu/~purcell/plink/>
- PubMed <http://pubmed.com/>
- QuantiSNP <http://www.well.ox.ac.uk/QuantiSNP/>

- R <http://www.r-project.org/>
- SLEP <http://gbrowse.csbio.unc.edu/cgi-bin/gb2/gbrowse/slep/>
- Ubuntu Linux <http://www.ubuntu.com/>
- UCSC Genome Bioinformatics <http://genome.ucsc.edu/>

3.2 Molekularbiologische Methoden

3.2.1 Probandenmaterial und DNA-Isolierung

Die DNA der Probanden wurde aus venösem Blut extrahiert. Diese Extraktion erfolgte entweder mit der Aussalzmethode nach Miller et al. (Miller et al., 1988) oder mit dem Chemagic Magnetic Separation Module I (PerkinElmer, Baesweiler).

3.2.2 Qualitäts- und Konzentrationsbestimmung der DNA

Für die Bestimmung der DNA Qualität und Quantität wurde ein Photospektrometer (Peq Lab, Erlangen) verwendet. Die DNA-Konzentration wurde über die gemessene Absorption bei 260 nm ermittelt. Für die Ermittlung der DNA-Qualität wurden der Quotient der Absorption bei 260 nm und 280 nm (A_{260}/A_{280}) verwendet. Dieser sollte zwischen 1,7 und 2,0 liegen, damit Verunreinigungen der DNA mit Proteinen, Salzen, Phenolen oder Ähnlichem ausgeschlossen werden konnten.

3.2.3 Genomweite Genotypisierung mit Illumina

Für die Genotypisierung wurde die Infinium BeadArray Plattform der Firma Illumina (San Diego, Kalifornien, USA) verwendet. Dabei werden die Genotypen der einzelnen SNPs durch eine allelspezifische Einzelbasenerweiterung (SBE, *single base extension*; Steemers et al., 2006) ermittelt.

Für die Genotypisierung wurden Chips mit Siliziumkugeln, den *beads*, verwendet. An jeden *bead* waren mehrere hundert, 50 bp lange fluoreszenzmarkierte Sonden gekoppelt. Alle Sonden eines *beads* waren spezifisch für einen SNP, jeder SNP-spezifische *bead* war im Mittel 20-fach auf einem Chip vorhanden (Gunderson et al., 2005; Steemers und Gunderson, 2007). Neben der SNP-spezifischen Sequenz trug jeder *bead* auch eine einmalige Sequenz, mit der der *bead* identifiziert werden konnte.

Die Genotypisierung bestand aus vier Schritten (Gunderson et al., 2005; Steemers et al., 2006):

1) Genomweite Amplifizierung und Fragmentierung der DNA

Die Amplifizierung der DNA erfolgte in einem isothermalen Schritt über Nacht. Durch eine bis zu 1500-fache Amplifizierung der DNA waren nur relativ kleine Mengen genomischer DNA (250-750 ng) nötig, um qualitativ hochwertige Genotypisierungsdaten zu erhalten (Steemers et al., 2006). Die amplifizierte DNA wurde dann enzymatisch fragmentiert. Nach Präzipitation und Resuspendierung der DNA folgte die Hybridisierung der DNA an die Sonden der *beads*.

2) Hybridisierung der amplifizierten DNA an sequenzspezifische Sonden

Abhängig vom untersuchten SNP waren diese Sonden unterschiedlich gewählt (siehe Abb. 3.1 b+c). Beim Infinium I Design gab es für jeden SNP zwei Sonden, die sich in nur einer Base am 3'-Ende unterschieden. Die eine Sonde stimmte perfekt mit der Zielsequenz des einen SNP-Allels überein; die zweite Sonde stimmte exakt mit dem anderen Allel überein. Beim Infinium II Design gab es nur eine Sonde für jeden SNP. Die Sequenz der Sonde unterschied sich nur in der letzten Base vor dem zu untersuchenden SNP. In beiden Fällen folgte nach perfekter Hybridisierung der Zielsequenz an die Primer eine SBE (Steemers et al., 2006).

Mit dem Infinium I Design ließen sich SNPs jeglicher Ausprägung (A/T, A/G, A/C, C/T, C/G, G/T) detektieren, da die Allelverteilung aus dem Verhältnis der gemessenen Intensitäten beider verwendeten Sonden berechnet wurde (siehe Abb. 3.1 b). Allerdings waren hierfür zwei Sonden nötig. Beim Infinium II Design war nur eine Sonde nötig (siehe Abb 3.1 c). Hier konnten jedoch keine A/T- oder C/G-SNPs untersucht werden, da beide Nukleotide mit dem gleichen Fluoreszenzfarbstoff markiert waren.

3) Allelspezifische Primererweiterung

Nach Hybridisierung der DNA an die Sonden, wurden diese durch eine Polymerase erweitert. Da hierbei dideoxy-Nukleotide (ddATP, ddCTP, ddGTP und ddUTP) verwendet wurden, konnte nur ein Nukleotid eingebaut werden.

4) Signalamplifizierung und Signaldetektion

Um die gemessene Fluoreszenz der SNPs zu verstärken, wurde ein Antikörper Sandwich Assay verwendet (Pinkel et al., 1986). Die verwendeten ddCTPs und ddGTPs waren mit Biotin markiert, ddATPs und ddUTPs mit 2,4-dinitrophenol (DNP; Steemers et al., 2006). Biotin und DNP wurden dann immunohistochemisch mit Alexa555-markiertem Streptavidin beziehungsweise Alexa647-markiertem primären Antikörpern (gewonnen aus Kaninchen) gegen DNP angefärbt (siehe Abb. 3.1 a). Die Fluoreszenzmarkierung wurde mit einem zweiten Färbungsschritt, bei dem biotinyliertes anti-Streptavidin und sekundäre anti-Kaninchen Antikörper (gewonnen aus Ziege) verwendet wurden, abgeschlossen. Um die gemessene Fluoreszenz weiter zu steigern, wurde die Sandwichfärbung zweimal durchgeführt. Nachdem die zweifache Sandwichfärbung abgeschlossen war, wurden die BeadArrays in einem iScan Scanner der Firma Illumina ausgelesen. Die Auswertung der Scandaten erfolgte mit Hilfe der GenomeStudio Software (siehe hierzu Kap. 3.3.1).

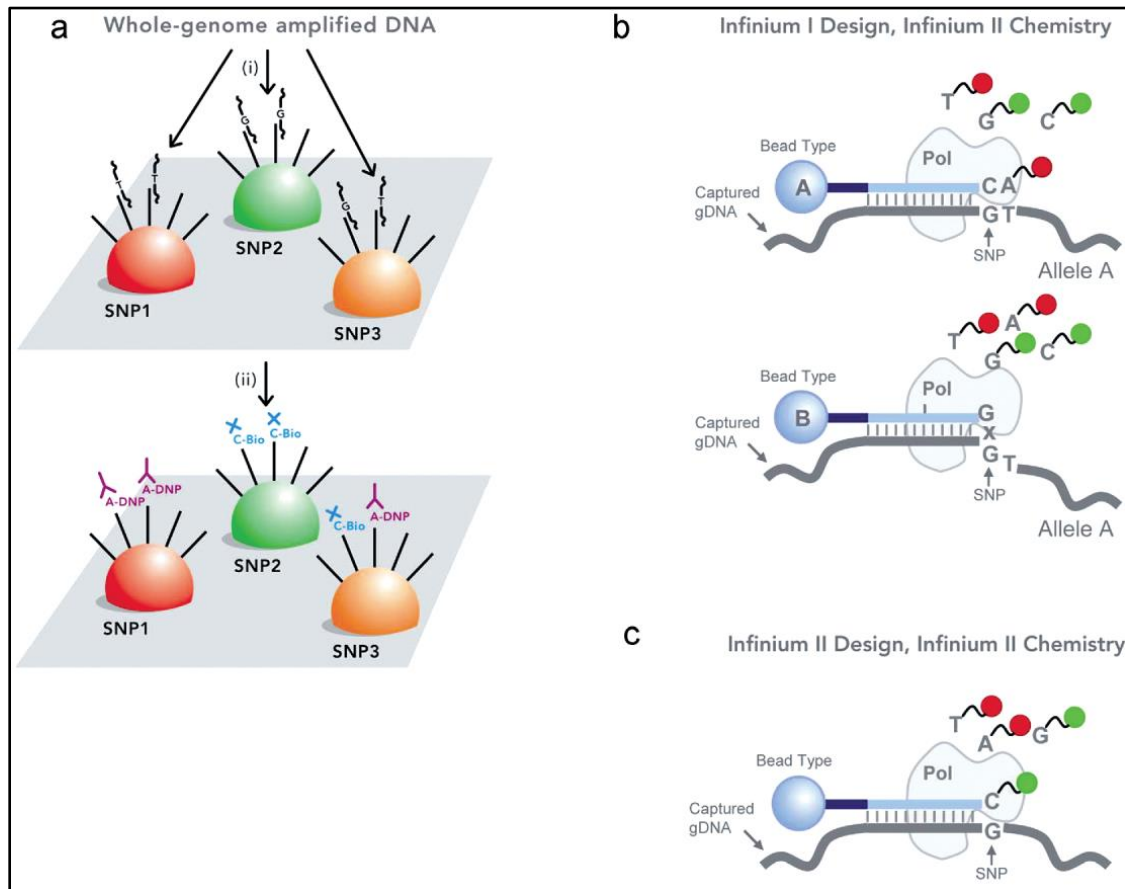


Abb. 3.1: Schematische Darstellung des Infinium Assays. a: Die amplifizierte und fragmentierte DNA bindet an mit spezifischen Sonden gekoppelte *beads* (hier dargestellt durch rote, grüne und orangene Halbkugeln). Anschließend erfolgt eine immunohistochemische Sandwichfärbung. b: Beim Infinium I Design werden SNPs mit zwei verschiedenen Sonden detektiert. Jede der Sonden bindet spezifisch an eines der beiden Allele des SNPs. Nur bei perfektem Übereinstimmen der Sonde mit der Zielsequenz wird eine Fluoreszenz gemessen. c: Für einen Großteil der untersuchten SNPs ist nur ein Sondentyp nötig. Bei diesem Infinium II Design können zwei unterschiedliche Farbstoffe durch eine Einzebasenverlängerung an die Sonde binden (Abbildung entnommen aus: Steemers und Gunderson, 2007).

In der vorliegenden Arbeit wurden drei unterschiedliche Illumina BeadArrays verwendet: (i) HumanHap550v3 (HH550), (ii) Human610-Quadv1 (H610Q) und (iii) Human660W-Quadv1 (H660W). Die Auswahl der SNPs auf diesen Chips wurde von Illumina vorgenommen, in die Selektion der SNPs wurden Daten aus dem HapMap Projekt (siehe Kap 1.1; The International HapMap Consortium, 2007) und vom Wellcome Trust Case Control Consortium (WTCCC; The Wellcome Trust Case Control Consortium, 2007) mit einbezogen. Beim HH550 Chip sind alle SNPs nach dem Infinium II Design ausgewählt worden, beim H610Q Chip und H660W Chip wurden außerdem SNPs mit dem Infinium I Design mit einbezogen. Alle drei Chiptypen zeigten ein hohes Maß an Überlappung bezüglich der Markerauswahl, insgesamt

waren 546.098 gemeinsame Marker auf allen drei Chiptypen vorhanden. Einen Vergleich der wichtigsten Charakteristika der drei Chiptypen zeigt Tabelle 3.1.

Tab. 3.1: Charakteristika der verwendeten BeadArrays

	HH550	H610Q	H660W
Anzahl SNPs	561.466	620.901	657.336
Anzahl der SNPs innerhalb von 10 kb eines <i>RefSeq</i> Gens	252.793	309.978	332.756
Mittelwert des Abstandes zwischen zwei Markern (in kb)	5,3	4,7	4,4
Median des Abstandes zwischen zwei Markern (in kb)	2,9	2,7	2,3

Durchführung:

Für die Prozessierung der Chips wurde das Herstellerprotokoll verwendet (*Infinium HD Assay Super Protocol Guide*, verfügbar über www.illumina.com). Die dabei verwendeten Zusammensetzungen von Puffern und Enzymen war teilweise proprietär. Das Protokoll konnte innerhalb von drei Tagen durchgeführt werden:

Tag 1: Amplifizierung der DNA

Die Amplifizierung benötigte 20 bis 24 Stunden. Sie wurde daher am ersten Tag angesetzt und erfolgte dann über Nacht in einem Hybridisierungssofen bei 37 °C.

Tag 2: Fragmentierung und Fällung der amplifizierten DNA

Nach der Amplifizierung wurde die DNA enzymatisch fragmentiert. Die fragmentierte DNA wurde dann von der Lösung abgeschieden, der Überstand verworfen und die zurück bleibende DNA Ablagerung resuspendiert und weiter verwendet (Dauer: 4 Stunden).

Hybridisierung der DNA auf BeadArrays

Die fragmentierte und resuspendierte DNA wurde nun auf BeadArrays überführt und es folgte die Hybridisierung der DNA-Fragmente an die chipgebundenen *beads*. Diese erfolgte wiederum über Nacht im Hybridisierungssofen bei 48 °C (mindestens 16 Stunden, höchstens 24 Stunden).

Tag 3: Waschen der BeadArrays

Nun wurden die BeadArrays für die Färbung vorbereitet. Dazu wurden die Chips zunächst gewaschen und in Durchfluskkammern überführt (Dauer: 30 Minuten).

Einzelbasenverlängerung und Färbung der BeadArrays

Im letzten Schritt vor dem Scannen wurde die Einzelbasenerweiterung durchgeführt. Im Anschluss wurde die hybridisierte DNA mit Formamid entfernt. Dies gewährleistete, dass nur die an die *beads* gekoppelten Sonden im abschließenden Schritt immunhistochemisch angefärbt wurden (Dauer: 1 Stunde).

Scannen der BeadArrays

Schließlich wurden die fertig prozessierten Chips mit Hilfe des iScan Scanners von Illumina ausgelesen (Scanzeit je Chip ca. 90 Minuten).

3.2.4 Verifikation von *in silico* CNVs

Für die Bestätigung der *in silico* detektierten CNVs wurden drei quantitative Methoden verwendet: multiplex Ligation-abhängige Sondenamplifikation (MLPA, *multiplex ligation-dependent probe amplification*), quantitative Echtzeit-PCR (qRT-PCR, *quantitative real-time PCR*) mit TaqMan-Sonden und qRT-PCR mit SYBR-Green. Mit diesen quantitativen Methoden konnte die exakte DNA-Kopienzahl einer Person bestimmt werden.

3.2.5 *Multiplex ligation-dependent probe amplification*

Die Quantifizierung von DNA-Abschnitten durch MLPA (MRC-Holland, Amsterdam, Niederlande) beruht auf einer Amplifikation von Sonden, die spezifisch für die zu untersuchenden Regionen sind (Schouten et al., 2002).

Jede der sequenzspezifischen Sonden besteht aus zwei Teilen: einem kürzeren Teil, der zwischen 40 und 49 Nukleotide lang ist, und einem längerem Teil mit 80 bis 420 Nukleotide. Beide Teile der Sonde haben einen Sequenz der bei allen kurzen (19 Nukleotide lang) beziehungsweise langen (36 Nukleotide lang) Teilen identisch ist und einem Teil, der spezifisch für die zu untersuchenden Region ist (Schouten et al., 2002). Der lange Teil der Sonde hat außerdem zwischen 19 und 341 Nukleotide, die für spezifisch für die Sonde

sind und durch die diese später identifiziert werden kann (Schouten et al., 2002).

Beide Teile der Sonde wurden so ausgewählt, dass sie mit nur wenigen Nukleotiden Abstand an die zu untersuchende DNA Sequenz hybridisieren. Nach der Hybridisierung wurden die beiden Teile der Sonde durch ein Ligase verbunden. In der nachfolgenden PCR wurden alle verwendeten Sonden amplifiziert. Durch die unterschiedliche Länge der Sonden konnten die PCR-Produkte dann durch Kapillarelektrophorese aufgetrennt werden. Die Höhe der so gemessenen Signale wurde für die Bestimmung der Kopienzahl verwendet.

Durchführung:

Entsprechend den Herstellervorgaben (<http://www.mlpa.com>, Schouten et al., 2002) wurde die MLPA Reaktion in fünf Arbeitsschritte unterteilt (siehe Abb. 3.2):

1) Denaturierung

Die Proben-DNA wurde eine Minute bei 95 °C denaturiert.

2) Hybridisierung

Die denaturierte DNA und beide Teile der Sonde wurden 16 Stunden bei 60 °C inkubiert. In dieser Zeit konnten sich die Sonden an die Zielsequenz anlagern.

3) Ligation

Die beiden Teile der Sonde wurden durch eine Ligase verbunden. Hierzu wurde die Probe 15 Minuten bei 54 °C mit 40 µl Ligationspuffer (bestehend aus 2,6 mM MgCl₂, 5 mM Tris-HCl, 0,2 mM NAD und 1 U Ligase-65) inkubiert. Anschließend wurde die Ligase inaktiviert (5 Minuten bei 98 °C).

4) PCR

Die Sonden wurden nun amplifiziert. Hierzu wurden 10 µl aus dem vorangegangenen Schritt zu 30 µl PCR-Puffer (mit 10 pmol PCR-Primer, 2,5 nmol dNTPs und 2,5 U Taq/SALSA Polymerase) hinzugefügt. Die PCR-Primer waren für alle untersuchten Sonden die gleichen und hybridisierten an die 19 beziehungsweise 36 Nukleotid langen Teile der Sonden, dadurch konnten alle verwendeten Sonden mit den gleichen Primern amplifiziert werden. Es wurde ein Standard-PCR-Programm verwendet:

Temperatur	Zeit	33 Zyklen
95 °C	30 Sek.	
60 °C	30 Sek.	
72 °C	60 Sek.	

5) Auftrennung der Produkte

Die PCR-Produkte wurden durch Kapillarelektrophorese (3130xl Genetic Analyzer, Life Technologies, Carlsbad, Kalifornien, USA) aufgetrennt. Da die einzelnen Sonden unterschiedliche Längen hatten, erzeugte jedes Amplikon ein spezifisches Signal. Durch Vergleich dieses Signals mit dem Signal einer Probe deren Kopienzahl bekannt war, ließ sich die Kopienzahl der untersuchten Probe bestimmen.

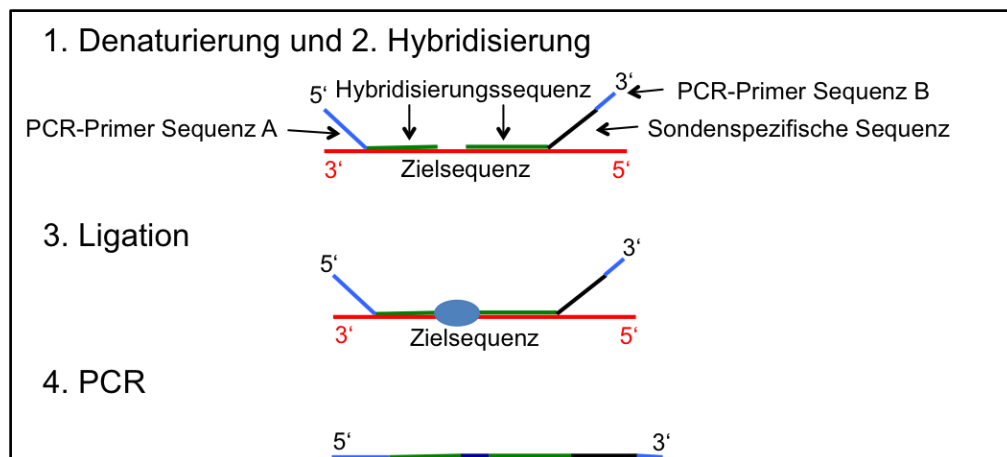


Abb. 3.2: Schematischer Ablauf der MLPA. Die MLPA erfolgte in fünf Schritten: 1) Die DNA wurde zunächst denaturiert; 2) an die einzelsträngige DNA der Zielregion (rote Balken) hybridisierten die beiden Teile der Sonde mit gleichbleibenden (hellblaue Balken), sequenzspezifischen (grüne Balken) und sondenspezifischen (dunkelblaue Balken) Abschnitten; 3) die beiden Teile der Sonde wurden durch ein Ligase verbunden; 4) alle verwendeten Sonden wurden durch eine PCR amplifiziert; 5) die PCR-Produkte wurden durch Kapillarelektrophorese aufgetrennt (nicht dargestellt, Abbildung modifiziert nach Schouten et al., 2002).

3.2.6 Quantitative real-time PCR mit TaqMan-Sonden

Bei der qRT-PCR wird nach jedem PCR-Zyklus eine Fluoreszenz-Messung durchgeführt. Bei der qRT-PCR mit TaqMan-Sonden werden für jede zu untersuchende Region spezifische Sonden eingesetzt. Die Quantifizierung der DNA erfolgt mit Hilfe des Fluoreszenz-Resonanzenergietransfer (FRET). Beim FRET wird die Fluoreszenzenergie eines angeregten Farbstoffes, dem Reporter, von einem weiteren Farbstoff, dem Quencher, aufgenommen (Förster, 1948). Der Quencher emittiert Licht in einem anderen Spektrum als

der Reporter. Der FRET kann allerdings nur über einen kurzen Abstand erfolgen, liegen Reporter und Quencher zu weit auseinander emittiert der Reporter und nicht der Quencher Licht. Die TaqMan-Sonden tragen einen Reporter an ihrem 5'-Ende und einen Quencher am 3'-Ende (siehe Abb. 3.3). Mit diesen Sonden wird eine PCR durchgeführt. Bei dieser PCR wird eine Polymerase verwendet, die zusätzlich über eine 5'-3'-Exonuklease-Funktionalität verfügt, dadurch werden Reporter und Quencher voneinander getrennt und es kann keine FRET mehr stattfinden (siehe Abb. 3.3). Nach jedem PCR Zyklus erfolgt eine Messung der Fluoreszenz im Spektrum des Reporters, diese steigt proportional zur Menge des vorhandenen PCR-Produkts an. Der Anstieg der Fluoreszenz verläuft zunächst linear und tritt dann in eine exponentielle Phase über. Der Zeitpunkt des Übertritts in diese exponentielle Phase ist dabei abhängig von der zu Beginn der Reaktion vorhandenen DNA-Menge einer Probe. Dieser Zeitpunkt ist der Schwellenwert (*cycle threshold*, Ct), der für die Berechnung der initial vorhandenen Kopienzahl verwendet wird. Für diese Berechnung (siehe Kap. 3.2.7) wird der Ct wird der untersuchten Region mit einer Referenz verglichen.

Die von der Firma Applied BioSystems (Foster City, Kalifornien, USA) entwickelten TaqMan-Sonden tragen zwei Fluorophore (Heid, et al., 1996). Der Reporter ist mit den Fluorophoren VIC oder FAM markiert, als Quencher dient TAMRA (siehe Abb. 3.3).

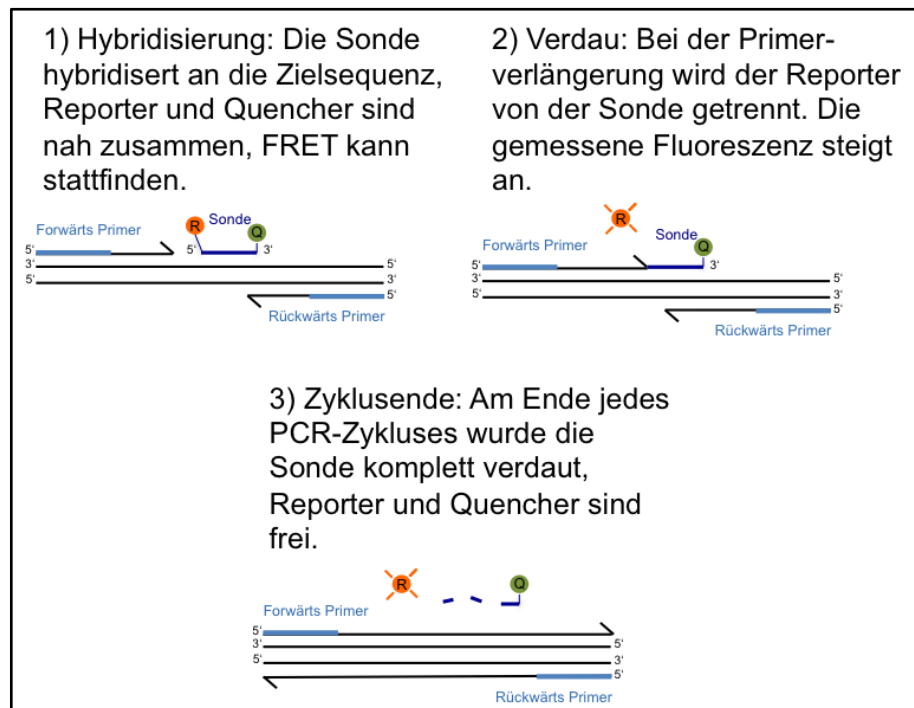


Abb. 3.3: Schematischer Ablauf der qRT-PCR mit TaqMan-Sonde. FRET – Fluoreszenz-Resonanzenergietransfer, R – Reporter, Q – Quencher (modifiziert nach <http://www.appliedbiosystems.com>)

Durchfuhrung:

Die Verifikation der *in silico* detektierten CNVs durch qRT-PCR mit TaqMan Sonden wurde gema der Herstellerangaben durchgefuhrt (siehe hierzu: TaqMan Copy Number Assays (PN 4397425)). Es wurden TaqMan Copy Number Assays (FAM markiert) von Applied BioSystems beziehungsweise TaqMan Copy Number Reference Assay RNase P (VIC markiert) verwendet. Alle Reaktionen wurden in Mikrotiterplatten mit 384 Vertiefungen (*wells*) durchgefuhrt, jede Probe wurde als Triplikat untersucht. Als Referenzassay wurde in allen Versuchen *RNaseP* verwendet. Das Protokoll besteht aus vier Schritten:

1) Vorbereitung eines Mastermixes

Ein Mastermix wurde vorbereitet, je *well* wurde benotigt:

TaqMan Genotyping Master Mix	5,0 μ l
TaqMan Copy Number Assay	0,5 μ l
TaqMan Copy Number Reference Assay	0,5 μ l
Nuklease freies Wasser	2,0 μ l
Gesamtvolumen	8,0 μ l

2) Hinzufügen der DNA

Je *well* wurden 8 μl des Mastermixes vorgelegt und 2 μl einer vorbereiteten DNA Lösung (5 ng/ μl) hinzugegeben.

3) PCR Reaktion

Die Reaktion lief in einem *real-time* PCR Gerät (7900HT, Applied BioSystems Foster City, Kalifornien, USA) mit diesen Parametern ab:

Temperatur	Zeit	
95 °C	10 Min.	
95 °C	15 Sek.	40 Zyklen
60 °C	60 Sek.	

4) Analyse der Ergebnisse

Im Anschluss an die PCR wurden die Ergebnisse exportiert und mit der CopyCaller Software analysiert (siehe hierzu Kapitel 3.2.7).

3.2.7 CopyCaller

Wie in Kapitel 3.2.6 beschrieben wurde die Auswertung der qRT-PCR Experimente mit der CopyCaller Software der Firma Applied Biosystems durchgeführt. Die Auswertung wurde entsprechend der Herstellerangaben (CopyCaller Software v1.0 User Guide) vorgenommen. Zunächst wurden die gemessenen Ct-Werte exportiert. Diese Werte geben den Zeitpunkt an, bei dem die PCR in die exponentielle Phase übergegangen ist (siehe Kap. 3.2.6). Bei der qRT-PCR mit TaqMan-Sonden wurde in jedem *well* zwei Assays analysiert; eines für die Zielregion, deren Kopienzahl ermittelt werden sollte und ein Referenzassay. In der vorliegenden Arbeit wurde *RNaseP* als Referenzassay verwendet. Die exportierten Ct-Werte wurden in die CopyCaller Software importiert. Diese berechnete dann anhand der Ct-Werte die Kopienzahl für jede Person. Für alle qRT-PCR Experimente wurden Kontrollpersonen verwendet, die zwei Kopien der untersuchten Regionen hatten.

Die CopyCaller Software berechnete die Kopienzahl in drei Schritten:

1) ΔCt -Berechnung

Für jedes *well* wurde zunächst der ΔCt -Wert berechnet. Dieser gab den Unterschied zwischen dem Referenzassay und dem Zielassay an:

$$(\Delta Ct)_w = (Ct)_{t,w} - (Ct)_{r,w}$$

mit $w = well$, $t = \text{Zielassay (target)}$, $r = \text{Referenzassay}$

Der ΔCt -Wert wurde zunächst für alle Replikate eines Zielassays ermittelt, für die weiteren Schritte wurde dann das arithmetische Mittel aller Replikate eines Zielassays $\mu(\Delta Ct)$ verwendet.

2) $\Delta\Delta Ct$ -Berechnung

Die Differenz zwischen dem $\mu(\Delta Ct)$ -Wertes des Zielassays einer betroffenen Person zu dem $\mu(\Delta Ct)$ -Wertes des gleichen Zielassays der Kontrollpersonen war der $\Delta\Delta Ct$ -Wert:

$$(\Delta\Delta Ct)_{s,t} = \mu(Ct)_{s,t} - \mu(Ct)_{k,t}$$

mit $s = \text{betroffene Person (sample)}$, $t = \text{Zielassay (target)}$, $k = \text{Kontrollperson}$

Der Vergleichswert $\mu(\Delta Ct)_{k,t}$ wurde über alle verwendeten Kontrollpersonen ermittelt.

3) Ermittlung der Kopienzahl

Mit dem so berechneten $\Delta\Delta Ct$ -Wert konnte schließlich die Kopienzahl der untersuchten Person für den Zielassay berechnet werden:

$$\text{Kopienzahl} = cn_k * 2^{-\Delta\Delta Ct}$$

mit $cn_k = \text{Kopienzahl der Kontrollperson (entsprach 2)}$

3.2.8 Quantitative real-time PCR mit SYBR-Green

Ähnlich wie bei der qRT-PCR mit TaqMan-Sonden wird bei der qRT-PCR mit SYBR-Green nach jedem durchgeführten PCR-Zyklus die emittierte Fluoreszenz gemessen. Auch bei dieser Methode erfolgt die Bestimmung der ursprünglichen Kopienzahl eines untersuchten DNA-Abschnittes durch den Vergleich der Ct-Werte mit einer Referenz. Jedoch wird bei dieser Methode das Cyanin SYBR-Green als Fluoreszenzfarbstoff verwendet.

SYBR-Green bindet an doppelsträngige DNA und emittiert dann Licht im grünen Spektrum (Ponchel et al., 2003). Zwar kann SYBR-Green auch an einzelsträngige DNA binden, dann ist das Fluoreszenzsignal aber deutlich geringer. Diese Eigenschaft von SYBR-Green wird für die qRT-PCR genutzt. Da sich bei jedem PCR-Zyklus die Menge an doppelsträngiger DNA erhöht, erhöht sich im gleichen Maße auch die gemessene Intensität.

Für jede zu untersuchende Region werden spezifische Primer hergestellt, die eine ausschließliche Amplifikation dieser Region gewährleisten. Da SYBR-Green unspezifisch an doppelsträngige DNA bindet würden eventuell entstehende Nebenprodukte das gemessene Signal verfälschen. Wie bei der qRT-PCR mit TaqMan-Sonden wird der Zeitpunkt C_t bestimmt, bei dem die gemessene Fluoreszenzintensität in die exponentielle Phase übergeht (siehe Kap. 3.2.6). Durch die gezielte Amplifikation der Zielregion und dem Vergleich zu Referenzregionen kann so die Kopienzahl der Zielregion bestimmt werden. Als interne Referenzen dienten Regionen in den Genen *BNC1* (*basonuclin 1*), *CFTR* (*cystic fibrosis transmembrane conductance regulator*) und *RPP38* (*ribonuclease P/MRP 38kDa subunit*, Primersequenzen jeweils im Anhang). Dabei handelt es sich um ubiquitär exprimierte Gene, in denen die untersuchten Personen zwei Kopien aufwiesen. Alle Verifikationsexperimente mit SYBR-Green wurden als Triplikate durchgeführt. Diese internen werden für die Normalisierung der Messwerte verwendet.

Es wurde ein PCR-Programm für alle Reaktionen verwendet:

Temperatur	Zeit	
95 °C	10 Min.	
95 °C	15 Sek.	40 Zyklen
60 °C	60 Sek.	

Die Reaktionen erfolgten in einem *real-time* PCR Gerät (7900HT, Applied Biosystems Foster City, Kalifornien, USA). Nach abgeschlossener Reaktion wurden die C_t -Werte exportiert. Die Kopienzahl wurde dann mit der ΔC_t -Methode ermittelt:

$$\Delta Ct = Ct (\text{Ziel}) - Ct (\text{Referenz})$$

$$\text{Kopienzahl} = 2 * 2^{-\Delta Ct}$$

3.2.9 Auswahl der Verifikationsmethoden

Für die in Kapitel 4 beschriebenen CNVs wurden unterschiedliche Verifikationsmethoden gewählt. Die Auswahl der Verifikationsmethode war abhängig von der zur Verfügung stehenden DNA-Menge, den Kosten der Methode und der Verfügbarkeit von Sonden in der zu verifizierenden chromosomalen Region. Sonden für eine Verifikation mittels qRT-PCR mit TaqMan-Sonden oder SYBR-Green können für jede Region spezifisch hergestellt werden. Die Sondauswahl für eine Verifikation mittels MLPA erfolgt in der Regel durch den Hersteller. Individualisierte Sonden können zwar hergestellt werden, dies ist jedoch aufwändig und mit hohen Kosten verbunden.

Bei der Entscheidung zwischen einer Verifikation mit qRT-PCR mit TaqMan-Sonden oder SYBR-Green sind die Kosten und Menge an verfügbarer DNA wichtig. Die qRT-PCR mit SYBR-Green ist wesentlich günstiger als mit TaqMan-Sonden, allerdings wird für jedes Experiment deutlich mehr DNA (~200 ng) benötigt als bei der qRT-PCR mit TaqMan-Sonden (~10 ng).

1) qRT-PCR mit TaqMan-Sonden

Alle CNVs in den beiden mit der Bipolaren Störung assoziierten Regionen (siehe Kap 4.2.3) wurden mit qRT-PCR und jeweils drei TaqMan-Sonden (siehe Kap. 3.2.6) verifiziert. Für Duplikationen in 6q27 wurden die Assays Hs02902865_cn, Hs04311265_cn und Hs03589363_cn verwendet, für Duplikationen in 10q11.22 die Assays Hs03749977_cn, Hs03744366_cn und Hs03283591_cn. Als Referenz wurde jeweils *RNaseP* verwendet.

2) qRT-PCR mit SYBR-Green-Sonden

Deletionen in den chromosomalen Subbanden 7p21.3 und 18p11.32, die mit Majorer Depression assoziiert waren (siehe Kap. 4.3.3), wurden mit jeweils drei SYBR-Green Sonden verifiziert (Sequenzen: siehe Anhang). Für die Verifikation der Duplikationen in der chromosomalen Subbande 15q26.3 (siehe Kap. 4.3.3) mit qRT-PCR wurden vier SYBR-Green Sonden gewählt (Sequenzen siehe Anhang).

3) MLPA

Für die Verifikation der CNVs in 16p11.2 (siehe Kap. 4.3.3) wurde das MLPA Kit P297 verwendet, dieses hatte fünf Sonden innerhalb der assoziierten Region (siehe Anhang).

3.3 Bioinformatische und statistische Methoden

3.3.1 GenomeStudio

GenomeStudio ist ein von der Firma Illumina entwickeltes Softwarepaket, mit dem die durch BeadArray Technologie (siehe Kap. 3.2.3) generierten genetischen Daten ausgewertet wurden. Im Folgenden wird das Genotypisierungsmodul näher erklärt. Insbesondere werden die für die Identifikation von CNVs relevanten Aspekte erörtert werden.

Grundlage für die Analysen waren die gemessenen Fluoreszenzintensitäten der beiden SNP Allele (*raw X* bzw. *raw Y*). GenomeStudio unterscheidet bei einem SNP nur zwischen A- und B-Allel. Beim Infinium I Design (siehe Kap. 3.2.3) entsprach das A-Allel dem häufiger in der europäischen Population auftretenden Allel. Für SNPs, die mit dem Infinium II Design (siehe Kap. 3.2.3) detektiert wurden, entsprach das A-Allel den rot markierten Nukleotiden A und T, das B-Allel den grün markierten Nukleotiden G und C. *Raw X* gibt die Intensität des A-Allels an, *raw Y* die des B-Allels. Diese Rohwerte wurden in fünf Schritten normalisiert (Peiffer et al., 2006):

1. Entfernung von Ausreißern
2. Bestimmung des Hintergrundrauschens
3. Bestimmung der Winkelabweichung
4. Bestimmung der Scherung
5. Einheitliche Skalierung der Daten

Die normalisierten Werte (X_{norm} bzw. Y_{norm}) waren wiederum Grundlage für die Berechnung verschiedener Größen, die für die Interpretation der Genotypisierungsdaten wichtig sind. Für die CNV-Analyse sind *B allele frequency* (BAF) und *Log R Ratio* (LRR) die relevanten Größen. Beide Werte wurden jeweils für jeden SNP einer Person berechnet.

B allele frequency

Die BAF beschreibt den Genotyp eines SNPs einer Person. Nach der Normalisierung der Rohdaten erfolgt eine Auftragung in einem polarem Koordinatensystem (Peiffer et al., 2006). Dabei wurden die Werte Theta und R jedes SNPs aufgetragen. Theta ist die Winkelabweichung eines Signals vom reinen A-Allel-Signal. Ein SNP, der homozygot für das A-Allel ist zeigt ein Theta von 0, bei einem homozygoten B-Allel war Theta gleich 1.

$$\text{Theta} = \left(\frac{2}{\pi}\right) * \arctan\left(\frac{Y_{norm}}{X_{norm}}\right)$$

R ist die Summe der normalisierten Intensitätswerte und gibt die Intensitätsstärke an.

$$R = X_{norm} + Y_{norm}$$

Bei der Auftragung im polaren Koordinatensystem kann man für jeden SNP die Allelverteilung und die Intensitätsstärke ablesen.

Die BAF für einen SNP ist der Wert Theta dieses SNP, korrigiert für die Position der Probe innerhalb des Genotyp-Clusters. Proben mit gleichem Genotyp, also entweder AA, AB oder BB, werden in Gruppen, den Clustern aufgetragen. Die BAF gab also die allelische Zusammensetzung an, sie ist ein kontinuierlicher Wert zwischen 0 und 1. Dabei hat der Genotyp AA eine BAF von 0, der Genotyp AB eine BAF von 0,5 und BB einen Wert von 1. Die experimentell ermittelten Werte konnten leicht von diesen theoretischen Werten abweichen. Im Gegensatz zu Theta zeigte die BAF weniger Variation.

Log R Ratio

Das LRR ergibt sich aus dem Verhältnis der gemessenen Intensität für einen bei der untersuchten Person ($R_{subject}$) zu einer Referenz ($R_{reference}$, Peiffer et al., 2006). In der vorliegenden Arbeit wurde die Illuminareferenz verwendet, welche von Illumina für alle Chips zur Verfügung gestellt wird.

$$LRR = \log_2\left(\frac{R_{subject}}{R_{reference}}\right)$$

Wenn die Intensität einer Probe der Referenz entspricht, so ist das LRR 0, zeigt die Probe eine erhöhte Intensität, so ist das LRR größer als 0, ist die

Intensität der Probe im Vergleich zur Referenz kleiner ergibt sich ein negatives LRR. Auch hier führten experimentell Schwankungen zu leichten Abweichungen von diesen theoretischen Werten.

3.3.2 Detektion von CNVs in GenomeStudio

Wenn keine Veränderung der Kopienzahl (d.h. zwei Kopien) vorliegt, streut das LRR aller SNPs einer Person um 0, die BAF streut in drei Clustern um 0, 0,5 und 1. Liegt jedoch ein CNV vor, zeigen sich charakteristische Veränderungen in LRR und BAF. Diese lassen sich mit GenomeStudio graphisch darstellen.

Bei einer Deletion (Kopienzahl 0 oder 1) ist die Zahl der Allele in der betroffenen chromosomalen Region bei der untersuchten Person reduziert. Dadurch verringert sich auch die gemessene Intensität der analysierten SNPs und das LRR ist kleiner als 0 (siehe Kap. 3.3.1). Die BAF ist bei heterozygoten Deletionen (Kopienzahl 1) entweder 0 oder 1 (siehe Abb. 3.4). Bei homozygoten Deletionen (Kopienzahl 0) dürfte es theoretisch kein Signal für die BAF geben. Da jedoch die SNP-spezifischen *beads* (siehe Kap. 3.2.3) immer noch vorhanden sind und ein Signal erzeugen, entstehen Artefakte. Homozygote Deletionen haben ein stark negatives LRR. Die BAF der SNPs verteilt sich zufällig zwischen 0 und 1, ohne erkennbare Cluster zu bilden (siehe Abb. 3.4).

Bei Amplifikationen (Kopienzahl 3 und größer) ist die Zahl der Allele erhöht. Die gemessene Intensität der betroffenen SNPs steigt somit an und das LRR wird größer als 1 (siehe Kap. 3.3.1). Die Verteilung der BAF hängt von der Art der Amplifikation ab. Bei einer Duplikation, bei der nur eines der beiden Allele dupliziert wurde (Kopienzahl 3), erwartet man theoretisch vier Genotypen: AAA, AAB, ABB oder BBB. In diesem Fall erwartet man für die betroffenen SNPs eine BAF von 0 (AAA), $\frac{1}{3}$ (AAB), $\frac{2}{3}$ (ABB) oder 1 (BBB; siehe Abb. 3.5). Bei komplexeren Amplifikationen steigt die Zahl der möglichen Genotypen und damit die zu erwartenden BAFs. Dabei ist die theoretische Zahl der BAF-Cluster um eins größer als die tatsächliche Kopienzahl.

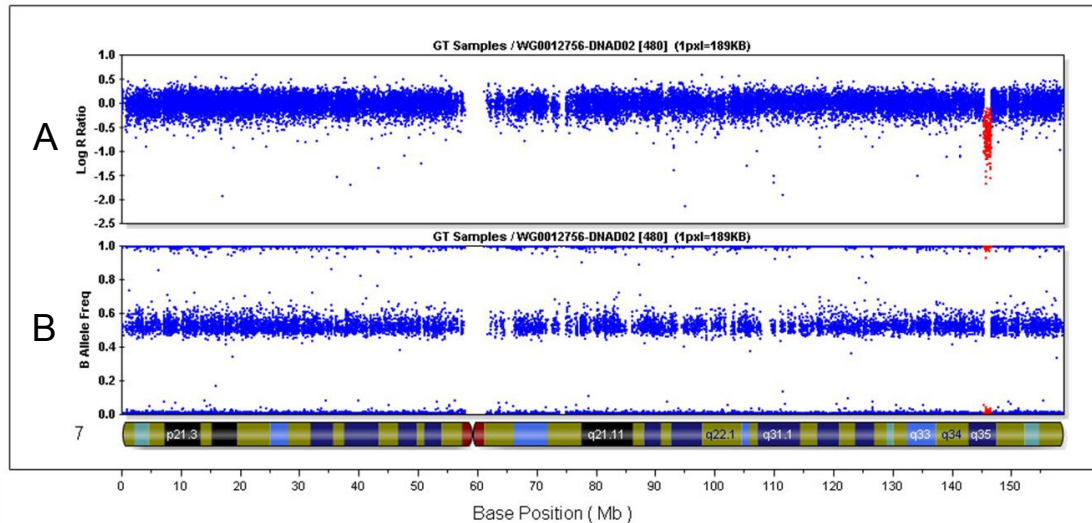


Abb. 3.4: Beispiel für eine Deletion in GenomeStudio. Jeder blaue und rote Punkt repräsentiert einen SNP. Das *Log R Ratio* (A) und die *B Allele Frequency* (B) aller SNPs auf Chromosom 7 einer Person sind auf der Ordinate aufgetragen. Auf der Abszisse ist die chromosomale Position der Marker aufgetragen. Man erkennt deutlich ein LRR kleiner als 0 und zwei BAF-Cluster bei 0 und 1 für die roten SNPs bei circa 145 Mb. An dieser Stelle liegt eine heterozygote Deletion vor.

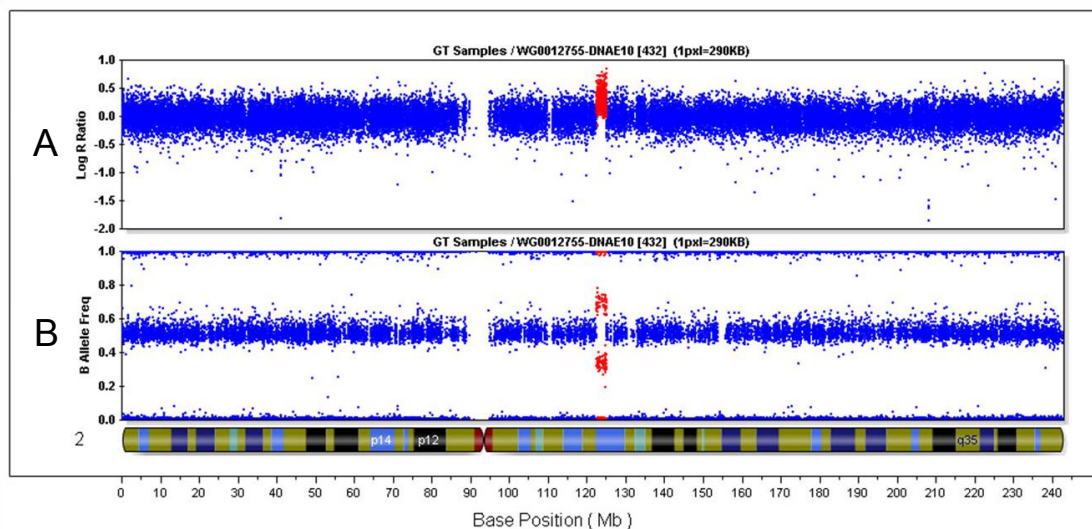


Abb. 3.5: Beispiel für eine Duplikation in GenomeStudio. Jeder blaue und rote Punkt repräsentiert einen SNP. Das *Log R Ratio* (A) und die *B Allele Frequency* (B) aller SNPs auf Chromosom 2 einer Person sind auf der Ordinate aufgetragen. Auf der Abszisse ist die chromosomale Position der Marker aufgetragen. Man erkennt deutlich ein LRR größer als 0 und vier BAF-Cluster bei circa 0, 0,33, 0,66 und 1 für die roten SNPs bei circa 125 Mb. An dieser Stelle liegt eine Duplikation vor.

3.3.3 Genomweite *in silico* Detektion von CNVs

Die visuelle Inspektion und Beurteilung von CNVs anhand der LRR und BAF (siehe Kap. 3.3.2) funktioniert problemlos für einzelne CNVs. Da in der vorliegenden Arbeit jedoch große Kollektive (mehr als 1.500 Personen mit etwa 0,8 CNVs pro Person) untersucht wurden, wäre eine manuelle Überprüfung

jedes einzelnen CNVs auf diese Weise zeitlich nicht effizient durchführbar gewesen. Wesentlich effektiver für die Analyse genomweiter CNV-Datensätze ist dagegen die Verwendung von automatisierten Computerprogrammen. Diese können, mit Hilfe verschiedener Algorithmen, CNVs *in silico* detektieren. Derzeit sind zahlreiche, unterschiedliche Programme verfügbar. Vergleichende Studien haben gezeigt, dass die verfügbaren Programme in identischen Datensätzen teilweise abweichende CNVs identifizieren (Winchester et al., 2009; Dellinger et al., 2010). Die Programme QuantiSNP (<http://www.well.ox.ac.uk/QuantiSNP/>, Colella et al., 2007) und PennCNV (<http://www.openbioinformatics.org/penncnv/>, Wang et al., 2007) gehörten zu den Programmen, die in diesen Studien am besten abschnitten (Winchester et al., 2009; Dellinger et al., 2010). Zur Reduzierung falsch positiver CNVs wurden in der vorliegenden Arbeit diese beiden Programme verwendet. Beide Programme detektierten CNVs aus SNP-Genotypisierungsdaten mit Hilfe eines Hidden Markov Models (HMM). Die Implementierung des HMMs unterschied sich in beiden Programmen.

Hidden Markov Model

Beim HMM handelt es sich um ein stochastisches Model mit dessen Hilfe man die unbekannt Zustände eines Systems bestimmen kann (Rabiner, 1989). Grundlage für das HMM sind Markovketten, welche aus verschiedenen Zuständen (X_1 und X_2 in Abb. 3.6) und den dazu korrespondierenden Wahrscheinlichkeiten der Übergänge zwischen den Zuständen (A_{12} und A_{21} in Abb. 3.6) bestehen. Beim HMM sind die einzelnen Zustände der Markovkette von außen nicht zu erkennen (die sogenannten *hidden states*), sondern es können nur die Ausgaben (Y_1 und Y_2 in Abb. 3.6) erkannt werden. Diese Ausgaben entstehen nur mit einer gewissen Wahrscheinlichkeit mit der ein Zustand ein bestimmtes Signal emittiert (die Emissionswahrscheinlichkeit, B_1 und B_2 in Abb. 3.6) abgebildet. Anstelle der einzelnen Zustände werden beim HMM also die Ereignisse, die durch den Zustand mit einer bestimmten Emissionswahrscheinlichkeit entstehen, beobachtet (siehe Abb. 3.6). Diese Beobachtung lässt dann wiederum Schlüsse auf die eigentlichen Zustände des HMMs zu.

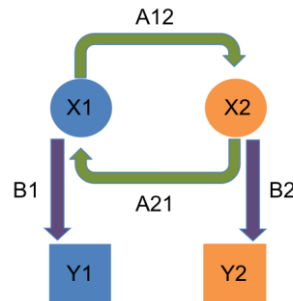


Abb. 3.6: Schematische Darstellung einer einfachen Markovkette. Die verborgenen Zustände des Markov Modells (X_1 , X_2) können nicht direkt erkannt werden, man beobachtet anstatt dessen die Ausgabe (Y_1 , Y_2), die mit einer Emissionswahrscheinlichkeit (B_1 , B_2) entsteht. Jeder Zustand X hat eine Übergangswahrscheinlichkeit (A_{12} , A_{21}) in einen anderen Zustand zu wechseln. Modifiziert nach Rabiner, 1989.

Sowohl bei QuantiSNP als auch bei PennCNV entsprechen die *in silico* detektierten Kopienzahlen den verborgenen Zuständen.

Tab. 3.2: Die verborgenen Zustände der Markovketten von QuantiSNP und PennCNV und die entsprechende Kopienzahl (nach Colella et al., 2007).

Verborgener Zustand	Beschreibung	Kopienzahl
1	Homozygote Deletion	0
2	Heterozygote Deletion	1
3	Kein CNV (homozygot)	2
4	Kein CNV (heterozygot)	2
5	Einfache Duplikation	3
6	Komplexere Amplifikation	≥ 4

Bei den beiden verborgenen Zuständen 3 und 4 ist die Kopienzahl jeweils zwei. Bei Zustand 3 findet sich auf jedem der beiden homologen Chromosomen eine Kopie des DNA-Segments. Im Gegensatz dazu befinden sich bei Zustand 4 beide Kopien nur einem Chromosom. Eine solche Situation ist beispielsweise denkbar, wenn auf dem einen Chromosom eine Deletion vorliegt, während die gleiche Region auf dem anderen Chromosom dupliziert wurde.

3.3.4 Qualitätskontrolle der untersuchten Personen und SNPs

Grundlage für die Identifikation von CNVs waren die genomweiten Genotypisierungsdaten. Wenn diese Daten bereits fehlerhaft sind, hat dies Auswirkungen auf die Ergebnisse. Insbesondere falsch-positive CNVs stellen ein Problem dar. Hierbei handelte es sich um CNVs, die von QuantiSNP und

PennCNV identifiziert wurden, allerdings nicht *in vivo* nachweisbar waren. Daher ist eine gründliche Qualitätskontrolle der Daten mit stringenten Filterkriterien absolut notwendig. Diese Filterkriterien können auf drei unterschiedlichen Ebenen greifen: (i) Personen, (ii) SNPs und (iii) CNVs (siehe hierzu Kap. 3.3.6).

SNP spezifische Filterkriterien

In der vorliegenden Arbeit wurden nur solche SNPs für die CNV Detektion verwendet, die in mindestens 98 % der untersuchten Personen einen Genotyp zugeordnet bekamen.

Personen spezifische Filterkriterien

Die Genotypisierungsdaten einzelner Personen wurden aus dem Datensatz entfernt, wenn eines der folgenden Kriterien erfüllt war:

- 1) Das phänotypische Geschlecht stimmte nicht mit dem genotypischen überein.
- 2) Bei der Person handelte es sich um eine Doublette mit einer anderen Person der Studie, in diesem Fall wurde die Person mit weniger zugeordneten Genotypen ausgeschlossen.
- 3) Zwei Personen waren nicht näher miteinander verwandt (nicht näher als Onkel und Nefte). Die Person mit weniger zugeordneten Genotypen wurde ausgeschlossen.
- 4) Weniger als 97 % der untersuchten SNPs wurde ein Genotyp zugeordnet.
- 5) Die Genotypzusammenstellung der Person wich zu stark von dem der Mitteleuropäer im HapMap Phase 2 Datensatz ab. Diese Populationsausreißer wurden mit multidimensionaler Skalierung detektiert.
- 6) Bezogen auf die Standardabweichung der LRR über alle SNPs war die Person ein extremer Ausreißer.

Extreme Ausreißer sind Personen, deren Werte, bezogen auf den jeweiligen Parameter größer als ein Schwellenwert war:

$$\text{Schwelle} = 75 \% \text{ Quantil} + 3 * \text{Interquartilabstand}$$

3.3.5 Durchführung und Ausgabe von QuantiSNP/PennCNV

QuantiSNP (Colella et al., 2007) und PennCNV (Wang et al., 2007) benötigen für die Identifikation von CNVs das LRR, die BAF der SNPs und deren chromosomale Position. Diese Werte wurden durch GenomeStudio berechnet (siehe Kap 3.3.1). Es wurde eine Ausgabedatei erstellt, die für die späteren Analysen wichtigen Parameter jedes SNPs von jeder Person enthielt:

SampleID: Identifikationsnummer der Person

Sex: Genotypisches Geschlecht der Person

Name: rsID des SNPs

Chromosome: Chromosom auf dem sich der SNP befindet

Position: exakte genomische Position des SNPs auf dem Chromosom

Log R ratio: siehe 3.3.1

B allele frequency: siehe 3.3.1

In einem *final report* waren die Daten mehrerer Personen enthalten. Beide Programme analysierten die Daten jeder Person jedoch einzeln. Daher musste die erzeugte Textdatei in einzelne, individuen-spezifische Textdateien unterteilt werden. Für QuantiSNP wurde hierzu in unserem Institut ein Skript (siehe Anhang) in der Programmiersprache Perl geschrieben; bei PennCNV war solch ein Skript Bestandteil des Softwarepakets.

Parallel zur Extraktion der personenspezifischen Daten erstellte das Skript für QuantiSNP eine Stapelverarbeitungsdatei, in der für jede zu untersuchende Person ein Eintrag mit der kompletten Kommandozeile für QuantiSNP notiert war. Mit dieser Stapelverarbeitungsdatei kann die Analyse von QuantiSNP automatisiert werden und so nacheinander alle zu untersuchenden Personen mit QuantiSNP analysiert werden. Bei PennCNV war nur eine Kommandozeile nötig, die dann für eine Liste von Dateien nacheinander ausgeführt wurde. Beispiele für die verwendeten Kommandozeilen finden sich im Anhang.

Die Berechnungszeit von QuantiSNP betrug durchschnittlich zehn Minuten pro Person beim HH550-Chip. PennCNV benötigte nur circa zwei Minuten pro Person. Die Ausgabe beider Programme erfolgte über Textdateien. Jede Zeile der Ausgabedatei beinhaltete einen *in silico* identifizierten CNV. Neben allgemeinen Parametern der CNVs (Chromosom, Start-/Endposition, Start-/End-SNP, Länge, Kopienzahl) und der Zahl der vom CNV betroffenen SNPs

lieferte QuantiSNP den *log Bayes Factor* (IBF, siehe Kap. 3.3.6). Bei PennCNV hieß dieser Parameter *confidence value* (CV, siehe Kap. 3.3.6).

3.3.6 Qualitätskontrolle der detektierten CNVs

Pro Person identifizierten QuantiSNP und PennCNV eine große Anzahl an CNVs. Bei einem Großteil dieser CNVs handelte es sich jedoch um technische Artefakte. Damit diese Artefakte für die Assoziationsanalyse ausgeschlossen werden konnten, mussten auf die *in silico* detektierten CNVs Filterkriterien angewendet werden. Diese Kriterien mussten so gewählt sein, dass nach Möglichkeit alle technischen Artefakte verworfen wurden ohne zu viele echte CNVs zu übersehen.

Prinzipiell waren unterschiedliche Arten von Filterkriterien denkbar. So könnte die Länge eines CNVs ein Kriterium sein (The International Schizophrenia Consortium, 2008; Grozeva et al., 2010). Ein weiteres Kriterium könnte die Lage des CNVs sein, hier wäre es zum Beispiel denkbar nur CNVs zu nehmen, die ein Gen betreffen oder solche, die in Regionen liegen, in denen bislang noch kein CNV berichtet wurde (siehe hierzu Kap. 3.3.9 Abschnitt Database of Genomic Variants).

Etablierung der Filterkriterien

Zu Beginn dieser Arbeit waren in unserem Institut keine Filterkriterien validiert worden. Daher war zunächst eine Etablierung solcher Kriterien nötig. Einzelne Personen wurden als Duplikate auf den verwendeten Chiptypen genotypisiert. Mit QuantiSNP und PennCNV wurden dann CNVs detektiert. Anschließend wurde analysiert, welche Filterkriterien angewendet werden müssen, damit in den Duplikaten identische CNVs identifiziert werden. In der vorliegenden Arbeit wurden die Anzahl der betroffenen SNPs und die Qualitätsparameter *log Bayes Factor* bzw. *confidence value* verwendet.

Anzahl der betroffenen SNPs

Die Anzahl der konsekutiven SNPs, die von einem CNV betroffen sind, ist ein Kriterium für die Verlässlichkeit des detektierten CNV. Beide Programme detektieren CNVs ausgehende von der Intensität und der allelischen Ausprägung der SNPs (siehe Kap. 3.3.1). Je mehr SNPs in einem CNV liegen desto mehr Datenpunkte stehen den Programmen für die Berechnung des

CNVs zur Verfügung. Sowohl QuantiSNP als auch PennCNV geben für jeden berechneten CNV die Anzahl der konsekutiven Markern an, die von dem jeweiligen CNV betroffen sind.

Die Verteilung der SNPs auf den BeadArrays ist nicht gleichmäßig. Es gibt Regionen in denen viele und Regionen in denen wenige oder keine SNPs vorhanden sind. Die Zahl der konsekutiven SNPs hängt also nicht nur von der Länge eines CNVs, sondern auch von der SNP-Dichte in der jeweiligen Region ab.

Log Bayes Factor/confidence value

QuantiSNP und PennCNV errechnen für jeden CNV den *log Bayes Factor* (IBF, bei QuantiSNP) beziehungsweise den *confidence value* (CV, bei PennCNV). In beiden Fällen geben diese statistischen Werte die Wahrscheinlichkeit an, dass eine Region eine bestimmte Kopienzahl hat im Vergleich dazu, dass sie diese Kopienzahl nicht hat. Je höher der Wert von IBF/CV, desto eher handelt es sich um einen biologisch echten CNV. Beide Werte werden auf ähnliche Weise berechnet (Colella et al., 2007; Wang et al., 2007). Da QuantiSNP und PennCNV ähnliche Algorithmen für die Detektion von CNVs verwenden haben der IBF und der CV ähnliche Größenordnungen.

3.3.7 Zusammenfügen der Datensätze

Für die statistische Untersuchung der genomweiten Last (siehe Kap. 3.3.8) wurden in dieser Arbeit nur solche CNVs verwendet werden, die sowohl die oben genannten Filterkriterien erfüllten als auch von beiden verwendeten Programmen worden waren. Die beiden separat generierten Datensätze mussten also miteinander fusioniert werden. Für die Fusionierung der Datensätze wurden zunächst die zuvor (siehe Kap. 3.3.6) beschriebenen Filterkriterien angewendet. Die CNVs konnten in unterschiedlicher Weise überlappen, dabei wurde jeweils unterschiedlich vorgegangen um die finale Liste zu bilden (siehe hierzu Abb. 3.7):

1) CNV hat die gleichen Bruchpunkte in beiden Programmen: Die Grenzen des CNVs werden so übernommen.

2) CNV unterscheidet sich in einem Bruchpunkt in beiden Programmen:

Der gemeinsame Bruchpunkt wird übernommen, außerdem der dazu näher liegende.

3) Eines der Programme detektiert einen CNV, der den CNV des anderen Programmes komplett überspannt: Die Bruchpunkte des kleineren, überspannten CNVs werden übernommen.

4) Die beiden CNVs überlappen, teilen aber keinen Bruchpunkt: Die Bruchpunkte werden durch die kleinste geteilte Überlappung definiert.

5) Eines der Programme detektiert einen CNV, das andere Programm detektiert zwei oder mehr CNVs innerhalb dieses größeren CNVs: Zunächst muss in GenomeStudio überprüft werden welches Szenario wahrscheinlich vorliegt (siehe Kap. 3.3.2), demensprechend werden dann die Bruchpunkte gewählt.

Die überlappende Region wird nur verwendet, wenn sie das Filterkriterium für die Anzahl konsekutiver SNPs erfüllt. Wird für die gleiche Person von einem Programm ein CNV detektiert, der die jeweiligen Filterkriterien erfüllt, beim anderen jedoch nicht, oder von diesem überhaupt kein CNV detektiert wird, so wird der CNV verworfen. Ebenso werden CNVs verworfen, die nicht gleichsinnig sind, das heißt eine Deletion im einen Programm und eine Duplikation im anderen Programm.

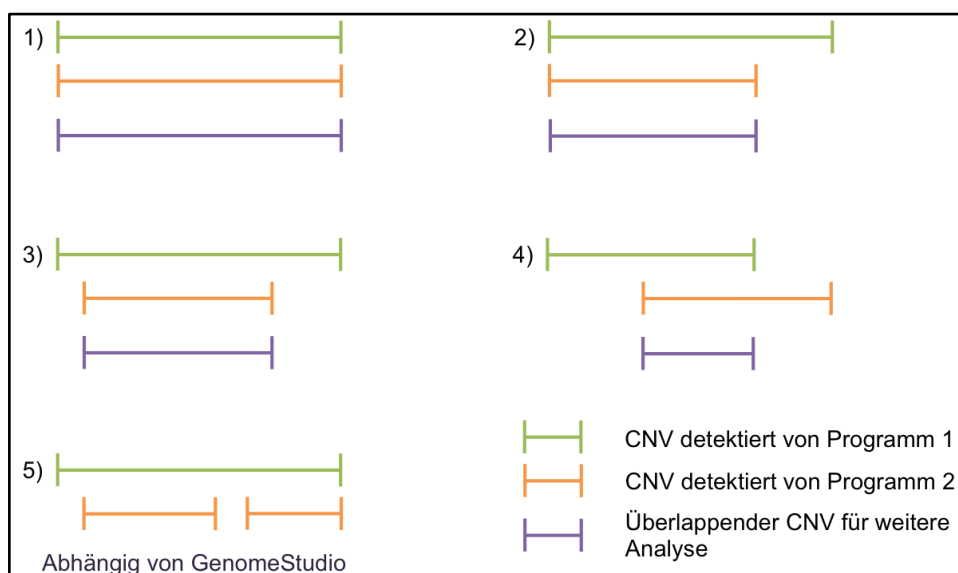


Abb. 3.7: Schematische Abbildung der möglichen Überlappung von CNVs.

3.3.8 Statistische Analysen

Für die genomweiten Untersuchungen für die Beteiligung von CNVs bei den untersuchten Erkrankungen wurden die CNVs in Patienten mit denen in Kontrollen verglichen. Bei der Bipolaren Störungen wurden zudem die Patienten entsprechend ihres Erkrankungsalters in Subgruppen unterteilt (siehe Kap. 3.4.1). Die beiden Subgruppen wurden separat mit Kontrollen verglichen. Es wurden zwei Arten von statistischen Tests durchgeführt: (i) Tests für genomweite Last von CNVs und (ii) Tests für spezifische Assoziationen. In beiden Fällen wurde eine Signifikanzschwelle von 5 % gewählt.

Tests für genomweiten *burden*

Bei einem Test auf genomweite Last wird überprüft, ob CNVs bei Patienten sich in Länge oder Anzahl signifikant von denen in Kontrollen unterscheiden.

In der vorliegenden Arbeit wurde die Software PLINK (v1.07, <http://pngu.mgh.harvard.edu/~purcell/plink/>, Purcell et al., 2007) zur Berechnung der genomweiten Last verwendet. PLINK führt vier unterschiedlichen Tests auf genomweite Last durch:

- 1) RATE: Gesamtzahl der CNVs aller Personen
- 2) PROP: Anteil der Personen mit mindestens einem CNV
- 3) TOTKB: Durchschnittliche Länge der CNVs
- 4) AVGKB: Durchschnittliche Länge der CNVs pro Person

Die beiden ersten Tests (RATE und PROP) vergleichen die Frequenz der CNVs, während die beiden anderen Tests (TOTKB und AVGKB) die Länge der CNVs untersuchen. Für alle vier Tests wurden von PLINK zweiseitige empirische (siehe Abschnitt P-Werte, *odds ratio*, multiples Testen) P-Werte berechnet.

Tests für spezifische Assoziationen

Spezifische Assoziationstests untersuchen, ob CNVs in einer bestimmten chromosomalen Region signifikant häufiger in einer Gruppe von Personen (z.B. Patienten) auftreten als in einer Vergleichsgruppe (z.B. Kontrollen). Auch hierfür wurde zunächst die Software PLINK verwendet. PLINK erstellt anhand der Liste aller zu untersuchenden CNVs eine Liste mit allen Start- und

Endpunkten von CNVs, die in dem Datensatz auftreten. Für diese Punkte wird dann ausgezählt, ob sie signifikant häufiger von CNVs in einer Gruppe betroffen sind. Regionen, die einen signifikanten empirischen P-Wert in der PLINK-Analyse zeigten, wurden weiter untersucht. Für diese Regionen wurde ein exakter Test nach Fisher verwendet, um den P-Wert und das *odds ratio* zu berechnen (siehe Abschnitt P-Werte, *odds ratio*, multiples Testen).

Neben diesen Tests von Regionen, die bislang keine Assoziation mit der untersuchten Erkrankung zeigten wurden auch Regionen untersucht, für die bereits zuvor eine Assoziation mit einer neuropsychiatrischen Erkrankung bekannt war. In diesen Regionen wurde geprüft ob die Patienten mit einer der drei untersuchten Erkrankungen signifikant häufiger einen CNV hatten als die Kontrollen.

P-Werte, *odds ratio*, multiples Testen

In der vorliegenden Arbeit wurden statistische Tests durchgeführt, die die Unterschiedlichkeit in zwei Gruppen einer Stichprobe überprüften. Sowohl beim Test für genomweite Last, als auch beim Test auf spezifische Assoziation wurden P-Werte berechnet. Der P-Wert gibt die Wahrscheinlichkeit an, dass die Nullhypothese eines statistischen Tests, in einer gegebenen Stichprobe, richtig ist. P-Werte kleiner als 0,05, also 5 % Wahrscheinlichkeit, dass die Nullhypothese richtig ist, werden als ausreichend signifikant betrachtet, um die Nullhypothese zu verwerfen und die Alternativhypothese anzunehmen. Für die spezifischen Assoziationstest lautete die Nullhypothese „Patienten und Kontrollen unterscheiden sich NICHT in der Frequenz von CNVs in dieser Region“. Im Gegensatz zu diesen einseitigen Tests wurden bei den Tests auf genomweite Last zweiseitige Tests durchgeführt. Die Nullhypothese kann dann zum Beispiel lauten: „Die Frequenz der CNVs ist in Patienten im Vergleich zu Kontrollen höher ODER niedriger“.

Für spezifische Assoziationstests wurde der exakte Test nach Fisher (*Fisher's exact test*, Fisher, 1922) angewendet. Dieser Test basierte auf einer Vierfeldertafel, in der die Anzahl von Patienten und Kontrollen mit und ohne CNV in der untersuchten Region aufgetragen sind:

	Mit CNV	Ohne CNV
Patienten	a	b
Kontrollen	c	d

Daraus ergibt sich dann der P-Wert (mit n = Stichprobengröße):

$$P = \frac{(a + b)! * (c + d)! * (a + c)! * (b + d)!}{a! * b! * c! * d! * n!}$$

Außerdem gibt das Kreuzprodukt dieser Vierfeldertafel das *odds ratio* an:

$$odds\ ratio = \frac{(a * d)}{(b * c)}$$

Das *odds ratio* (OR) gibt die Erhöhung des Risikos einer Person die untersuchte Erkrankung zu erleiden, im Vergleich zur Allgemeinbevölkerung. Während also der P-Wert angibt ob ein CNV mit einer Erkrankung assoziiert ist, ist das OR ein Maß für die Effektstärke.

Bei genomweiten Untersuchungen werden in einer Stichprobe mehrere unabhängige Tests durchgeführt. Um einer Inflation der Fehler erster Art (bei denen die Nullhypothese verworfen wird obwohl sie richtig ist) zu vermeiden, muss jeder berechnete P-Wert korrigiert werden. Die einfachste, aber auch konservativste Methode der Korrektur ist die Bonferroni-Korrektur (Dunn, 1961). Bei dieser Korrektur wird der individuelle P-Wert mit der Anzahl der durchgeführten unabhängigen Tests multipliziert.

Die P-Werte, die PLINK berechnet, sind empirische P-Werte und werden durch Permutation des Datensatzes bestimmt. Bei jeder Permutation wird die der Status eines Patienten oder einer Kontrolle, nicht aber CNV-Träger oder nicht CNV-Träger, zufällig vergeben. Wenn alle Permutationen erfolgt sind ergibt sich der P-Wert der tatsächlich beobachteten Verteilung aus dem Verhältnis aller beobachteten Permutationen, die die gleiche oder erhöhte Verteilung zeigten zur Gesamtzahl der durchgeführten Permutationen.

$$P_{emp} = \frac{n(\text{gleiche oder erhöhte Verteilungen})}{n(\text{alle Permutationen})}$$

In der vorliegenden Arbeit wurden jeweils eine Million Permutationen durchgeführt.

3.3.9 Einschätzung der biologischen Relevanz von CNVs

Um die biologische Relevanz der assoziierten CNVs besser einschätzen zu können, wurde auf unterschiedliche, öffentlich zugängliche Datenbanken zurückgegriffen: i) die Datenbank der *University of California Santa Cruz* (UCSC), die Datenbank des *Sullivan Lab Evidence Project* (SLEP), iii) die *Database of Genomic Variants* (DGV, siehe Kap. 1.3.3), iv) die Datenbank des *Children's Hospital of Philadelphia* (CHOP) und v) die *Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources* (DECIPHER).

Datenbank der *University of California Santa Cruz*

Die UCSC bietet eine Metadatenbank von insgesamt 58 Spezies (<http://genome.ucsc.edu/>). Das Spektrum der Spezies reicht von der Bäckerhefe über Invertebraten und Vertebraten bis hin zum Menschen. Für jede Spezies ist die aktuellste Referenzsequenz verfügbar, darüber hinaus werden auch ältere Referenzsequenzen angeboten. Für den Menschen sind vier Referenzsequenzen verfügbar (NCBI build 34, 35 und 36 und GRCh build 37). Die Metadatenbank ist mit verschiedenen anderen Datenbanken verbunden, so kann eine DNA-Region beispielsweise auf bereits bekannte Genexpressionsmuster, regulatorische Elemente oder Assoziation mit verschiedenen Phänotypen abgefragt werden. In der vorliegenden Arbeit wurde der NCBI build 36 verwendet, dies entsprach der von den Genotypisierungschips verwendeten Referenz. Die UCSC Datenbank wurde vor allem für die Überprüfung der Überlappung von spezifischen CNVs mit RefSeq Genen und die Abdeckung auf unterschiedlichen Genotypisierungschips verwendet. Außerdem wurde die Möglichkeit eigenen Daten in der grafischen Oberfläche darzustellen verwendet um die Übersichtsabbildungen der assoziierten Regionen im Ergebnisteil zu erstellen.

Sullivan Lab Evidence Project

Im SLEP (<http://gbrowse.csbio.unc.edu/cgi-bin/gb2/gbrowse/slep/>; Konneker et al., 2008) sind die Ergebnisse genetischer Studien zu verschiedenen neuropsychiatrischen Erkrankungen, unter anderem auch Schizophrenie, Bipolare Störung und Depression, archiviert. Bei den Studien handelt es sich um Kopplungsuntersuchungen, genomweite Assoziationsstudien und

Microarraystudien. Außerdem sind in der SLEP differentiell exprimierte Gene, CNVs und Inversionen annotiert. Diese Datenbank ist ein Katalog aller bislang publizierten Assoziationsstudien zu den unterschiedlichen Erkrankungen und bietet die Möglichkeit der graphischen Darstellung dieser Ergebnisse. Die SLEP Datenbank wurde verwendet, um zu überprüfen genetische Varianten in spezifischen CNVs bereits mit neuropsychiatrischen Erkrankungen assoziiert gefunden worden waren.

Database of Genomic Variants

Die DGV (<http://dgv.tcag.ca/dgv/app/home/>; lafrate et al., 2004; Zhang, et al., 2006) wurde bereits in Kapitel 1.3.3 genauer beschrieben. In der Datenbank sind Informationen von populationsbasierten Kontrollen aus 55 Studien enthalten. Mit Hilfe der DGV kann die Frequenz einer bestimmten CNVs in zusätzlichen Kontrollkollektiven überprüft werden.

Datenbank des Children's Hospital of Philadelphia

Das CHOP bietet eine Datenbank, in der genomweite CNV Daten von 2.026 gesunden Kindern zugänglich sind (<http://cnv.chop.edu/>; Shaikh et al., 2009). Alle Kinder wurden auf Illumina HumanHap550v3 SNP-Chip genotypisiert. Ähnlich wie in der DGV kann in der CHOP Datenbank die Frequenz eines CNVs in einem Kollektiv von gesunden Personen überprüft werden. Zu bedenken ist allerdings, dass in der CHOP Datenbank nur Daten von Kindern verfügbar sind, diese sind bei spät manifestierenden Erkrankungen nur bedingt aussagekräftig.

Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources

In der DECIPHER (<http://decipher.sanger.ac.uk/>; Firth et al., 2009) werden CNV Daten von Patienten mit Intelligenzminderung gespeichert. Neben den Informationen zu CNVs werden auch umfangreiche phänotypische Beschreibungen der Patienten angeboten. DECIPHER bietet, ähnlich wie die DGV und CHOP Datenbank, den Zugriff auf CNV-Informationen von zusätzlichen Personen. Anders als bei den beiden anderen Datenbanken haben diese Personen jedoch alle eine Intelligenzminderung.

3.4 Verwendete Kollektive

Bei Patienten Kollektiven mit den drei neuropsychiatrischen Erkrankungen Bipolare Störung, Majore Depression und Schizophrenie wurden Untersuchungen zur Beteiligung von CNVs bei der Krankheitsentstehung durchgeführt. Die untersuchten Patienten und Kontrollen wurden zum Teil auf unterschiedlichen SNP-Chips analysiert, da zunächst die Produktion des HumanHap550v3 SNP-Chips und später auch des Human610-Quadv1 SNP-Chips vom Hersteller eingestellt wurde.

3.4.1 Bipolare Störung

Es wurden 957 Patienten und 880 populationsbasierte Kontrollen für die CNV Analysen bei Bipolarer Störung verwendet. Alle untersuchten Personen waren deutscher Herkunft. Die Patienten wurden auf dem HumanHap550v3 (HH550) oder Human610-Quadv1 (H610Q) SNP-Chip (siehe Kap 3.2.3) genotypisiert. Alle Kontrollen wurden auf dem HH550 genotypisiert. Für eine detaillierte Aufstellung der verwendeten Kollektive siehe Tab. 3.3.

Die Patienten wurden in zwei klinischen Zentren rekrutiert: i) dem Zentralinstitut für Seelische Gesundheit in Mannheim und ii) der Klinik und Poliklinik für Psychiatrie und Psychotherapie des Universitätsklinikums Bonn. Alle Patienten erfüllten die Diagnosekriterien für Bipolare I Störung oder Bipolare II Störung des DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders fourth edition*; American Psychiatric Association, 2007).

Die Kontrollen waren Teil zweier populationsbasierter epidemiologischer Studien: i) der Heinz Nixdorf Recall (HNR) Studie (Schmermund et al., 2002) und ii) der PopGen (Population-Based Recruitment of Patients and Controls for the Analysis of Complex Genotype-Phenotype Relationships) Studie (Krawczak et al., 2006). In der HNR Studie wurden Personen zur Untersuchung von Herz-Kreislauf-Erkrankungen prospektiv rekrutiert. Die Teilnehmer der HNR Studie stammten aus Städten im Ruhrgebiet: Essen, Bochum und Mülheim an der Ruhr. Auch die PopGen Studie ist eine prospektive Studie, neben Herz-Kreislauf-Erkrankungen werden in dieser Studie zudem weitere Erkrankungen wie Asthma, chronische Darmerkrankungen und Krebs untersucht. Die Teilnehmer der PopGen Studie wurden in Schleswig-Holstein rekrutiert. Aus

der HNR Studie wurden 383 Personen verwendet, aus der PopGen Studie 497 Personen.

Patienten mit besonders frühem Erkrankungsalter

Klinische Studien haben gezeigt, dass Patienten, die früh an einer Bipolaren Störung erkranken im Vergleich zu Patienten, die spät erkranken einen eine andere familiäre Belastung haben (Grigoriu-Serbanescu et al., 2001). Patienten bilden wahrscheinlich eine genetische Subgruppe innerhalb der Patienten mit Bipolaren Störung, bei denen seltene genetische Varianten eine größere Rolle spielen (Malhotra und Sebat, 2012). Sie sind also sowohl klinisch als auch genetisch eine Untergruppe. Daher wurden Patienten mit besonders frühem Erkrankungsalter und Patienten mit späterem Erkrankungsalter getrennt untersucht. Patienten, die ein Erkrankungsalter 21 Jahre oder jünger waren, wurden in die Gruppe der Früherkrankten sortiert. Die übrigen Patienten bildeten die Gruppe der Späterkrankten. Das Erkrankungsalter der Patienten wurde definiert als das Alter in dem die erste manische oder depressive Phase auftrat.

3.4.2 Majore Depression

Ein Kollektiv von 604 Patienten und 1.643 Kontrollen wurde für die CNV Analysen bei Majorer Depression analysiert. Alle untersuchten Personen waren deutscher Herkunft. Die Patienten wurden auf dem H610Q, die Kontrollen wurden auf dem HH550, dem H610Q oder dem Human660W-Quadv1 (H660W) genotypisiert (siehe Tab. 3.3)

Die Patienten wurden in der Klinik und Poliklinik für Psychiatrie und Psychotherapie des Universitätsklinikums Bonn rekrutiert. Alle Patienten erfüllten die DSM-IV Diagnosekriterien für Majore Depression (American Psychiatric Association, 2007).

Die Kontrollen setzten sich zusammen aus populationsbasierten 1.357 Kontrollen und 286 Personen, die im Rahmen einer Bildgebungsstudie (MooDS Imaging study; Rietschel et al., 2012) rekrutiert wurden. Die populationsbasierten Kontrollen stammten aus drei Studien: i) der HNR Studie (siehe Kap 3.4.1; Schmermund et al., 2002), ii) der PopGen Studie (siehe Kap. 3.4.1; Krawczak et al., 2006) und iii) der Kooperative Gesundheitsforschung in

der Region Augsburg (KORA; Wichmann et al., 2005). Die KORA Studie ist ebenfalls eine prospektive, populationsbasierte Studie. Die Teilnehmer stammten aus der schwäbischen Region um Augsburg. Aus der KORA Studie wurden 481 Personen verwendet. Die Teilnehmer der Bildungsstudien wurden in drei Zentren rekrutiert: i) dem Zentralinstitut für Seelische Gesundheit in Mannheim, ii) der Klinik und Poliklinik für Psychiatrie und Psychotherapie des Universitätsklinikums Bonn und iii) der Klinik für Psychiatrie und Psychotherapie der Charité Universitätsmedizin Berlin. Alle Individuen der Bildungsstudie wurden von einem Psychiater untersucht. Anamnestisch erfüllte keiner der Probanden die DSM-IV Kriterien für eine neuropsychiatrische Erkrankung.

3.3.3 Schizophrenie

Insgesamt wurden 1.831 Patienten und 1.643 Kontrollen für die CNV Analysen bei der Schizophrenie verwendet. Patienten und Kontrollen waren deutscher Abstammung. Sowohl Patienten als auch Kontrollen wurden auf dem HH550, H610Q oder H660W genotypisiert (siehe Tab. 3.3).

Die Patienten wurden in vier klinischen Zentren rekrutiert: i) dem Zentralinstitut für Seelische Gesundheit in Mannheim, ii) der Klinik und Poliklinik für Psychiatrie und Psychotherapie der Ludwig-Maximilians-Universität München, iii) der Klinik für Psychiatrie und Psychotherapie des Universitätsklinikums Jena und iv) der Klinik und Poliklinik für Psychiatrie und Psychotherapie des Universitätsklinikums Bonn. Nach DSM-IV erfüllten alle Patienten die Diagnosekriterien für eine Schizophrenie.

Alle in dieser Studie verwendeten Kontrollen wurden auch bei der Untersuchung von CNVs bei der Majoren Depression verwendet (siehe Kap. 3.4.2). Die Kontrollen der HNR- und PopGen-Studien wurden bei allen drei neuropsychiatrischen Erkrankungen verwendet. Auf Grund Personenspezifischer Filterkriterien unterschieden sich Individuenzahl nach Anwendung der Filterkriterien

Tab. 3.3: Übersicht über die untersuchten Kollektive. Die angegebene Anzahl der Personen ist vor Anwendung von Filterkriterien. Abkürzungen: HH550 - HumanHap550v3, H610Q - Human610-Quadv1, H660W - Human660W-Quadv1.

Gruppe	Anzahl Personen	Chiptyp
Bipolare Störung	957	HH550 + H610Q
	725	HH550
	232	H610Q
Majore Depression	604	H610Q
Schizophrenie	1.831	HH550+ H610Q + H660W
	493	HH550
	643	H610Q
	695	H660W
Pop. Kontrollen	1.357	HH550
HNR	383	HH550
PopGen	497	HH550
KORA	481	HH550
Bildgebungskontrollen	286	H610Q + H660W
	122	H610Q
	164	H660W

3.5 Durchgeführte Analysen

3.5.1 Etablierung der Filterkriterien

Bei der *in silico* Identifikation von CNVs durch SNP-Chips kann es zu falsch positiv identifizierten CNVs kommen, da quantitative Daten ausgewertet werden. Einen nachgewiesenen Einfluß auf die CNV Detektion haben unter anderem die Herkunft der DNA, die Genotypisierung an verschiedenen Zentren oder die Verwendung unterschiedlicher Chemiechargen (The Wellcome Trust Case Control Consortium, 2010). Zum Ausschluss solcher falsch positiver

CNVs ist die Anwendung stringenter Filterkriterien auf den jeweiligen Datensatz notwendig. Zu Beginn dieser Arbeit waren keine einheitlichen Filterkriterien für die Analyse von CNVs, die mit Hilfe von SNP-Arrays identifiziert wurden, etabliert. Daher mussten zunächst Schwellenwerte für eine stringente Qualitätskontrolle ermittelt werden.

Als Filterkriterien für die einzelnen identifizierten CNVs wurde die Anzahl der betroffenen SNPs und die Qualitätsparameter *log Bayes Factor* (IBF) beziehungsweise *confidence value* (CV) gewählt. Für die Bestimmung der Schwellenwerte für diese Filterkriterien wurden insgesamt 27 Personen doppelt genotypisiert. Die erhobenen Daten jedes Duplikats wurden dann für die Identifikation potentieller CNVs verwendet. Die Übereinstimmung der detektierten CNVs aus beiden Duplikaten wurde als Maß für die Verlässlichkeit der gewählten Filterkombination gewählt. Als Standardqualitätsparameter wurde eine hausinterne Kontrolle mitgeführt. Diese anonymisierte DNA-Probe wurde bereits zu unterschiedlichen Zeitpunkten, mit unterschiedlichen Chemiechargen und auf unterschiedlichen Chiptypen prozessiert.

Drei Kombinationen der Filterkriterien wurden untersucht: i) mindestens 10 betroffene SNPs und ein IBF/CV von mindestens 10, ii) mindestens 20 betroffene SNPs und ein IBF/CV von mindestens 20, iii) mindestens 30 betroffene SNPs und ein IBF/CV von mindestens 30.

3.5.2 Beteiligung von CNVs an der Erkrankungsentstehung

In jedem der in Kapitel 3.4 beschriebenen Kollektive wurden drei unterschiedliche Analysen durchgeführt, um die Beteiligung von CNVs an der Entstehung der jeweiligen Erkrankung zu untersuchen:

- 1) Unterschiede im genomweiten *burden*
- 2) Genomweite Assoziationsanalyse spezifischer CNVs
- 3) Frequenz von CNVs in bekannten Kandidatenregionen

Unterschiede im genomweiten *burden*

Die Tests auf Unterschiede im genomweiten *burden* (siehe Kap. 3.3.8) wurden in sechs CNV Klassen (siehe Abb. 3.8) in Patienten mit denen in Kontrollen verglichen:

- 1) Alle CNVs (d.h. Deletionen und Duplikationen)
- 2) Deletionen
- 3) Duplikationen
- 4) CNVs, die nur ein einziges Mal im gesamten Datensatz vorhanden sind (*singleton CNVs*)
- 5) *Singleton* Deletionen
- 6) *Singleton* Duplikationen

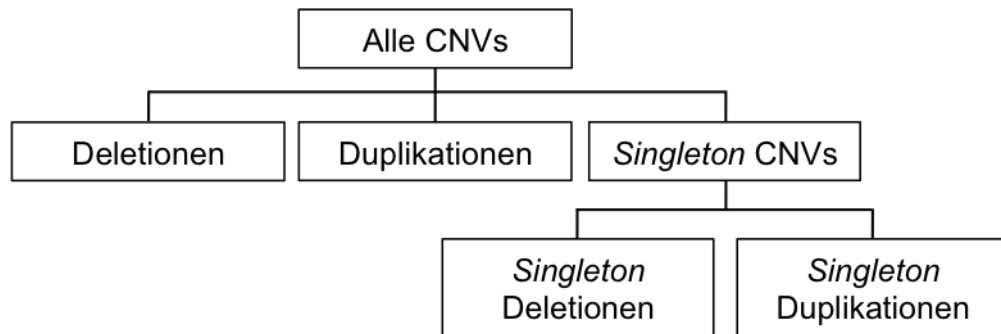


Abb. 3.8: Klassen der CNVs, in denen die Analysen durchgeführt wurden.

Spezifisch assoziierte CNV Regionen

Auch bei der genomweiten Suche nach mit der Erkrankung spezifisch assoziierten Regionen (siehe Kap. 3.3.8) wurden alle CNVs zusammen und getrennt nach Deletionen und Duplikationen analysiert. Dieser Ansatz war hypothesenfrei.

CNVs in bekannte Kandidatenregionen

Zahlreiche Studien konnten bereits eine Assoziation von CNVs in verschiedenen Regionen mit neuropsychiatrischen Erkrankungen feststellen. Die Beteiligung dieser Kandidatenregionen bei den untersuchten Kollektiven sollte ebenfalls geprüft werden. Die Kandidatenregionen für die dritte Analyse mussten zwei Kriterien erfüllen: i) Assoziation von CNVs in der Region mit Bipolarer Störung, Majorer Depression oder Schizophrenie in einer genomweiten CNV Studie und ii) mindestens eine weitere Studie, die eine Beteiligung von CNVs in der Region an der Entstehung von Autismus, Bipolarer Störung, Majorer Depression oder Schizophrenie zeigte. Diese Kriterien wurden von elf Regionen erfüllt (Details zu den Regionen in Tab. 3.4): 1q21.1, 2p16.3, 3q29, 7q36.3, 15q11.2, 15q13.3, 16p13.11, 16p11.2, 17p12, 17q12 und 22q11.21. In jeder dieser Regionen wurde die Frequenz von CNVs, die zumindest 80 % mit den vorbeschriebenen Regionen überlappten in Patienten und Kontrollen verglichen. Mit zwei Ausnahmen: i) CNVs in 2p16.3 mussten das Gen *Neurexin-1* (*NRXN1*) betreffen und ii) CNVs in 16p13.11 mussten

mindestens 80 % mit dem Intervall II aus der Studie von Ingason *et al.* überlappen (Ingason *et al.*, 2009). Nur CNVs, die dieses Intervall II betrafen, zeigten in der letztgenannten Studie eine Assoziation mit Schizophrenie.

Tab. 3.4: Übersicht bekannter Kandidatenregionen. Position in Mb nach NCBI build 36

Region	Position	Referenzen
1q21.1	144,94-146,29	The International Schizophrenia Consortium (2008), Stefansson <i>et al.</i> (2008), Levinson <i>et al.</i> (2011)
2p16.3	49,90-51,50	Kirov <i>et al.</i> (2009a), Rujescu <i>et al.</i> (2009), Levinson <i>et al.</i> (2011)
3q29	197,2–198,83	Mulle <i>et al.</i> (2010), Levinson <i>et al.</i> (2011), Vacic <i>et al.</i> (2011)
7q36.3	158,51–158,63	Levinson <i>et al.</i> (2011), Vacic <i>et al.</i> (2011)
15q11.2	20,31-20,78	Stefansson <i>et al.</i> (2008), Kirov <i>et al.</i> (2009b), Magri <i>et al.</i> (2010)
15q13.3	28,72-30,30	The International Schizophrenia Consortium (2008), Stefansson <i>et al.</i> (2008), Magri <i>et al.</i> (2010), Levinson <i>et al.</i> (2011), Vacic <i>et al.</i> (2011)
16p13.11	15,38-16,19	Ingason <i>et al.</i> (2009), Kirov <i>et al.</i> (2009b), Magri <i>et al.</i> (2010)
16p11.2	29,50-30,30	McCarthy <i>et al.</i> (2009), Levinson <i>et al.</i> (2011), Vacic <i>et al.</i> (2011)
17p12	14,0–15,4	Kirov <i>et al.</i> (2009b), Magri <i>et al.</i> (2010)
17q12	31,89-33,28	Magri <i>et al.</i> (2010), Moreno-De-Luca <i>et al.</i> (2010)
22q11.21	17,00-19,80	Karayiorgou <i>et al.</i> (1995), The International Schizophrenia Consortium (2008), Levinson <i>et al.</i> (2011), Vacic <i>et al.</i> (2011)

4 Ergebnisse

4.1 Etablierung von Filterkriterien

Zu Beginn dieser Arbeit gab es keine publizierten Referenzarbeiten zur CNV-Qualitätskontrolle. Zur Reduktion falsch positiver CNVs mussten daher zunächst Schwellenwerte für Filterkriterien etabliert werden.

Für die Etablierung der Schwellenwerte für die Filterkriterien Anzahl betroffener SNPs und log Bayes Factor (IBF) beziehungsweise confidence value (CV) wurden 27 Patienten doppelt genotypisiert. Die aus diesen Daten detektierten potentiellen CNVs wurden bei Anwendung verschiedener Kombinationen der Filterkriterien verglichen.

Bei den 27 untersuchten Doubletten wurden insgesamt 108 CNVs detektiert, die mindestens 10 SNPs betrafen und ein IBF/CV von mindestens 10 hatten. Von diesen wurden 73 (~68 %) bei beiden Replikaten detektiert. 43 CNVs betrafen mindestens 20 SNPs und hatten einen IBF/CV von mindestens 20, davon wurden 32 (~74 %) bei beiden Replikaten gefunden. Insgesamt 21 CNVs überstanden die stringenteste Filterkombination; sie betrafen mindestens 30 SNPs und hatten einen IBF/CV von mindestens 30, diese CNVs konnten alle bei beiden Doubletten detektiert werden.

Auf Grund der hundertprozentigen Übereinstimmung der Duplikate bei der Filterkombination mindestens 30 betroffenen SNPs und IBF/CV von mindestens 30 wurde diese Kombination für die weiteren Analysen verwendet.

Tab. 4.1: Übersicht detektierter CNVs und Anteil der CNVs, die in beiden Replikaten detektiert wurden, bei unterschiedlichen Filterkombinationen. IBF – log Bayes Factor, CV – confidence value

Filterkriterien (IBF/CV - SNPs)	Detektierte CNVs	CNVs bei beiden Replikaten
10 - 10	108	73 (~68 %)
20 - 20	43	32 (~74 %)
30 - 30	21	21 (100 %)

4.2 Beteiligung von CNVs bei der Bipolaren Störung

4.2.1 Qualitätskontrollierter Datensatz

Die in den Kapiteln 3.3.4 sowie 3.3.6 geschilderten und in Kapitel 4.1 etablierten Filterkriterien wurden für die genomweiten CNV-Analysen bei der Bipolaren Störung angewendet. Zusätzlich wurden zwei weitere Filterkriterien auf den Datensatz angewendet: Personen wurden ausgeschlossen, wenn sie extreme Ausreißer (siehe Kapitel 3.3.4) i) in Bezug auf die Standardabweichung (SA) ihrer *B allele frequency* ($SA > 0,12$) oder ii) bezüglich der Anzahl von CNV (mehr als sieben CNVs) waren. Beide Parameter sind Hinweise auf schlechte DNA-Qualität.

Von den initialen 957 Patienten verblieben nach Anwendung aller Filterkriterien 882 Patienten im analysierten Datensatz. Das durchschnittliche Alter dieser Patienten war 44,0 Jahre ($SA = 13,4$) bei Rekrutierung, das durchschnittliche Erkrankungsalter betrug 27,9 Jahre ($SA = 11,3$). Bei der Diagnosestellung waren 291 Patienten höchstens 21 Jahre alt.

Im Kontrolldatensatz blieben nach Anwendung aller Filterkriterien 872 Personen übrig, mit einem Durchschnittsalter von 48,0 Jahren ($SA = 11,4$) bei Rekrutierung.

In den 1.754 Personen wurden insgesamt 1.044 CNVs detektiert, das entspricht etwa 0,6 CNVs pro Person.

4.2.2 Genomweite *burden* Tests

Insgesamt wurden 72 genomweite *burden* Tests durchgeführt. Hierbei wurde getestet ob sich die Länge oder Anzahl von CNVs bei Patienten und Kontrollen signifikant unterschieden. Im Folgenden werden die wichtigsten Ergebnisse genauer betrachtet werden. Tabelle 4.2 zeigt alle Ergebnisse.

Vergleich zwischen allen Patienten und Kontrollen

Singleton Deletionen, die nur ein einziges Mal im gesamten Datensatz von Patienten und Kontrollen auftraten, waren bei Patienten signifikant länger im Vergleich mit allen Kontrollen (472,6 kb in Patienten, 249,3 kb in Kontrollen; P-Wert: 0,0141). Alle übrigen genomweite *burden* Tests waren nicht signifikant (siehe Tab. 4.2).

Vergleich zwischen Patienten mit besonders frühem Erkrankungsalter und Kontrollen

Beim Vergleich der Subgruppe von Patienten mit besonders frühem Erkrankungsalter (siehe Kap 3.4.1) mit allen Kontrollen war die Länge der *singleton* Deletionen ebenfalls signifikant größer bei Patienten (679,6 kb in Patienten, 249,3 kb in Kontrollen; P-Wert: 0,0084). Außerdem war die Frequenz von Personen mit mindestens einem CNV bei Patienten mit frühem Erkrankungsalter signifikant größer als im Kontrollkollektiv (52,76 % in Patienten, 42,32 % in Kontrollen, P-Wert: 0,0009). Die Signifikanz des Frequenzunterschieds war noch stärker, wenn Duplikationen allein betrachtet wurden (44,48 % in Patienten, 33,14 % in Kontrollen, P-Wert: 0,0004). Die letztgenannte Assoziation war auch nach Korrektur für multiples Testen mit der Bonferroni-Methode (siehe Kap. 3.3.8) signifikant mit Bipolarer Störung assoziiert (Korrektur für 72 durchgeführte Tests; korrigierter P-Wert: 0,0288).

Vergleich zwischen Patienten mit späterem Erkrankungsalter und Kontrollen

Keiner der genomweiten *burden* Tests erreichte die Signifikanzschwelle beim Vergleich zwischen Patienten jenseits des 21ten Lebensjahrs und Kontrollen.

Tab. 4.2: Übersicht der *burden* Tests bei Bipolarer Störung. Es werden nur die P-Werte kleiner als 0,1 dargestellt, die Gesamtergebnisse befinden sich im Anhang. Die fett markierten P-Werte werden im Text genauer beschrieben. * Dieser P-Wert blieb nach Korrektur für multiples Testen signifikant (korrigierter P-Wert: 0,0288). Modifiziert nach Priebe et al., 2012.

Vergleichsgruppe	Test			
	CNV Frequenz		CNV Länge	
	RATE-Test	PROP-Test	TOTKB-Test	AVGKB-Test
Alle Patienten (882) gegen alle Kontrollen (872)				
Alle CNVs	--	--	--	--
Nur Duplikationen	--	--	--	--
Nur Deletionen	--	--	--	--
Alle <i>singleton</i> CNVs	--	--	--	--
Nur <i>singleton</i> Duplikationen	0,0633	0,0534	--	--
Nur <i>singleton</i> Deletionen	--	--	0,0141	0,0139
Patienten mit frühem Erkrankungsalter (291) gegen alle Kontrollen (872)				
Alle CNVs	0,0875	0,0009	--	0,0919
Nur Duplikationen	0,0238	0,0004 *	--	--
Nur Deletionen	--	--	--	--
Alle <i>singleton</i> CNVs	0,0453	0,0874	0,0812	0,0939
Nur <i>singleton</i> Duplikationen	0,0102	0,0166	--	--
Nur <i>singleton</i> Deletionen	--	--	0,0084	0,0056

4.2.3 Spezifische Assoziationen

Bei der genomweiten Untersuchung des Datensatzes auf spezifische CNVs wurden zwei chromosomale Regionen identifiziert, die bei Patienten mit Bipolarer Störung signifikant häufiger von CNVs betroffen waren als bei Kontrollen.

Assoziation von Duplikationen in 6q27

Duplikationen in der chromosomalen Bande 6q27 waren mit Bipolarer Störung mit frühem Erkrankungsalter assoziiert. Von 291 Patienten mit einem frühen Erkrankungsalter hatten 17 (5,84 %) eine Duplikation in dieser Region. Im Vergleich dazu hatten 22 von 872 Kontrollen (2,52 %) eine Duplikation (P-Wert: 0,0039; OR: 2,40; 95-KI: 1,18-4,80). Die potentiellen Bruchpunkte der Duplikationen unterschieden sich zum Teil um einige SNPs. Die bei allen Personen assoziierte CNV-Region umfasste 110 SNPs und war etwa 248 kb groß (168.086.687-168.334.983; siehe Abb. 4.1). Drei Gene waren in dieser Region lokalisiert: *kinesin family member 25* (*KIF25*) und *FERM domain containing 1* (*FRMD1*) komplett, und von *mixed-lineage leukemia translocated 4* (*MLLT4*) nur ein Teil des 3'-Terminus. Alle Duplikationen konnten mit qRT-PCR verifiziert werden.

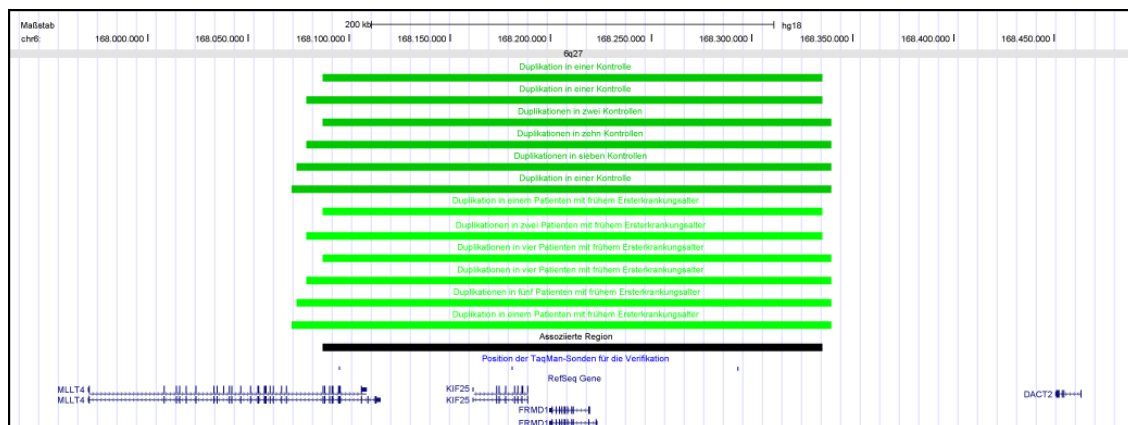


Abb. 4.1: CNVs in der chromosomalen Bande 6q27. Bei 872 untersuchten Kontrollen (dunkel grüne Balken) wurden 22 Duplikationen mit sechs unterschiedlichen Bruchpunkten detektiert. In 291 Patienten mit frühem Erkrankungsalter (hell grüne Balken) wurden 17 Duplikationen detektiert. Die assoziierte Region (schwarzer Balken) überspannte etwa 248 kb. Für die Verifikation der CNVs wurden drei TaqMan-Sonden (blaue vertikale Striche) verwendet. Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Priebe et al., 2012.

Assoziation von Duplikationen in 10q11.22

In der chromosomalen Subbande 10q11.22 waren Duplikationen mit Bipolarer Störung assoziiert. Duplikationen wurden in 53 von 882 Patienten (6,01 %) und in 32 von 872 Kontrollen (3,67 %) beobachtet (P-Wert: 0,0354; OR: 1,53 95-KI: 1,05-2,72). In der Gruppe der Patienten mit frühem Erkrankungsalter hatten 28 von 291 Patienten (9,62 %) die Duplikation (P-Wert: 0,0005; OR: 2,79; 95-KI: 1,59-4,89). Die identifizierten Duplikationen zeigten die gleichen *in silico* berechneten genomischen Bruchpunkte, sie überspannten 30 SNPs und etwa 160 kb (47.013.328-47.173.619; siehe Abb. 4.2).

Die Duplikationen umfassten das gesamte Gen *anthrax toxin receptor-like* (*ANTXRL*). Dieses Gen wird seit der Identifizierung einer Assoziation von CNVs in dieser Region mit der Bipolaren Störung (Priebe et al., 2012) nicht mehr in der *RefSeq* Datenbank geführt. Nach neuestem Datenstand (Dezember 2012) befindet sich somit kein *RefSeq* Gen in der assoziierten Region. Alle Duplikationen konnten mit qRT-PCR mit Taqman-Sonden verifiziert werden.

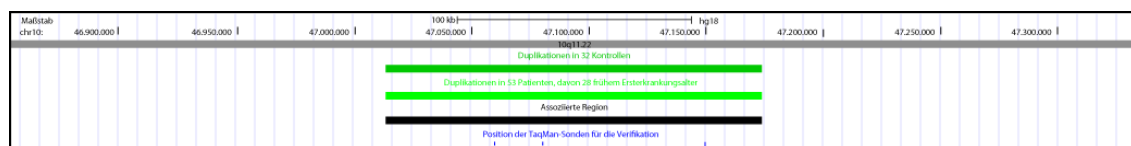


Abb. 4.2: CNVs in der chromosomalen Bande 10q11.22. Bei 872 untersuchten Kontrollen (dunkel grüne Balken) wurden 32 Duplikationen detektiert. In 882 Patienten wurden 53 Duplikationen detektiert, davon 28 in Patienten mit frühem Erkrankungsalter (hell grüne Balken). Alle detektierten Duplikationen hatten die gleichen Bruchpunkte. Die assoziierte Region (schwarzer Balken) überspannte etwa 160 kb. Für die Verifikation der CNVs wurden drei TaqMan-Sonden (blaue vertikale Striche) verwendet. Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Priebe et al., 2012.

4.2.4 Frequenz von CNVs in vorbeschriebenen Regionen

Keine der elf mit neuropsychiatrischen Erkrankungen vorbeschriebenen Regionen (siehe Kap. 3.5.2) war im vorliegenden Datensatz mit Bipolarer Störung assoziiert (siehe Tab. 4.3). Nur in einer der elf Regionen (15q11.2) konnten CNVs in Patienten detektiert werden, die zu mindestens 80 Prozent mit den vorbeschriebenen CNVs überlappten.

Region	Position *	Patienten (882)	Kontrollen (872)	P	OR	OR-KI	P-dup	OR-dup	KI-dup	P-del	OR-del	KI-del
1q21.1	144.94-146.29	--	--	--	--	--	--	--	--	--	--	--
2p16.3	49.90-51.50	--	--	--	--	--	--	--	--	--	--	--
3q29	197.2-198.83	--	--	--	--	--	--	--	--	--	--	--
7q36.3	158.51-158.63	--	--	--	--	--	--	--	--	--	--	--
15q11.2	20.31-20.78	4 dup, 1 del	2 dup, 3 del	1	0,99	0,23-4,31	0,69	1,98	0,28-21,95	0,37	0,33	0,01-4,11
15q13.3	28.72-30.30	--	--	0,25	**	**	--	--	--	0,25	**	**
16p13.11	15.38-16.19	--	2 del	--	--	--	--	--	--	--	--	--
16p11.2	29.50-30.30	--	--	--	--	--	--	--	--	--	--	--
17p12	14.0-15.4	--	--	--	--	--	--	--	--	--	--	--
17q12	31.89-33.28	--	--	--	--	--	--	--	--	--	--	--
22q11.21	17.00-19.80	--	--	--	--	--	--	--	--	--	--	--
Summe	--	4 dup, 1 del	3 dup, 5 del	0,42	0,62	0,16-2,15	1	1,32	0,22-9,03	0,12	0,20	0,01-1,77

Tab. 4.3: CNVs in Regionen mit Vorbefunden bei neuropsychiatrischen Erkrankungen bei der Bipolaren Störung. CNVs im untersuchten Kollektiv für Bipolare Störung die mit vorbeschriebenen Regionen überlappten. P – P-Wert; OR – *odds ratio*, OR-KI – 95%-Konfidenzintervall des ORs; dup – Duplikationen; del – Deletionen; * Positionsangaben entsprechend NCBI build 36; ** Wegen fehlendem Ereignis in Patienten kann weder das OR noch das Konfidenzintervall geschätzt werden. Modifiziert nach Priebe et al., 2012.

4.3 Beteiligung von CNVs bei der Majorer Depression

4.3.1 Qualitätskontrollierter Datensatz

Nach Anwendung aller Filterkriterien (siehe Kap. 3.3.4 und 4.1) verblieben 575 Patienten mit Majorer Depression und 1.618 Kontrollen für die weiteren Analysen. Das durchschnittliche Alter der Patienten bei Rekrutierung war 47,6 Jahre (SA = 13,9), das durchschnittliche Erkrankungsalter 36,6 Jahre (SA = 13,3). Das durchschnittliche Alter der Kontrollen bei Rekrutierung lag bei 47,1 Jahren (SA = 12,7).

Nach Anwendung der in Kapitel 4.1 etablierten Filterkriterien, waren genomweite Daten von 2.193 Personen verfügbar. In diesen wurden 1.311 CNVs von beiden Programmen QuantiSNP und PennCNV detektiert. Dies entspricht einer Frequenz von etwa 0,6 CNVs pro Person.

4.3.2 Genomweite *burden* Tests

Keiner der genomweiten *burden* Tests zeigte eine signifikante Assoziation mit der Majorer Depression (siehe Anhang).

4.3.3 Spezifische Assoziationen

Zur Identifikation von Regionen, in denen bis dato keine CNVs mit der Majorer Depression assoziiert gefunden worden waren, wurden die genomweiten Daten von Patienten und Kontrollen statistisch analysiert. In vier spezifischen chromosomalen Regionen waren CNVs mit Majorer Depression assoziiert:

Assoziation von Deletionen in 7p21.3

Fünf Patienten (0,87 %) und drei Kontrollen (0,19 %) hatten eine Deletion in der chromosomalen Subbande 7p21.3 (P-Wert: 0,0331; OR: 4,72; 95-KI: 0,91 - 30,48). Die Deletionen in allen Patienten und in zwei Kontrollen hatten die gleichen potentiellen Bruchpunkte und überspannten 33 SNPs und etwa 95 kb (9.094.766-9.189.993, siehe Abb. 4.3). Die dritte Deletion in einer Kontrolle zeigte andere Bruchpunkte und überlappte nur teilweise mit den anderen Deletionen (8.923.566-9.160.323). Innerhalb der assoziierten Region war kein Gen lokalisiert. Alle potentiellen CNV-Träger konnten mit qRT-PCR bestätigt werden. Der abweichende 3'-Bruchpunkt in einer Kontrolle konnte ebenfalls bestätigt werden.

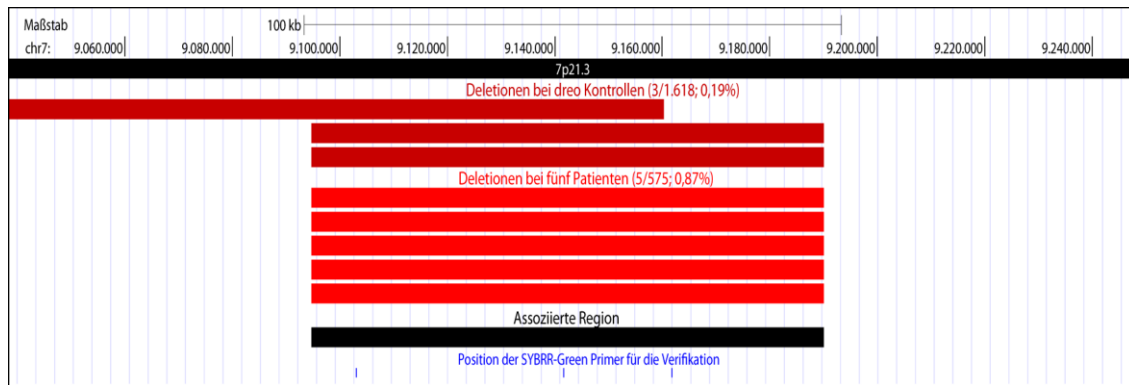


Abb. 4.3: CNVs in der chromosomalen Bande 7p21.3. Drei von 1.618 untersuchten Kontrollen (dunkel rote Balken) hatten eine Deletion in der Region. Bei fünf von 575 Patienten wurde eine Deletion detektiert (hell rote Balken). Die assoziierte Region (schwarzer Balken) überspannte etwa 95 kb. Für die Verifikation der CNVs wurden drei SYBR-Green Primer (blaue vertikale Striche) verwendet. Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Degenhardt und Priebe et al., 2012.

Assoziation von Duplikationen in 15q26.3

In der chromosomalen Subbande 15q26.3 wurden Duplikationen in fünf Patienten (0,87 %) und drei Kontrollen detektiert (0,19 %; P-Wert: 0,0331; OR: 4,72; 95-KI: 0,91 - 30,48). Alle Duplikationen unterschieden sich in ihren potentiellen Bruchpunkten, überlappten jedoch in der assoziierten Region von etwa 13 kb Länge (99.918.818-99.931.589, siehe Abb. 4.4). Von der geteilten Region war kein Gen direkt überspannt, allerdings waren die Gene *TM2 domain containing* und *threonyl-tRNA synthetase-like 2 (TM2D3)* und *Threonyl-TRNA Synthetase-Like 2 (TARSL2)* von Duplikationen in zwei Patienten und zwei Kontrollen betroffen. Die CNVs in allen acht Personen konnten erfolgreich mit qRT-PCR verifiziert werden.

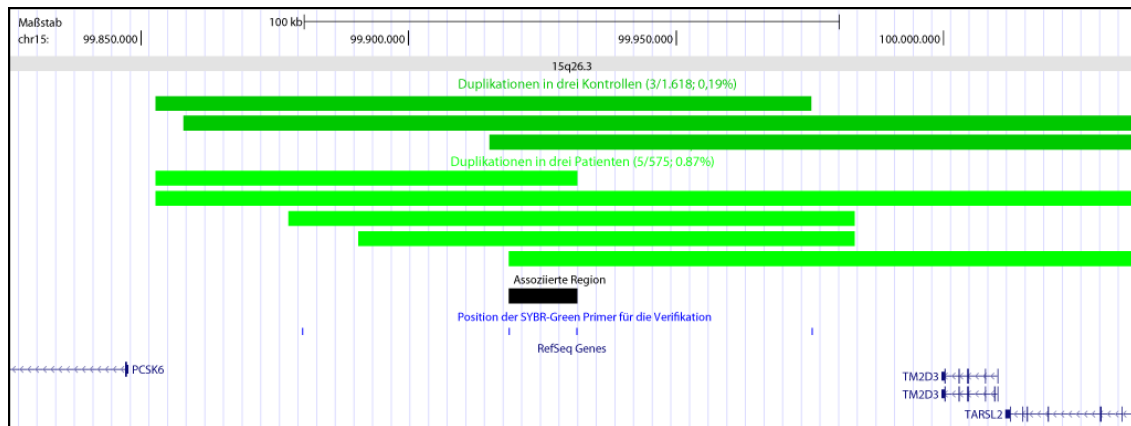


Abb. 4.4: CNVs in der chromosomalen Bande 15q26.3. Drei von 1.618 untersuchten Kontrollen (dunkel grüne Balken) hatten eine Duplikation in der Region. Bei fünf von 575 Patienten wurde eine Duplikation detektiert (hell grüne Balken). Alle Duplikationen hatten unterschiedliche Bruchpunkte. Die assoziierte Region (schwarzer Balken) überspannte etwa 13 kb. Für die Verifikation der CNVs wurden drei SYBR-Green Primer (blaue vertikale Striche) verwendet. Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Degenhardt und Priebe et al., 2012.

Assoziation von Deletionen und Duplikationen in 16p11.2

Zwei Patienten hatten eine Duplikation (0,35 %) und ein Patient eine Deletion (0,17 %) in der chromosomalen Subbande 16p11.2 (P-Wert: $\leq 0,0179$; OR: $\geq 8,48$ (wegen des fehlenden Ereignisses in Kontrollen kann kein 95-KI angegeben werden)). In dieser Region konnten bei den untersuchten Kontrollen ein CNV mit mehr als 30 SNPs nachgewiesen werden. Die assoziierte Region war etwa 530 kb lang (29.554.843-30.085.308, siehe Abb. 4.5) und betraf mehr als 20 Gene. Die in einem Patienten identifizierte Deletion war mit etwa 1,6 Mb (28.447.349-30.085.308) größer als die assoziierte Region hatte aber den gleichen 3'-Bruchpunkt. Alle CNVs konnten mit MLPA bestätigt werden.

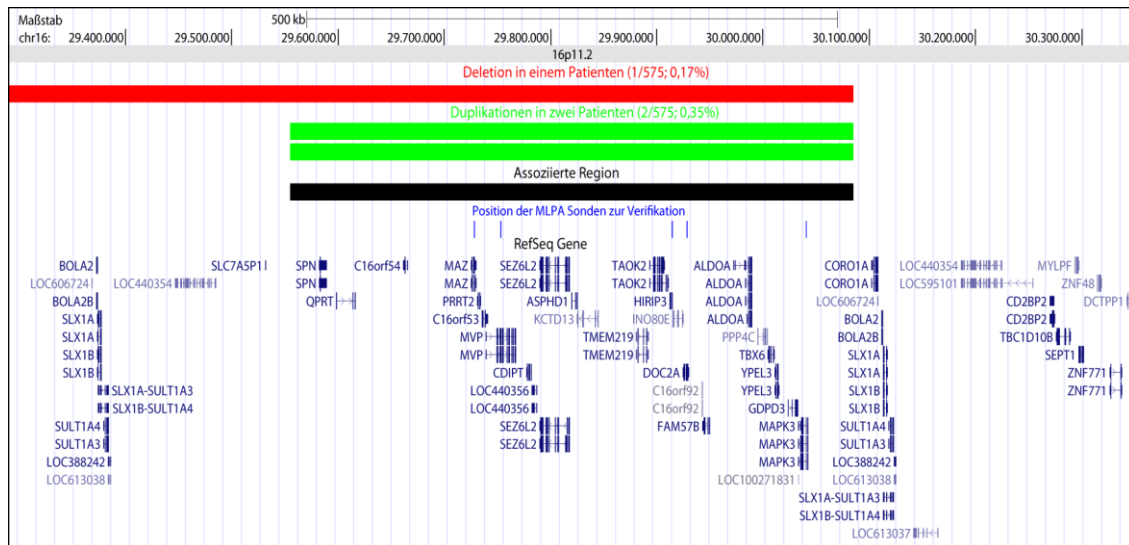


Abb. 4.5: CNVs in der chromosomalen Bande 16p11.2. In keiner der 1.618 untersuchten Kontrollen wurde ein CNV in der Region detektiert. Bei zwei von 575 Patienten wurde eine Duplikation detektiert (grüne Balken) und bei einem Patienten eine Deletion (roter Balken). Alle Duplikationen hatten die gleichen Bruchpunkte, die Deletion hatte einen abweichenden 3' Bruchpunkt. Die assoziierte Region (schwarzer Balken) überspannte etwa 530 kb. Für die Verifikation der CNVs wurden das MLPA Kit P297 verwendet, dieses hat fünf Sonden in der assoziierten Region (blaue vertikale Striche). Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Degenhardt und Priebe et al., 2012.

Assoziation von Deletionen in 18p11.32

Neun Patienten und neun Kontrollen hatten eine Deletion in der chromosomalen Subbande 18p11.32 (P-Wert: 0,0298; OR: 2,84; 95-KI: 0.99-8.12). Die Deletionen in Kontrollen zeigten alle die gleichen potentiellen Bruchpunkte, in Patienten waren vier Deletionen etwas kleiner. Die assoziierte Region überspannte 33 SNPs und etwa 118 kb (1.721.154-1.829.387, siehe Abb. 4.6). In der assoziierten Region war kein Gen lokalisiert. Alle CNVs konnten mit qRT-PCT bestätigt werden.



Abb. 4.6: CNVs in der chromosomalen Bande 18p11.32. Bei neun der 1.618 untersuchten Kontrollen (dunkelrote Balken) und bei neun von 575 Patienten wurde eine Deletion detektiert (hellrote Balken). Die Deletionen bei Patienten zeigten teilweise unterschiedliche Bruchpunkte. Die assoziierte Region (schwarzer Balken) überspannte etwa 121 kb. Für die Verifikation der CNVs wurden drei SYBR-Green Primer verwendet (blaue vertikale Striche). Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Degenhardt und Priebe et al., 2012.

4.3.4 Phänotypische Charakterisierung der 16p11.2-CNV-Träger

Die Assoziation von CNVs in 16p11.2 mit Majorer Depression war aus zwei Gründen besonders interessant: 1) war eine Assoziation dieser Region bereits mit unterschiedlichen Erkrankungen, darunter auch neuropsychiatrische, berichtet worden: Bipolare Störung (McCarthy et al., 2009; Grozeva et al., 2010), Schizophrenie (McCarthy et al., 2009; Levinson et al., 2011; Vacic et al., 2011), Autismus (Kumar et al., 2008; Weiss et al., 2008), Intelligenzminderung (Bijlsma et al., 2009) und 2) war dies die einzige Region, die zahlreiche Gene überspannte.

Zur weiteren phänotypischen Beschreibung der CNV-Träger wurden deren Krankenakten auf Informationen bezüglich der Eigen- und Familienanamnese eingesehen.

Patientin A, die Deletionsträgerin, hatte ihre erste depressive Episode mit 26 Jahren. Die Patientin besuchte die Hauptschule, sie erreichte dort einen Abschluss, hatte jedoch keine abgeschlossene Berufsausbildung. Der *body mass index* (BMI) der Patientin lag bei 43,71, gleichbedeutend mit der höchsten Stufe einer Adipositas (morbide Adipositas, Grad III). Die Familienanamnese der Patientin war positiv für neuropsychiatrische Erkrankungen; beide

Elternteile hatten eine gesicherte Major Depression und Angststörung. Für eine Tante väterlicherseits war die Verdachtsdiagnose einer Angststörung vermerkt.

Patientin B trug eine Duplikation, sie hatte ihre erste depressive Episode mit 37 Jahren. Die Patientin besuchte eine Sonderschule, blieb jedoch ohne Schulabschluss und Berufsausbildung. Die Schwester der Patientin hatte eine Bipolare Störung und ihre Großmutter mütterlicherseits die Verdachtsdiagnose für eine nicht näher bezeichnete neuropsychiatrische Störung. Der BMI dieser Patientin lag im Normbereich.

Patient C trug ebenfalls eine Duplikation, er hatte seine erste depressive Episode mit 43 Jahren. Der Patient hatte einen Sonderschulabschluss und eine abgeschlossene Ausbildung zum Kraftfahrzeugmechaniker. Seine Familienanamnese war leer für neuropsychiatrische Störungen. Der BMI dieses Patienten war unauffällig.

Alle Patienten hatten zum Zeitpunkt der Rekrutierung bereits mehrere depressive Episoden erlebt. Weiterhin erfüllte keiner der Patienten die DSM-IV Kriterien für eine Intelligenzminderung.

4.3.5 Frequenz von CNVs in vorbeschriebenen Regionen

Außer in der chromosomalen Subbande 16p11.2 (siehe Kap. 4.3.3) konnte in keiner der elf mit neuropsychiatrischen Erkrankungen vorbeschriebenen Regionen (siehe Kap. 3.5.2) eine Assoziation mit Majorer Depression im vorliegenden Datensatz gefunden werden. Nur in drei der elf Regionen wurden überhaupt CNVs in Patienten mit Majorer Depression gefunden (siehe Tab. 4.4).

Region	Position *	Patienten (575)	Kontrollen (1.618)	P	OR	OR-CI	P-dup	OR-dup	CI-dup	P-del	OR-del	CI-del
1q21.1	144.94-146.29	1 dup	1 del	0,46	2,82	0,04-220,87	≤ 0,26	≥ 2,81	***	1	**	**
2p16.3	49.90-51.50	3 del	5 del	0,44	1,69	0,26-8,73	--	--	--	0,44	1,69	0,26-8,73
3q29	197.2-198.83	--	--	--	--	--	--	--	--	--	--	--
7q36.3	158.51-158.63	--	--	--	--	--	--	--	--	--	--	--
15q11.2	20.31-20.78	2 dup, 1 del	3 dup, 3 del	0,71	1,41	0,23-6,62	0,61	1,88	0,16-16,44	1	0,94	0,02-11,71
15q13.3	28.72-30.30	--	1 dup	1	**	**	1	**	**	1	**	**
16p13.11	15.38-16.19	--	3 del	0,57	**	**	--	--	--	0,57	**	**
16p11.2	29.50-30.30	2 dup, 1 del	--	≤ 0,018	≥ 8,48	***	≤ 0,07	≥ 5,64	***	≤ 0,26	≥ 2,82	***
17p12	14.0-15.4	--	--	--	--	--	--	--	--	--	--	--
17q12	31.89-33.28	--	--	--	--	--	--	--	--	--	--	--
22q11.21	17.00-19.80	--	--	--	--	--	--	--	--	--	--	--
Summe	--	5 dup, 5 del	4 dup, 12 del	0,18	1,77	0,71-4,18	0,06	3,54	0,76-17,89	0,78	1,17	0,32-3,60

Tab. 4.4: CNVs in Regionen mit Vorbefunden bei neuropsychiatrischen Erkrankungen bei der Majoren Depression. CNVs im Datensatz für Majore Depression die mit vorbeschriebenen Regionen überlappten. Die Ergebnisse in der Region 16p11.2 (P-Wert fett) sind im Text genauer beschrieben. P – P-Wert; OR – odds ratio, OR-KI – 95%-Konfidenzintervall des OR; dup – Duplikationen; del – Deletionen; * Positionsangaben entsprechend NCBI build 36; ** Wegen fehlendem Ereignis in Patienten kann weder das OR noch das OR-KI geschätzt werden; *** Wegen fehlendem Ereignis in Kontrollen, kann das OR-KI nicht geschätzt werden. Modifiziert nach Degenhardt und Priebe et al., 2012.

4.4 Beteiligung von CNVs der Schizophrenie

4.4.1 Qualitätskontrollierter Datensatz

Insgesamt 1.637 Patienten und 1.627 Kontrollen verblieben nach Anwendung aller Filterkriterien. Die Patienten hatten ein durchschnittliches Alter bei Rekrutierung von 34,1 Jahren (SA = 11,0) und ein durchschnittliches Erkrankungsalter von 21,9 Jahren (SA = 6,5). Das durchschnittliche Alter der Kontrollen bei Rekrutierung lag bei 47,53 Jahren (SA = 12,2).

In diesem Kollektiv wurden drei Analysen durchgeführt, um die Beteiligung von CNVs bei der Entstehung der Schizophrenie zu untersuchen: i) genomweite *burden* Tests (Kap. 4.4.2), ii) Tests auf die Assoziation bis dato unbekannter spezifischer CNVs (Kap. 4.4.3) und iii) Überprüfung der Frequenz von CNV in Regionen, die zuvor als mit neuropsychiatrischen Erkrankungen assoziiert berichtet worden waren (siehe Kap. 4.4.4).

4.4.2 Genomweite *burden* Tests

Bei den genomweiten *burden* Tests zeigte sich eine keine signifikante Assoziation von CNVs mit der Schizophrenie. Die stärkste Assoziation wurde für die Frequenz von Deletionen gefunden, diese erreichte jedoch nicht das Signifikanzniveau von 0,05 (P-Wert: 0,07; siehe Anhang).

4.4.3 Spezifische Assoziationen

Die individuellen CNV-Daten von Patienten und Kontrollen wurden genomweit verglichen, um spezifische Regionen zu detektieren, in denen CNVs signifikant häufiger in Patienten mit einer Schizophrenie auftraten als in Kontrollen. Bei dieser Suche nach neuen spezifischen Regionen, die mit der Schizophrenie assoziiert sein könnten, zeigte sich kein nominell signifikantes Ergebnis.

4.4.4 Frequenz von CNVs in vorbeschriebenen Regionen

In elf Regionen war eine Assoziation von CNVs mit neuropsychiatrischen Erkrankungen bereits berichtet worden (siehe Kap. 3.5.2): 1q21.1, 2p16.3, 3q29, 7q36.3, 15q11.2, 15q13.3, 16p13.11, 16p11.2, 17p12, 17q12 und 22q11.21. Die Frequenz von CNVs, die zu mindestens 80 Prozent (vgl. Kap. 3.5.2) überlappten, wurde im vorliegenden Datensatz überprüft. In keiner der elf Regionen zeigte sich eine signifikante Assoziation von CNVs mit der

Schizophrenie (siehe Tab. 4.5). Bei 1.637 Patienten wurden 20 Deletionen und elf Duplikationen gefunden, die mit einer der elf Regionen überlappte. Bei 1.627 Kontrollen wurden zwölf Deletionen und fünf Duplikationen gefunden. Diese Überrepräsentation von CNVs war gerade nicht signifikant (P-Wert: 0,06; OR: 1,83). In den Regionen auf Chromosom 1q21.1 und 22q11.21, sowie in 2p16.3/*NRXN1* wurden häufiger CNVs in Patienten im Vergleich zu Kontrollen gefunden, auch dieser Unterschied war nicht signifikant (siehe Tab. 4.5; 1q21.1: P-Wert: 0,62; OR: 2,98; 22q11.2: P-Wert: 0,5; OR: > 1,99; 2p16.3/*NRXN1*: P-Wert: 0,21, OR: 2,19). Im Gegensatz dazu war die Frequenz von CNVs in der Region auf Chromosom 15q11.2 größer in Kontrollen als in Patienten (P-Wert: 0,37; OR: 0,33).

Region	Position *	Patienten (1.637)	Kontrollen (1.627)	P	OR	OR-CI	P-dup	OR-dup	CI-dup	P-del	OR-del	CI-del
1q21.1	144,94-146,29	3 del	1 del	0,62	2,98	0,24-156,7	0,62	2,98	0,24-156,7	--	--	--
2p16.3	49,90-51,50	1 dup, 10 del	5 del	0,21	2,19	0,70-8,07	0,30	1,99	0,62-7,45	1	>1	0,01-78,1
3q29	197,2-198,83	--	--	--	--	--	--	--	--	--	--	--
7q36.3	158,51-158,63	1 dup	--	1	>1	**	--	--	--	1	>1	**
15q11.2	20,31-20,78	4 dup, 2 del	4 dup, 3 del	0,58	0,71	0,18-2,60	0,37	0,33	0,01-4,13	1	0,99	0,18-5,35
15q13.3	28,72-30,30	1 dup	1 dup	1	0,99	0,01-78,02	--	--	--	1	0,99	0,01-78,0
16p13.11	15,38-16,19	2 dup, 3 del	3 del	0,73	1,66	0,32-10,70	0,50	>1,99	**	1	0,99	0,13-7,43
16p11.2	29,50-30,30	1 dup	--	1	>1	**	--	--	--	1	>1	**
17p12	14,0-15,4	--	--	--	--	--	--	--	--	--	--	--
17q12	31,89-33,28	1 dup	--	1	>1	**	--	--	--	1	>1	**
22q11.21	17,00-19,80	2 del	--	0,50	>1,99	**	0,50	>1,99	**	--	--	--
Summe	--	11 dup, 20 del	5 dup, 12 del	0,06	1,83	0,97-3,54	0,21	2,19	0,70-8,10	0,21	1,66	0,77-3,75

Tab. 4.5: CNVs in Regionen mit Vorbefunden bei neuropsychiatrischen Erkrankungen bei der Schizophrenie. CNVs im Datensatz für Schizophrenie die mit vorbeschriebenen Regionen überlappten. P – P-Wert; OR – odds ratio, OR-KI – 95%-Konfidenzintervall des ORs; dup – Duplikationen; del – Deletionen; * Positionsangaben in Mb entsprechend NCBI build 36; ** Wegen fehlendem Ereignis in Kontrollen, konnte das OR-KI nicht geschätzt werden.

5 Diskussion

Ziel der vorliegenden Arbeit war die Untersuchung der Bedeutung von Kopienzahlvarianten (CNVs, *copy number variants*) bei der Entstehung der neuropsychiatrischen Erkrankungen Bipolare Störung, Majore Depression und Schizophrenie. In einem ersten Schritt musste zunächst eine standardisierte Auswertestrategie zur genomweiten *in silico* Detektion von CNVs aus bereits vorhandenen Genotypisierungsdaten etabliert werden.

5.1 Etablierung der Auswertestrategie

Die Identifizierung von CNVs aus genomweiten Genotypisierungsdaten ist technisch anspruchsvoll und war zu Beginn der Arbeit in der Fachliteratur nicht standardisiert. Wie weiter unten detaillierter beschrieben, hatten die Strategien unterschiedliche CNV-Detektionsraten und nahmen entweder eine erhöhte Rate falsch-positiver oder falsch-negativer Ergebnisse in Kauf. Die Anwendung einer stringenten Auswertestrategie war daher notwendig, um die Anzahl von *in silico* detektierten, falsch-positiven CNVs auf ein Minimum zu verringern und gleichzeitig die Detektionsrate tatsächlich vorhandener CNVs nicht zu sehr abzuschwächen. In der vorliegenden Arbeit erfüllte die Kombination aus mindestens 30 vom CNV überspannten SNPs und einem *log Bayes Factor* (IBF) beziehungsweise *confidence value* (CV) von mindestens 30 die hohen Ansprüche an die geforderte Stringenz.

Seit Beginn der Arbeit wurden zahlreiche genomweite CNV Studien publiziert, deren Auswahlstrategien sich von denen der vorliegenden Arbeit unterschieden (u.a. Zhang et al., 2009; Grozeva et al., 2010; Rucker et al., 2011). Ein wesentliches Unterscheidungsmerkmal zur vorliegenden Arbeit war die Verwendung eines Längenfilters anstelle eines SNP-Filters. So schlossen beispielsweise Grozeva et al. und Rucker et al. alle CNVs mit einer Länge von weniger als 100 kb - unabhängig von der Anzahl der vom CNV überspannten SNPs - aus ihren Studien aus (Grozeva et al., 2010; Rucker et al., 2011).

Die SNPs auf den verwendeten Chips waren nicht gleichmäßig über das gesamte Genom verteilt. Die *in silico* Detektion von CNVs mit QuantiSNP und

PennCNV basiert jedoch auf den Daten jedes einzelnen genotypisierten SNPs. Je mehr SNPs eine bestimmte Region abdecken, desto mehr Datenpunkte stehen den Programmen für ihre Berechnungen zur Verfügung. Daher kann ein CNV in einer besonders dicht abgedeckten Region verlässlicher identifiziert werden. Als Beispiel: ein CNV mit einer Länge von 150 kb, der nur 17 SNPs betrifft, würde bei dem Filterkriterium „Mindestlänge von 100 kb“ eingeschlossen werden, während ein CNV mit 75 kb Länge und 34 überspannten SNPs ausgeschlossen würde. Wegen der höheren SNP-Dichte des zuletzt genannten CNVs handelt es sich bei diesem jedoch mit höherer Wahrscheinlichkeit um einen biologisch validen CNV.

Da nicht von allen untersuchten Personen genügend genomische DNA für eine technische Verifikation der identifizierten CNVs vorhanden war, wurde die Anwendung eines SNP-Filters gegenüber einem Längenfilter favorisiert. Alle CNVs, die die Kombination aus SNP-Filter und IBF/CV-Filter erfüllten, konnten mit einer unabhängigen, quantitativen Methode verifiziert werden. Allerdings werden durch diese stringente Kombination vermutlich auch viele kleinere, biologisch valide CNVs verworfen. Es ist davon auszugehen, dass die Weiterentwicklung der CNV-Identifizierungsmethoden, insbesondere durch neue Sequenzierverfahren, zukünftig eine valide Detektion dieser kleineren CNVs ermöglichen wird.

5.2 Bipolare Störung

Nach Anwendung aller Filterkriterien verblieben 882 Patienten mit Bipolarer Störung und 872 populationsbasierte Kontrollen im Datensatz. In dieser Stichprobe wurden die folgenden drei Analysen durchgeführt:

- i) Genomweiter burden der Länge und Frequenz von CNVs
- ii) Assoziation spezifischer Regionen
- iii) Frequenz von CNVs in vorbeschriebenen Regionen

Die Ergebnisse wurden bereits veröffentlicht (Priebe et al., 2012).

5.2.1 Genomweiter *burden*

Die burden Analyse erzielte drei signifikante Hauptergebnisse: (i) *singleton* Deletionen waren bei den Patienten im Vergleich zu den Kontrollprobanden

signifikant länger, (ii) in der Subgruppe der Patienten mit frühem Erkrankungsalter war der zuvor genannte Unterschied noch stärker ausgeprägt und (iii) die Frequenz von Duplikationen war bei den Patienten mit frühem Erkrankungsalter signifikant höher im Vergleich zur Kontrollstichprobe.

Die veröffentlichten Ergebnisse zum genomweiten burden von CNVs bei der Bipolaren Störung ergeben bisher kein einheitliches Bild. Während zwei Studien keinen statistischen Zusammenhang zwischen der Krankheit und CNVs nachweisen konnten (Grozeva et al., 2010; McQuillin et al., 2011), zeigten drei weitere Studien eine signifikant höhere Frequenz von CNVs bei Patienten im Vergleich zur jeweiligen Kontrollstichprobe (Zhang et al., 2009; Malhotra et al., 2011; Bergen et al., 2012).

Die im vorliegenden Datensatz identifizierte Assoziation der *singleton* Deletionen steht im Einklang mit anderen publizierten Studien (Zhang et al., 2009; Bergen et al. 2012).

Ebenfalls berichteten Zhang et al., dass in ihrem Kollektiv, bestehend aus 1.001 Patienten mit Bipolarer Störung und 1.034 Kontrollen, der beobachtete Frequenzunterschied für *singleton* Deletionen – analog zu den Ergebnissen der vorliegenden Arbeit – noch stärker in der Gruppe der Patienten mit frühem Erkrankungsalter war (Zhang et al., 2009). Diese Ergebnisse deuten auf eine Beteiligung von seltenen CNVs insbesondere bei der Entstehung der frühmanifesten Bipolaren Störung hin. Diese mögliche Schlussfolgerung wird unterstützt durch die Studie von Malhotra et al. (Malhotra et al., 2011). Im Gegensatz zu den zuvor genannten Studien verfolgte diese Arbeit einen Trio-Ansatz und fokussierte sich auf die Identifizierung von de novo CNVs. Insgesamt wurden 185 Eltern-Kind-Trios mit einer neuropsychiatrischen Diagnose (Indexpersonen an einer Bipolare Störung erkrankt) und 426 gesunde Eltern-Kind-Trios untersucht (Malhotra et al., 2011). Im Patientenkollektiv war die Frequenz von de novo CNVs (4,3 %) signifikant höher im Vergleich zu den Kontroll-Trios (0,9 %; Malhotra et al., 2011). Bei den Patienten mit frühem Erkrankungsalter war die Frequenz von de novo CNVs sogar noch höher als in der übrigen Patientenstichprobe (5,6 %; Malhotra et al., 2011).

Die in der vorliegenden Arbeit detektierte Assoziation zwischen Duplikationen und frühem Erkrankungsalter wurde in den Studien von Zhang et al. beziehungsweise Malhotra et al. nicht nachgewiesen. Grund hierfür war, dass diese Arbeiten nur *singleton* CNVs beziehungsweise *de novo* CNVs untersuchten (Zhang et al., 2009; Malhotra et al., 2011). Die genomweite Analyse aller Duplikationen, unabhängig von deren Frequenz, wurde nicht durchgeführt. Bergen et al. führten keine Subgruppenanalyse durch (Bergen et al. 2012).

In zwei weiteren Studien konnte keine signifikante Assoziation zwischen dem genomweiten burden und der Bipolaren Störung identifiziert werden (Grozeva et al., 2010; McQuillin et al., 2011). Potentielle Gründe für die von der vorliegenden Arbeit abweichenden Ergebnisse sind: (i) eine geringere Auflösung der verwendeten Detektionsmethoden, (ii) abweichende Filterkriterien, (iii) eine zu kleine Anzahl an untersuchten Patienten mit einem frühem Erkrankungsalter (iv) unerkannte biologische Unterschiede zwischen unseren untersuchten Patienten und den Patienten in den anderen Studien und (v) ein falsch-positiver Befund in unserer Studie. Grozeva et al. und McQuillin et al. verwendeten jeweils das GeneChip Human Mapping 500K Array Set der Firma Affymetrix (Santa Clara, Kalifornien, USA), welches eine geringere Auflösung, im Vergleich zu den in dieser Arbeit verwendeten Chips, hatte (Grozeva et al., 2010; McQuillin et al., 2011). Es ist möglich, dass dadurch CNVs übersehen wurden, die in der vorliegenden Arbeit detektiert werden konnten. Die Frequenz von CNVs war kein Filterkriterium dieser Arbeit. Grozeva et al. hingegen untersuchten alle CNVs deren Frequenz nicht größer als 1 % in dem untersuchten Kollektiv war (Grozeva et al. 2010).

Die Ergebnisse der vorliegenden Arbeit liefern weitere Evidenz für die Beteiligung von seltenen CNVs an der Entstehung der Bipolaren Störung. Es zeigte sich, dass diese CNVs insbesondere bei Patienten mit frühem Erkrankungsalter eine Rolle zu spielen scheinen. Dies wurde auch bereits in weiteren Studien gezeigt (Zhang et al., 2009; Malhotra et al., 2011). Der Befund einer stärkeren Beteiligung von CNVs bei der Bipolaren Störung mit frühem Erkrankungsalter könnte darauf hinweisen, dass diese klinische

Untergruppe auch einen anderen genetischen Hintergrund hat als Patienten mit späterem Erkrankungsalter.

5.2.2 Assoziation spezifischer Regionen

Spezifische Duplikationen in 6q27 (OR: 2,40) und in 10q11.22 (OR: 1,53) waren mit der Bipolaren Störung signifikant assoziiert.

Die Duplikationen auf Chromosom 6 waren mit einem besonders frühen Erkrankungsalter assoziiert und hatten unterschiedliche in silico identifizierte Bruchpunkte. Allen CNVs teilen jedoch eine gemeinsame Kernregion (ca. 248 kb). In dieser Kernregion lagen drei Gene: *kinesin family member 25 (KIF25)*, *FERM domain containing 1 (FRMD1)* wurden komplett überspannt, von *myeloid/lymphoid or mixed-lineage leukemia translocated to 4 (MLLT4)* war nur das 3'-Ende betroffen. Für keines der drei Gene wurde bisher ein Zusammenhang zu neuropsychiatrischen Phänotypen beschrieben, sie sind sowohl im embryonalen wie auch im adulten Gehirn nur schwach exprimiert. Ein direkter Zusammenhang zwischen den Duplikationen und der Entwicklung einer Bipolaren Störung ist daher derzeit nicht ersichtlich.

Interessanterweise war eine Häufung von Duplikationen in 6q27 bei Patienten mit Bipolarer Störung bereits in einer anderen Studie beobachtet worden. Yang et al. untersuchten eine Großfamilie der Amischen, bestehend aus drei Generationen mit 46 Familienmitgliedern. Insgesamt 16 Familienmitglieder waren an einer Bipolarer Störung und drei an einer Majorer Depression erkrankt (Yang et al., 2009). In dieser Familie untersuchten Yang et al. die Kosegregation von CNVs mit neuropsychiatrischen Erkrankungen (Yang et al., 2009). Bei elf Personen, davon acht mit einer psychiatrischen Störung, fanden sich Duplikationen in 6q27 mit ähnlichen Bruchpunkten, wie die in dieser Arbeit beschriebenen. Diese Studie gibt gemeinsam mit den Ergebnissen der vorliegenden Arbeit einen guten Hinweis darauf, dass Duplikationen in 6q27 eine Rolle bei der Entstehung der Bipolaren Störung spielen, auch wenn der zugrunde liegende Mechanismus bisher nicht identifiziert werden konnte. Theoretisch könnten die Duplikationen einen regulatorischen Effekt auf Gene haben, die außerhalb der durch den CNV abgedeckten Region liegen. So liegt das Gen *dapper, antagonist of beta-catenin homolog 2 (DACT2)* etwa 115 kb

distal von der assoziierten Region auf Chromosom 6q27 entfernt. *DACT2* ist Teil des Wnt-Signalwegs, welcher unter anderem die Neurogenese und Neuroprotektion reguliert. Somit wäre *DACT2* – sofern beispielsweise seine Expression durch den naheliegenden CNV beeinflusst wird – ein interessantes Kandidatengen bei zentralnervösen Erkrankungen wie der Bipolaren Störung.

Auch häufige Duplikationen in 10q11.22 waren signifikant mit der Bipolaren Störung assoziiert. Diese Assoziation war noch signifikanter in der Subgruppe der Patienten mit jungem Erkrankungsalter (OR: 2,79).

Das einzige Gen in dieser Region war *anthrax toxin receptor-like (ANTXRL)*. Bei diesem Gen handelt es sich um ein Pseudogen, über das nur wenig bekannt ist. Bisher wurde keine Assoziation dieses Genes mit einer neuropsychiatrischen oder somatischen Erkrankung beschrieben.

Auch diese Region wurde in der oben beschriebenen Familienstudie von Yang et al. detektiert (Yang et al., 2009). Insgesamt 18 Personen, davon neun mit einer Bipolaren Störung oder Majorer Depression, hatten eine Duplikation in 10q11.22 mit den gleichen potentiellen Bruchpunkten wie die in dieser Arbeit detektierten (Yang et al., 2009).

Die signifikante Assoziation von Duplikationen in 6q27 und 10q11.22 mit der Bipolaren Störung gehören zu den ersten spezifischen Assoziationen von CNVs bei der Bipolaren Störung. In beiden Fällen war die Assoziation in der Subgruppe von Patienten mit frühem Erkrankungsalter am stärksten. Zusammen mit dem Ergebnis der genomweiten burden Untersuchung könnte dies auf Unterschiede bei den genetischen Faktoren hinweisen, die bei Patienten mit frühem und spätem Erkrankungsalter eine Rolle spielen. Interessanterweise konnten auch klinische Studien bereits zeigen, dass Patienten mit einem frühem Erkrankungsalter eine andere familiäre Belastung haben als Patienten, die spät erkranken (Grigoriu-Serbanescu et al., 2001). Funktionelle Studien werden nötig sein, um die Wirkmechanismen der CNVs zu verstehen.

5.3 Majore Depression

Nach Anwendung aller Filterkriterien verblieben 575 Patienten mit Majorer Depression und 1.618 Kontrollen im Datensatz. In diesem Kollektiv wurden die drei in Kap. 5.2 genannten Analysen durchgeführt. Die Ergebnisse wurden bereits veröffentlicht (Degenhardt und Priebe et al., 2012).

5.3.1 Genomweiter *burden*

Keiner der genomweiten *burden* Tests war signifikant mit der Erkrankung assoziiert.

Bislang gibt es nur eine weitere Studie, welche den genomweiten *burden* von CNVs bei Patienten mit Majorer Depression untersucht hat (Rucker et al., 2011). Die Analyse von knapp 2.700 Patienten und 4.800 Kontrollen zeigte eine signifikant höhere Frequenz von Deletionen bei den Patienten im Vergleich zu den Kontrollen (OR: 1,25; Rucker et al., 2011).

Für die unterschiedlichen Ergebnisse der beiden Studien gibt es wahrscheinlich zwei Hauptgründe: (i) Anwendung voneinander abweichender Auswertestrategien und (ii) Analyse ungleicher Kollektivgrößen. In der Studie von Rucker et al. wurden im Vergleich zur vorliegenden Arbeit weniger stringente Filterkriterien angewendet (CNVs mussten mindestens 10 SNPs betreffen und 100 kb lang sein, eine Filterung nach IBF/CV fand nicht statt; Rucker et al., 2011). Diese weichen Auswahlkriterien führten dazu, dass CNVs bei der *burden* Analyse berücksichtigt wurden, welche in der vorliegenden Arbeit ausgeschlossen worden wären. Weiterhin hatte die von Rucker et al. beobachtete Assoziation eine geringe Effektstärke (OR: 1,25). Wahrscheinlich hatte das in dieser Arbeit verwendete Kollektiv aufgrund seiner begrenzten Größe zu wenig statistische Aussagekraft, um ähnliche Effekte zu identifizieren.

Weitere Untersuchungen in größeren, unabhängigen Kollektiven sind notwendig, um die genomweite Beteiligung von CNVs bei der Entstehung der Majoren Depression näher einschätzen zu können.

5.3.2 Assoziation spezifischer Regionen

Spezifische Deletionen in 7p21.3 (OR: 4,72), Duplikationen in 15q26.3 (OR: 4,72), CNVs in 16p11.2 (OR: $\geq 8,48$) und Deletionen in 18p11.32 (OR: 2,84) waren mit der Erkrankung Major Depression signifikant assoziiert.

Die CNVs in den assoziierten Regionen 7p21.3, 15q26.3 und 18p11.32 überspannten jeweils keine Gene. Weiterhin erbrachte eine Datenbankrecherche keine Hinweise darauf, dass häufige genetische Varianten (SNPs) in diesen Regionen mit neuropsychiatrischen Phänotypen assoziiert waren. Somit bleibt der potentielle Wirkmechanismus dieser CNVs derzeit unklar. Denkbar wäre, dass die Deletionen und Duplikationen einen regulatorischen Effekt auf Gene außerhalb der CNVs haben. Weitere Studien sind nötig, um die Assoziation von CNVs in diesen drei genomischen Regionen genauer einordnen zu können.

Einer Interpretation besser zugänglich sind hingegen die Deletion/Duplikationen in 16p11.2, da sie (i) > 25 Gene überspannen und (ii) CNVs in dieser spezifischen Region mit den neuropsychiatrischen Phänotypen Autismus (Kumar et al., 2008; Weiss et al., 2008), Intelligenzminderung (McCarthy et al., 2009; Bijlsma et al., 2009), Schizophrenie (McCarthy et al., 2009; Levinson et al., 2011; Vacic et al., 2011) und Bipolare Störung (McCarthy et al., 2009; Grozeva et al., 2010) assoziiert waren. Ebenfalls ist bekannt, dass Deletionen in 16p11.2 mit einer Makrocephalie und Adipositas (Bochukova et al., 2010; Jacquemont et al., 2011; Shinawi et al., 2010; Walters et al. 2010) während Duplikationen derselben Region mit einer Mikrocephalie und Untergewicht (Jacquemont et al., 2011; Shinawi et al., 2010) assoziiert sind.

Im Einklang mit der bisher publizierten Literatur hat die in unserer Stichprobe identifizierte Deletionsträgerin in 16p11.2 einen massiv erhöhten *body mass index*. Ebenfalls zeigen alle CNV-Träger dieser Arbeit ein gewisses Maß an pre-morbiden kognitiven Defiziten. Angaben zum Kopfumfang lagen uns nicht vor, so dass bezüglich dieses Merkmals keine Aussagen möglich sind.

Welches bzw. welche der mehr als 20 Gene in der Region 16p11.2 ursächlich für die beobachteten Phänotypen sind, ist bisher nicht vollständig aufgeklärt. In einer ersten funktionellen Studie zu CNVs in 16p11.2, konnten Golzio et al. in

Zebrafischen zeigen, dass die Überexpression des Gens *potassium channel tetramerisation domain containing 13 (KCTD13)* den Kopfumfang der Tiere signifikant verringerte während eine Inhibition desselben Gens zu einer signifikanten Zunahme des Kopfumfanges führte (Golzio et al., 2012). Da der Kopfumfang einen Rückschluss auf das Hirnvolumen ermöglicht, legt diese Studie nahe, dass *KCTD13* an der Entwicklung der assoziierten neuropsychiatrischen Phänotypen beteiligt ist.

In der vorliegenden Arbeit wurde zum ersten Mal eine signifikante Assoziation von CNVs in der chromosomalen Subregion 16p11.2 mit Majorer Depression berichtet. Dieses Ergebnis erweitert das Spektrum der Erkrankungen, bei denen CNVs in der Region 16p11.2 ein genetischer Risikofaktor sind.

Bislang waren nur zwei Studien publiziert, welche das genomweite Auftreten von spezifischen CNVs bei Patienten mit Majorer Depression untersuchten (Glessner et al., 2010; Rucker et al., 2011). In keiner dieser Studien wurde eine signifikante Assoziation von CNVs in 7p21.3, 15q26.3, 16p11.2 oder 18p11.32 berichtet. Möglicherweise wurde aber eine Assoziation der Regionen in diesen Studien übersehen, was unter anderem in den spezifischen Analysestrategien begründet liegen könnte:

Die Studie von Glessner et al. beschränkte sich auf die Analyse von CNVs, die ausschließlich im Patientenkollektiv identifiziert wurden (Glessner et al., 2010). In drei der vier (7p21.3, 15q26.3 und 18p11.32) in dieser Arbeit assoziierten Regionen trugen auch Kontrollen einen CNV und wären somit in der Studie von Glessner et al. ausgeschlossen worden. Im Kontrollkollektiv der vorliegenden Arbeit erfüllte kein CNV in der Region 16p11.2 die Filterkriterien. Allerdings wurde ein kleiner CNV mit 14 SNPs detektiert. Dieser CNV hätte die von Glessner et al. angewendeten Filterkriterien (mindestens 10 SNPs und eine Länge von mindestens 100 kb) erfüllt und die Region wäre wiederum ausgeschlossen worden.

In der Studie von Rucker et al. wurde nur die Frequenz von CNVs in sieben mit Schizophrenie assoziierten Regionen analysiert (Rucker et al., 2011). Ein Vergleich der Frequenz von CNVs in den Regionen auf 7p21.3, 15q26.3 und 18p11.32 war also nicht möglich. In 16p11.2 identifizierten Rucker et al. zwei

Deletionen und drei Duplikationen bei 2.723 Patienten (0,18 %) sowie drei Deletionen und eine Duplikation bei 4.828 populationsbasierten Kontrollen (0,08 %). Diese CNVs hatten ähnliche Bruchpunkte wie die CNVs der vorliegenden Arbeit. Die Überrepräsentation von CNVs in 16p11.2 bei Patienten mit Majorer Depression unterschritt zwar nicht die Signifikanzgrenze, was vermutlich in der begrenzten statistischen Aussagekraft der Stichprobe begründet liegt; in Zusammenschau mit den Vorbefunden in dieser chromosomalen Region liefern unsere Ergebnisse aber weitere Evidenz für eine Beteiligung von CNVs in 16p11.2 bei der Pathogenese der Majoren Depression.

5.4 Schizophrenie

Nach Anwendung aller Filterkriterien verblieben 1.637 Patienten und 1.627 Kontrollen im Datensatz. In diesem Kollektiv wurden die drei in Kap. 5.2 genannten Analysen durchgeführt. Die Ergebnisse wurden bereits veröffentlicht (Priebe et al., 2013). Keine dieser statistischen Tests war signifikant.

Allerdings war die Frequenz von CNVs in sieben der elf mittlerweile bekannten risikoassoziierten Regionen – analog zur publizierten Literatur – bei den Patienten im Vergleich zu den Kontrollen erhöht: 1q21.1, 2p15.3/NRXN1, 7q36.3, 16p13.11, 16p11.2, 17q12 und 22q11.21. Wenn eine statistische Signifikanz auch nicht erreicht wird, so liefert die Überrepräsentation dieser CNVs in unseren Patienten gegenüber Kontrollen dennoch unterstützende Hinweise auf einen Einfluss auf die Schizophrenie. Dass keine der vorbeschriebenen Regionen im vorliegenden Datensatz signifikant assoziiert war, ist wahrscheinlich in der zu geringen Teststärke der vorliegenden Arbeit begründet. Eine Abschätzung der statistischen Aussagekraft einer Stichprobe wäre allerdings a priori schwierig und mit großen Unsicherheiten behaftet gewesen, da die Frequenzschätzungen aus den ursprünglichen Publikationen wegen ihrer begrenzten Stichprobengrößen ungenau (möglicherweise unterschätzt) sind. Zudem ist nicht auszuschließen, dass es Populationsunterschiede in den CNV-Frequenzen gibt und dass auch Unterschiede in der Erhebung bzw. Diagnosestellung einen Einfluss auf die beobachteten CNV-Frequenzen haben könnten. Letztlich sind Daten aus

möglichst vielen Patienten- und Kontrollstichproben nötig, um zu genaueren Frequenzschätzungen zu bekommen. Sullivan et al. zeigte in einem neueren Übersichtsartikel über die aktuelle Literatur, dass die bisher bekannten Schizophrenie-assoziierten CNVs zusammen genommen bei Patienten mit einer Frequenz von maximal 0,3% auftreten (Sullivan et al., 2012), während sie in Kontrollen lediglich mit 0,04% beobachtet wurden.

Es ist wichtig, dass die Datenbasis noch größer wird. Daher wurden die Daten der vorliegenden Arbeit bereits publiziert und somit anderen Arbeitsgruppen zur Verfügung gestellt (Priebe et al., 2013). Außerdem sind die Daten Teil der weltweit größten CNV Untersuchungen innerhalb des Psychiatric Genomics Consortium (PGC; Sullivan, 2010). Das PGC ist die bislang größte internationale Kollaboration im Feld der neuropsychiatrischen Genetik.

5.5 Schlussfolgerungen

In der bislang veröffentlichten Literatur wurde von mehreren unabhängigen Arbeitsgruppen zum einen eine genomweite Häufung von CNVs (besonders Mikrodeletionen) sowie auch eine hochsignifikante Assoziation mehrerer spezifischer CNVs mit der Schizophrenie berichtet (Sullivan et al., 2012; Rucker und McGuffin, 2012). Auch die Befunde in unserer Stichprobe sind unterstützend, obwohl die implizierten Regionen aufgrund der begrenzten Stichprobengröße nicht signifikant wurden. Die Konsistenz der Befunde über insgesamt viele tausend Patienten und Kontrollen in verschiedenen Studien lassen sie als gesichert erscheinen.

Über die Beteiligung von CNVs bei der Bipolaren Störung (Sullivan et al., 2012; Rucker und McGuffin, 2012) und Majoren Depression (Rucker und McGuffin, 2012) ist man sich hingegen bisher nicht exakt im Klaren. Die Ergebnisse der vorliegenden Arbeit ergeben allerdings Hinweise auf eine Beteiligung von CNVs bei der Entstehung sowohl der Bipolaren Störung als auch der Majoren Depression.

In den untersuchten Stichproben zeigte sich bei der Bipolaren Störung eine signifikante Beteiligung sowohl von spezifischen CNVs, wie auch im genomweiten burden (siehe Kap. 5.2). Die Befunde sind aber im Vergleich mit anderen Studien nicht einheitlich. So gibt es Studien, die keine Beteiligung

feststellten (Grozeva et al., 2010; McQuillin et al., 2011), wie auch solche, in denen es Hinweise auf eine Beteiligung gab (Zhang et al., 2009; Malhotra et al., 2011; Bergen et al., 2012). Tatsächlich muss also die Rolle von CNVs bei der Bipolaren Störung in großen, unabhängigen Kollektiven geklärt werden. Nicht zuletzt vor dem Hintergrund der Ergebnisse dieser Arbeit sollte bei solchen Studien Information über das Ersterkrankungsalter einbezogen werden, da Patienten mit frühem und spätem Erkrankungsalter möglicherweise unterschiedliche Ätiologien haben.

Für die Major Depression gab es bislang nur eine Studie, die CNVs auf genomweiter Ebene untersucht haben (Rucker et al., 2011). Sowohl die Ergebnisse der vorliegenden Arbeit (siehe Kap. 5.3), wie auch die von Rucker et al. deuten auf eine Beteiligung von CNVs bei der Entstehung der Erkrankung hin. Allerdings sind auch hier weitere Untersuchungen in großen, unabhängigen Stichproben nötig um die Rolle von CNVs bei der Major Depression zu klären.

Die zuvor erwähnten größeren, unabhängigen Stichproben können wahrscheinlich nur durch internationale Kollaborationen zusammengetragen werden. Zu schwierig ist es für einen einzelnen klinischen Standort, große Zahlen von Patienten und Kontrollpersonen in einem überschaubaren Zeitraum für Studien zu rekrutieren. Aus diesem Grund gibt es mittlerweile das *Psychiatric Genomics Consortium* (PGC; Sullivan, 2010). In diesem Zusammenschluss internationaler Forschungsgruppen werden mehrere zehntausend Patienten mit neuropsychiatrischen Erkrankungen analysiert. Unter anderem sind auch die in der vorliegenden Arbeit untersuchten Stichproben Teil des PGCs. Eine weitere Einschränkung der bislang veröffentlichten Studien war auch die Verwendung einer Chip-basierten Detektion von CNVs. Dadurch können sehr kleine CNVs nicht mit ausreichender Verlässlichkeit detektiert werden. Aussagen zur genomweiten Beteiligung solcher kleiner CNVs bei der Entstehung der Erkrankungen konnten daher bislang nicht gemacht werden. Durch die Weiterentwicklung der Sequenziermethoden der zweiten Generation wird die Analyse dieser CNVs jedoch bald ebenfalls möglich sein.

6 Zusammenfassung

Die neuropsychiatrischen Erkrankungen Bipolare Störung, Majore Depression und Schizophrenie gehören zu den genetisch komplexen Erkrankungen. Zu Beginn dieser Arbeit gab es überraschende Befunde, dass spezifische, seltene Kopienzahlvarianten (CNVs) höherpenetrante Risikofaktoren für die Schizophrenie sind (u.a. Karayiorgou et al., 1995; Stefansson et al., 2008). Die Relevanz von CNVs für die Entstehung der Bipolaren Störung und Majoren Depression war hingegen noch kaum bzw. gar nicht untersucht (Zhang et al., 2009).

Ziel dieser Arbeit war es daher, die Beteiligung von CNVs bei der Entwicklung der Bipolaren Störung und Majoren Depression erstmalig zu untersuchen und auch weitere genetische Evidenz für die Relevanz von CNVs bei der Schizophrenie zu identifizieren.

Die Identifizierung von CNVs aus SNP-Array-Daten ist technisch anspruchsvoll und war zu Beginn der Arbeit noch kaum standardisiert. In einem ersten Schritt wurde daher eine valide und replizierbare Auswertestrategie zur genomweiten *in silico* Detektion von CNVs etabliert. In einem zweiten Schritt wurde diese auf drei Kollektive angewendet: (i) 882 Patienten mit Bipolarer Störung und 872 Kontrollen, (ii) 575 Patienten mit einer Majoren Depression und 1.618 Kontrollen und (iii) 1.637 Patienten mit einer Schizophrenie und 1.627 Kontrollen.

Die Analyse der Kohorte mit Bipolarer Störung wies darauf hin, dass CNVs - insbesondere in der Subgruppe der Patienten mit frühem Erkrankungsalter - zur genetischen Architektur der Erkrankung beitragen: unter anderem waren spezifische Duplikationen in 6q27 (P-Wert: 0,004) und die Frequenz aller genomweit identifizierten Duplikationen (P-Wert: 0,008) ausschließlich in dieser Subgruppe mit der Erkrankung assoziiert. Weiterhin waren spezifische Duplikationen in 10q11.22 (P-Wert: 0,035) signifikant häufiger im Patientenkollektiv nachweisbar.

Bei den Patienten mit Majorer Depression waren CNVs in vier spezifischen Regionen mit der Erkrankung assoziiert: (i) Deletionen in 7p21.3 (P-Wert:

0,033), (ii) Duplikationen in 15q26.3 (P-Wert: 0,033), (iii) Deletionen und Duplikationen in 16p11.2 (P-Wert: $\leq 0,018$) und (iv) Deletionen in 18p11.32 (P-Wert: 0,03).

Im Schizophrenie-Kollektiv wurde eine Häufung von Duplikationen und Deletionen in den mittlerweile elf bekannten Risiko-assoziierten chromosomalen Regionen identifiziert. Keine dieser Regionen war in diesem Datensatz statistisch signifikant assoziiert, wahrscheinlich aufgrund einer für die geringe Frequenz dieser CNVs zu geringen Teststärke des Kollektives. Der vorliegende Datensatz liefert jedoch weitere, unabhängige Evidenz für die Relevanz der elf Risiko-assoziierten Regionen.

In den drei in dieser Arbeit untersuchten Kollektiven zeigten sich Hinweise für eine Beteiligung von CNVs bei der Entstehung der jeweiligen neuropsychiatrischen Erkrankung. Die Frequenz dieser CNVs war zum Teil sehr gering und die statistische Aussagekraft der untersuchten Stichproben daher begrenzt. Diese müssen in deutlich größeren Kollektiven überprüft werden. Zudem waren die hier untersuchten CNVs, technisch bedingt relativ groß, durch nun verfügbare Methoden ist auch die Untersuchung deutlich kleinerer CNVs möglich, welche ebenfalls eine Rolle in der Krankheitsentstehung haben könnten.

7 Ausblick

In der vorliegenden Arbeit wurde die Beteiligung von CNVs bei der Bipolaren Störung, der Majoren Depression und der Schizophrenie untersucht. Unter anderem wurden spezifische Regionen identifiziert, in denen CNVs mit der Bipolaren Störung beziehungsweise der Majoren Depression assoziiert waren. Bei der Schizophrenie zeigte sich eine Überrepräsentation von CNVs in Regionen, die bereits mit neuropsychiatrischen Erkrankungen assoziiert gefunden worden waren.

Um diese Ergebnisse zu bestätigen, ist eine Replikation in unabhängigen Kollektiven nötig. Diese Replikation könnte beispielsweise im Kollektiv des *Psychiatric Genomics Consortium* (PGC; Sullivan, 2010) erfolgen. Beim PGC handelt es sich um eine internationale Kollaboration, bei der genetische Daten von mehreren zehntausend Patienten mit neuropsychiatrischen Erkrankungen zusammengefasst werden. Ein Teil der vom PGC durchgeführten Untersuchungen ist die genomweite Identifikation von CNVs. Die in der vorliegenden Arbeit untersuchten Kollektive sind Teil des PGCs, eine Replikation in den übrigen Kollektiven wäre jedoch möglich. Damit würde die Replikation in den derzeit größten verfügbaren Datensätzen für neuropsychiatrische Erkrankungen erfolgen.

Bislang wurden für die Identifizierung von CNVs meist Chip-basierte Methoden verwendet. Allerdings sind mit diesen Methoden sehr kleine CNVs nur schwer detektierbar. Daher konnten in bisherigen Studien keinerlei Aussagen zur genomweiten Beteiligung solcher kleinen CNVs bei der Entstehung der Erkrankungen gemacht werden. Ein weiteres Manko der genomweiten CNV Untersuchungen mit Hilfe von Chip-basierten Methoden war die Bruchpunktbestimmung der CNVs. Momentan ist die exakte Bestimmung der Bruchpunkte noch schwierig. Allerdings wird diese durch die Weiterentwicklung der Sequenziermethoden der zweiten Generation und durch neue Methoden bald möglich sein.

Außerdem muss die Wirkungsweise eines CNVs bei einer Erkrankung geklärt werden. Durch Expressionsuntersuchungen kann geklärt werden ob und in

welchem Maße ein CNV die Expression eines oder mehrerer Gene verändert. Diese Gene wären dann mögliche Kandidatengene für weitere klinische Untersuchungen.

Neben Expressionsuntersuchungen bieten Studien an Tiermodellen eine weitere Möglichkeit den Effekt einzelner CNVs auf den Phänotypen zu untersuchen. Dieser Ansatz wurde bereits erfolgreich angewendet um die Rolle von CNVs in der chromosomalen Region 16p11.2 genauer zu verstehen (Golzio et al., 2012).

Die Identifikation von Regionen, in denen CNVs spezifisch mit einer Erkrankung assoziiert sind und die Aufklärung der zugrunde liegenden Mechanismen bietet auch Potential für eine Anwendung in der Klinik und damit auch einen Nutzen für die Patienten. Durch die Identifikation der CNVs beziehungsweise der Gene, die der Assoziation zu Grunde liegen, könnten neue Therapiemöglichkeiten gefunden werden. So konnten zum Beispiel bereits die Wutausbrüche eines Patient mit einer Deletion in der chromosomalen Region 15q13.3 erfolgreich durch die Gabe des Medikaments Galantamin, dass die Reduzierung eines Neurotransmitters durch die Deletion kompensiert, behandelt werden (Cubells et al., 2011). Weitere potentielle Ziele von Medikamenten für die Behandlung von neuropsychiatrischen Erkrankungen könnten durch CNV Studien identifiziert werden. Zum Beispiel kodiert das Gen *VIPR2* für einen G-Protein-gekoppelten Rezeptor, dessen Überexpression einen direkten Einfluss auf die Pathophysiologie der Schizophrenie haben könnte.

Ein weiterer Aspekt, der weiterer Studien bedarf ist die Assoziation spezifischer CNVs mit unterschiedlichen Erkrankungen. So sind zum Beispiel CNVs in der chromosomalen Region 16p11.2 unter anderem mit Autismus (u.a. Weiss et al., 2008), Bipolarer Störung (u.a. McCarthy et. al., 2009), Majorer Depression (Degenhardt und Priebe et al., 2012), Intelligenzminderung (Bijlsma et al., 2009) oder Schizophrenie (u.a. McCarthy et. al., 2009) assoziiert gefunden worden. Zu klären wäre jetzt ob diese Erkrankungen einen zugrunde liegenden intermediären Phänotypen teilen oder ob vielleicht pleiotrope Effekte im Zusammenspiel mit anderen krankheitspezifischen Faktoren zur Ausbildung der jeweiligen Erkrankung führen. Eine weitere Erklärung wäre, dass die

diagnostische Abgrenzung der neuropsychiatrischen Erkrankungen nicht immer klar ist. Dies wurde auch in der aktuellen Auflage des *Diagnostic and Statistical Manual of Mental Disorders* vom Mai 2013 berücksichtigt.

Das genaue Verständnis der Wirkmechanismen von CNVs in den einzelnen Erkrankungen könnte schließlich dazu führen, dass CNVs Teil der Diagnostik von neuropsychiatrischen Erkrankung werden. Sobald genügend häufige und seltene Risikofaktoren identifiziert wurden, die zusammen einen substantiellen Teil der Heritabilität erklären, können diese analysiert werden, um eine Diagnose zu beschleunigen. Für die Patienten würde diese schnellere Diagnose auch einen frühzeitigen Beginn einer möglichen Therapie bedeuten.

8 Abbildungs- und Tabellenverzeichnis

8.1 Abbildungen

Abb. 1.1: Schematische Darstellung von CNVs. In der Referenzsequenz liegt ein DNA-Segment (grüne Kästen) in zwei Kopien vor (1). Bei einer homozygoten Deletion (2) fehlen beide, bei einer heterozygoten Deletion (3) fehlt eine Kopie des betroffenen DNA-Segments (In diesen Beispielen Segment B). Bei Amplifikationen in die Kopienzahl des DNA-Segments erhöht, z.B. liegen drei Kopien vor (4). 3

Abb. 1.2: Schematische Darstellung der *nonallelic homologous recombination*. Zwischen DNA-Segmenten mit hoher Homologie, den segmentalen Duplikationen (SD) kann es zu einer Rekombination kommen. Sind diese segmentalen Duplikationen in die gleiche Richtung ausgerichtet, entsteht eine Deletion und eine Duplikation. Die DNA-Abschnitte sind mit A, B und C bezeichnet. Hier ist der Abschnitt B von NAHR betroffen. Die homologen Chromosome sind grün bzw. rot markiert. (modifiziert nach Malhotra und Sebat, 2012). 7

Abb. 1.3: Schematische Darstellung des *nonhomologous DNA end joinings*. Bei der Reparatur von Doppelstrangbrüchen durch NHEJ kann es zu Fehlern mit verschiedenen Resultaten kommen: A: korrekte Reparatur; B: Deletion durch Rekombination zwischen Mikrohomologien (MH). Farbige vertikale Striche repräsentieren Basenpaare (modifiziert nach Malhotra und Sebat, 2012). 8

Abb. 1.4: Schematische Darstellung des *Fork Stalling and Template Switching*. Stößt die Replikationsgabel auf einen Einzelstrangbruch (blauer Stern) dann kommt es zu einem Stopp der Replikation. Die aktive Replikationsgabel kann durch Anlagern an Mikrohomologien (MH) an anderer Stelle die Replikation fortsetzen. Die DNA-Abschnitte zwischen den Mikrohomologien sind mit A, B, C und D gekennzeichnet. Je nach Lokalisation der Mikrohomologien entsteht eine Duplikation oder Deletion (modifiziert nach Malhotra und Sebat, 2012). 9

Abb. 1.5: Schematische Darstellung der Funktion eines Retrotransposons. Ein aktives Retrotransposon wird zunächst translatiert und transkribiert. An einer geeigneten Zielsequenz wird die, durch reverse Transkription entstandene Retrotransposon-DNA eingefügt. Die dafür nötigen Enzyme (eine reverse Transkriptase und eine Endonuklease) sind auf dem Retrotransposon selbst kodiert. Das ursprüngliche Retrotransposon verbleibt unverändert in der DNA, durch die Retrotransposition wird seine Kopienzahl erhöht (modifiziert nach Malhotra und Sebat, 2012). 10

Abb. 1.6: Mögliche Auswirkungen von CNVs auf die Genexpression 1) Ist ein komplettes Gen (Oranges/braunes Rechteck) von einem CNV (grüne, gestrichelte Linie) überspannt, ist die Zahl der funktionellen Genkopien erhöht (bei einer Duplikation) oder verringert (bei einer Deletion). 2) Ist ein Gen nur teilweise betroffen, verringert sich die Zahl der aktiven Genkopien. 3) Liegen die Bruchpunkte eines CNVs in zwei Genen kann ein Fusionsgen entstehen. 4) Wenn ein regulatorisches Element (braun-weißes Rechteck) von einem CNV überspannt wird, kann dies einen Einfluss auf die Expression eines oder mehrerer Gene haben. 5) Eine Deletion kann dazu führen, dass eine rezessive Mutation oder ein funktioneller Polymorphismus demaskiert wird, wenn das kompensierende Allel deletiert wird. Abbildung modifiziert nach: Lupski und Stankiewicz, 2005. 12

Abb. 3.1: Schematische Darstellung des Infinium Assays. a: Die amplifizierte und fragmentierte DNA bindet an mit spezifischen Sonden gekoppelte *beads* (hier dargestellt durch rote, grüne und orangene Halbkugeln). Anschließend erfolgt eine immunhistochemische Sandwichfärbung. b: Beim Infinium I Design werden SNPs mit zwei verschiedenen Sonden detektiert. Jede der Sonden bindet spezifisch an eines der beiden Allele des SNPs. Nur bei perfektem Übereinstimmen der Sonde mit der Zielsequenz wird eine Fluoreszenz gemessen. c: Für einen Großteil der untersuchten SNPs ist nur ein Sondentyp nötig. Bei diesem Infinium II Design können zwei unterschiedliche Farbstoffe durch eine Einzebasenverlängerung an die Sonde binden (Abbildung entnommen aus: Steemers und Gunderson, 2007). 25

Abb. 3.2: Schematischer Ablauf der MLPA. Die MLPA erfolgte in fünf Schritten: 1) Die DNA wurde zunächst denaturiert; 2) an die einzelsträngige DNA der Zielregion (rote Balken) hybridisierten die beiden Teile der Sonde mit gleichbleibenden (hellblaue Balken), sequenzspezifischen (grüne Balken) und sondenspezifischen (dunkelblaue Balken) Abschnitten; 3) die beiden Teile der Sonde wurden durch ein Ligase verbunden; 4) alle verwendeten Sonden wurden durch eine PCR amplifiziert; 5) die PCR-Produkte wurden durch Kapillarelektrophorese aufgetrennt (nicht dargestellt, Abbildung modifiziert nach Schouten et al., 2002). 29

Abb. 3.3: Schematischer Ablauf der qRT-PCR mit TaqMan-Sonde. FRET – Fluoreszenz-Resonanzenergietransfer, R – Reporter, Q – Quencher (modifiziert nach <http://www.appliedbiosystems.com>) 31

Abb. 3.4: Beispiel für eine Deletion in GenomeStudio. Jeder blaue und rote Punkt repräsentiert einen SNP. Das *Log R Ratio* (A) und die *B Allele Frequency* (B) aller SNPs auf Chromosom 7 einer Person sind auf der Ordinate aufgetragen. Auf der Abszisse ist die chromosomale Position der Marker aufgetragen. Man erkennt deutlich ein LRR kleiner als 0 und zwei BAF-Cluster bei 0 und 1 für die roten SNPs bei circa 145 Mb. An dieser Stelle liegt eine heterozygote Deletion vor. 39

Abb. 3.5: Beispiel für eine Duplikation in GenomeStudio. Jeder blaue und rote Punkt repräsentiert einen SNP. Das *Log R Ratio* (A) und die *B Allele Frequency* (B) aller SNPs auf Chromosom 2 einer Person sind auf der Ordinate aufgetragen. Auf der Abszisse ist die chromosomale Position der Marker aufgetragen. Man erkennt deutlich ein LRR größer als 0 und vier BAF-Cluster bei circa 0, 0,33, 0,66 und 1 für die roten SNPs bei circa 125 Mb. An dieser Stelle liegt eine Duplikation vor. 39

Abb. 3.6: Schematische Darstellung einer einfachen Markovkette. Die verborgenen Zustände des Markov Modells (X_1, X_2) können nicht direkt erkannt werden, man beobachtet anstatt dessen die Ausgabe (Y_1, Y_2), die mit einer Emissionswahrscheinlichkeit (B_1, B_2) entsteht. Jeder Zustand X hat eine Übergangswahrscheinlichkeit (A_{12}, A_{21}) in einen anderen Zustand zu wechseln. Modifiziert nach Rabiner, 1989. 41

Abb. 3.7: Schematische Abbildung der möglichen Überlappung von CNVs. 46

Abb. 3.8: Klassen der CNVs, in denen die Analysen durchgeführt wurden. 57

Abb. 4.1: CNVs in der chromosomalen Bande 6q27. Bei 872 untersuchten Kontrollen (dunkel grüne Balken) wurden 22 Duplikationen mit sechs unterschiedlichen Bruchpunkten detektiert. In 291 Patienten mit frühem Erkrankungsalter (hell grüne Balken) wurden 17 Duplikationen detektiert. Die assoziierte Region (schwarzer Balken) überspannte etwa 248 kb. Für die Verifikation der CNVs wurden drei TaqMan-Sonden (blaue vertikale Striche) verwendet. Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Priebe et al., 2012. 62

Abb. 4.2: CNVs in der chromosomalen Bande 10q11.22. Bei 872 untersuchten Kontrollen (dunkel grüne Balken) wurden 32 Duplikationen detektiert. In 882 Patienten wurden 53 Duplikationen detektiert, davon 28 in Patienten mit frühem Erkrankungsalter (hell grüne Balken). Alle detektierten Duplikationen hatten die gleichen Bruchpunkte. Die assoziierte Region (schwarzer Balken) überspannte etwa 160 kb. Für die Verifikation der CNVs wurden drei TaqMan-Sonden (blaue vertikale Striche) verwendet. Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Priebe et al., 2012. 63

Abb. 4.3: CNVs in der chromosomalen Bande 7p21.3. Drei von 1.618 untersuchten Kontrollen (dunkel rote Balken) hatten eine Deletion in der Region. Bei fünf von 575 Patienten wurde eine Deletion detektiert (hell rote Balken). Die assoziierte Region (schwarzer Balken) überspannte etwa 95 kb. Für die Verifikation der CNVs wurden drei SYBR-Green Primer (blaue vertikale Striche) verwendet. Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Degenhardt und Priebe et al., 2012. 66

Abb. 4.4: CNVs in der chromosomalen Bande 15q26.3. Drei von 1.618 untersuchten Kontrollen (dunkel grüne Balken) hatten eine Duplikation in der Region. Bei fünf von 575 Patienten wurde eine Duplikation detektiert (hell grüne Balken). Alle Duplikationen hatten unterschiedliche Bruchpunkte. Die assoziierte Region (schwarzer Balken) überspannte etwa 13 kb. Für die Verifikation der CNVs wurden drei SYBR-Green Primer (blaue vertikale Striche) verwendet. Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Degenhardt und Priebe et al., 2012. 67

Abb. 4.5: CNVs in der chromosomalen Bande 16p11.2. In keiner der 1.618 untersuchten Kontrollen wurde ein CNV in der Region detektiert. Bei zwei von 575 Patienten wurde eine Duplikation detektiert (grüne Balken) und bei einem Patienten eine Deletion (roter Balken). Alle Duplikationen hatten die gleichen Bruchpunkte, die Deletion hatte einen abweichenden 3' Bruchpunkt. Die assoziierte Region (schwarzer Balken) überspannte etwa 530 kb. Für die Verifikation der CNVs wurden das MLPA Kit P297 verwendet, dieses hat fünf Sonden in der assoziierten Region (blaue vertikale Striche). Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Degenhardt und Priebe et al., 2012. 68

Abb. 4.6: CNVs in der chromosomalen Bande 18p11.32. Bei neun der 1.618 untersuchten Kontrollen (dunkel rote Balken) und bei neun von 575 Patienten wurde eine Deletion detektiert (hell rote Balken). Die Deletionen bei Patienten zeigten teilweise unterschiedliche Bruchpunkte. Die assoziierte Region (schwarzer Balken) überspannte etwa 121 kb. Für die Verifikation der CNVs wurden drei SYBR-Green Primer verwendet (blaue vertikale Striche). Die Abbildung wurde mit dem UCSC *Genome Browser* basierend auf dem NCBI build 36 erstellt. Modifiziert nach Degenhardt und Priebe et al., 2012. 69

8.2 Tabellen

Tab. 3.1: Charakteristika der verwendeten BeadArrays. 26

Tab. 3.2: Die verborgenen Zustände der Markovketten von QuantiSNP und PennCNV und die entsprechende Kopienzahl (nach Colella et al., 2007). 41

Tab. 3.3: Übersicht über die untersuchten Kollektive. Die angegebene Anzahl der Personen ist vor Anwendung von Filterkriterien. Abkürzungen: HH550 - HumanHap550v3, H610Q - Human610-Quadv1, H660W - Human660W-Quadv1 55

Tab. 3.4: Übersicht bekannter Kandidatenregionen. Position in Mb nach NCBI build 36 58

Tab. 4.1: Übersicht detektierter CNVs und Anteil der CNVs, die in beiden Replikaten detektiert wurden, bei unterschiedlichen Filterkombinationen. IBF – log Bayes Factor, CV – confidence value 59

Tab. 4.2: Übersicht der burden Tests bei Bipolarer Störung. Es werden nur die P-Werte kleiner als 0,1 dargestellt, die Gesamtergebnisse befinden sich im Anhang. Die fett markierten P-Werte werden im Text genauer beschrieben. * Dieser P-Wert blieb nach Korrektur für multiples Testen signifikant (korrigierter P-Wert: 0,0288). Modifiziert nach Priebe et al., 2012. 61

Tab. 4.3: CNVs in Regionen mit Vorbefunden bei neuropsychiatrischen Erkrankungen bei der Bipolaren Störung. CNVs im untersuchten Kollektiv für Bipolare Störung die mit vorbeschriebenen Regionen überlappten. P – P-Wert; OR – *odds ratio*, OR-KI – 95%-Konfidenzintervall des ORs; dup – Duplikationen; del – Deletionen; * Positionsangaben entsprechend NCBI build 36; ** Wegen fehlendem Ereignis in Patienten kann weder das OR noch das Konfidenzintervall geschätzt werden. Modifiziert nach Priebe et al., 2012. 64

Tab. 4.4: CNVs in Regionen mit Vorbefunden bei neuropsychiatrischen Erkrankungen bei der Majoren Depression. CNVs im Datensatz für Majore Depression die mit vorbeschriebenen Regionen überlappten. Die Ergebnisse in der Region 16p11.2 (P-Wert fett) sind im Text genauer beschrieben. P – P-Wert; OR – *odds ratio*, OR-KI – 95%-Konfidenzintervall des OR; dup – Duplikationen; del – Deletionen; * Positionsangaben entsprechend NCBI build 36; ** Wegen fehlendem Ereignis in Patienten kann weder das OR noch das OR-KI geschätzt werden; *** Wegen fehlendem Ereignis in Kontrollen, kann das OR-KI nicht geschätzt werden. Modifiziert nach Degenhardt und Priebe et al., 2012. 71

Tab. 4.5: CNVs in Regionen mit Vorbefunden bei neuropsychiatrischen Erkrankungen bei der Schizophrenie. CNVs im Datensatz für Schizophrenie die mit vorbeschriebenen Regionen überlappten. P – P-Wert; OR – *odds ratio*, OR-KI – 95%-Konfidenzintervall des ORs; dup – Duplikationen; del – Deletionen; * Positionsangaben in Mb entsprechend NCBI build 36; ** Wegen fehlendem Ereignis in Kontrollen, konnte das OR-KI nicht geschätzt werden. 74

9 Literaturverzeichnis

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell*, 4th edition, *Garland Science*.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders 4th Edition March 2007.
- Bergen, S. E., O'Dushlaine, C. T., Ripke, S., Lee, P. H., Ruderfer, D. M., Akterin, S., Moran, J. L., Chambert, K. D., Handsaker, R. E., Backlund, L., Osby, U., McCarroll, S., Landen, M., Scolnick, E. M., Magnusson, P. K. E., Lichtenstein, P., Hultman, C. M., Purcell, S. M., Sklar, P., & Sullivan, P. F. (2012). Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Molecular psychiatry*, 17(9):880-6.
- Bijlsma, E. K., Gijsbers, A. C. J., Schuurs-Hoeijmakers, J. H. M., van Haeringen, A., Fransen van de Putte, D. E., Anderlid, B.-M., Lundin, J., Lapunzina, P., Pérez Jurado, L. A., Delle Chiaie, B., Loeys, B., Menten, B., Oostra, A., Verhelst, H., Amor, D. J., Bruno, D. L., van Essen, A. J., Hordijk, R., Sikkema-Raddatz, B., Verbruggen, K. T., Jongmans, M. C. J., Pfundt, R., Reeser, H. M., Breuning, M. H., & Ruivenkamp, C. A. L. (2009). Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *European Journal of Medical Genetics*, 52(2-3), 77–87.
- Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, Saeed S, Hamilton-Shield J, Clayton-Smith J, O'Rahilly S, Hurles ME, Farooqi IS (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*. 463: 666-70.
- Buizer-Voskamp, J. E., Muntjewerff, J.-W., Strengman, E., Sabatti, C., Stefansson, H., Vorstman, J. A. S., & Ophoff, R. a. (2011). Genome-wide analysis shows increased frequency of copy number variation deletions in Dutch schizophrenia patients. *Biological Psychiatry*, 70(7), 655–62.
- Chance, P. F., Mary, I., Alderson, K., Leppig, K. A., Lensch, M. W., Matsunami, N., Smith, B., Swansoq, P. D., Odelberg, S. J., Disteche, C. M., & Bird, T. D. (1993). DNA Deletion Associated with Hereditary Neuropathy with Liability to Pressure Palsies. *Cell*, 72(162500), 143–151.
- Colella, S., Yau, C., Taylor, J. M., Mirza, G., Butler, H., Clouston, P., Bassett, A. S., Seller, A., Holmes, C. C., & Ragoussis, J. (2007). QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Research*, 35(6), 2013–25.
- Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E., & Pritchard, J. K. (2006). A high-resolution survey of deletion polymorphism in the human genome. *Nature Genetics*, 38(1), 75–81.

- Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C. H., Kristiansson, K., Macarthur, D. G., Macdonald, J. R., Onyiah, I., Pang, A. W. C., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N. P., Lee, C., Scherer, S. W., & Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289), 704–12.
- Cubells, J. F., Deoreo, E. H., Harvey, P. D., Garlow, S. J., Garber, K., Adam, M. P., & Martin, C. L. (2011). Pharmaco-genetically guided treatment of recurrent rage outbursts in an adult male with 15q13.3 deletion syndrome. *American Journal of Medical Genetics. Part A*, 155A(4), 805–10.
- Degenhardt, F., Priebe, L., Herms, S., Mattheisen, M., Mühleisen, T. W., Meier, S., Moebus, S., Strohmaier, J., Groß, M., Breuer, R., Lange, C., Hoffmann, P., Meyer-Lindenberg, A., Heinz, A., Walter, H., Lucae, S., Wolf, C., Müller-Myhsok, B., Holsboer, F., Maier, W., Rietschel, M., Nöthen, M. M., & Cichon, S. (2012). Association between copy number variants in 16p11.2 and major depressive disorder in a German case-control sample. *American Journal of Medical Genetics. Part B, Neuropsychiatric genetics*. 159B(3), 263-73.
- Dellinger, A. E., Saw, S.-M., Goh, L. K., Seielstad, M., Young, T. L., & Li, Y.-J. (2010). Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Research*, 38(9), e105.
- Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN, Herausgeber, 2012). S3-Leitlinie zur Diagnostik und Therapie Bipolarer Störungen. <http://www.leitlinie-bipolar.de/>
- Dunn, O.J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, 56, 52-64
- Eaton, W. W., Martins, S. S., Nestadt, G., Bienvenu, O. J., Clarke, D., & Alexandre, P. (2008). The burden of mental disorders. *Epidemiologic Reviews*, 30, 1–14.
- Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics*, 7(2), 85–97.
- Firth, H. V., Richards, S. M., Bevan, a P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R. M., & Carter, N. P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American Journal of Human Genetics*, 84(4), 524–33.
- Fisher, R. A. (1922). On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society*, 85(1), 87.
- Förster, T. (1948). Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik*, 437(1-2), 55-75.
- Glessner, J. T., Wang, K., Sleiman, P. M. A., Zhang, H., Kim, C. E., Flory, J. H., Bradfield, J. P., Imielinski, M., Frackelton, E. C., Qiu, H., Mentch, F., Grant, S. F. A., & Hakonarson, H. (2010). Duplication of the SLIT3 Locus on 5q35.1 Predisposes to Major Depressive Disorder. *PLoS ONE*, 5(12), e15463.

- Golzio, C., Willer, J., Talkowski, M. E., Oh, E. C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J. F., Kamiya, A., Beckmann, J. S., & Katsanis, N. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature*, 485(7398), 363–7.
- Grigoriou-Serbanescu, M., Martinez, M., Nöthen, M. M., Grinberg, M., Sima, D., Propping, P., Marinescu, E., & Hrestic, M. (2001). Different Familial Transmission Patterns in Bipolar I Disorder With Onset Before and After Age 25. *American Journal of Medical Genetics*, 105(8), 765-73.
- Grozeva, D., Kirov, G., Ivanov, D., Jones, I. R., Jones, L., Green, E. K., St Clair, D. M., Young, A. H., Ferrier, N., Farmer, A. E., McGuffin, P., Holmans, P. A., Owen, M. J., O'Donovan, M. C., & Craddock, N. (2010). Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Archives of general psychiatry*, 67(4), 318–27.
- Gudmundsson, J., Sulem, P., Gudbjartsson, D. F., Masson, G., Agnarsson, B. A., Benediktsdottir, K. R., Sigurdsson, A., Magnusson, O. T., Gudjonsson, S. A., Magnúsdóttir, D. N., Johannsdóttir, H., Helgadóttir, H. T., Stacey, S. N., Jonasdóttir, A., Olafsdóttir, S. B., Thorleifsson, G., Jonasson, J. G., Tryggvadóttir, L., Navarrete, S., Fuertes, F., Helfand, B. T., Hu, Q., Csiki, I. E., Mates, I. N., Jinga, V., Aben, K. K. H., van Oort, I. M., Vermeulen, S. H., Donovan, J. L., Hamdy, F. C., Ng, C.-F., Chiu, P. K. F., Lau, K.-M., Ng, M. C. Y., Gulcher, J. R., Kong, A., Catalona, W. J., Mayordomo, J. I., Einarsson, G. V., Barkardóttir, R. B., Jonsson, E., Mates, D., Neal, D. E., Kiemeneý, L. A., Thorsteinsdóttir, U., Rafnar, T., & Stefansson, K. (2012). A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nature Genetics*, (October, Epub ahead of print), 1–5.
- Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G., & Chee, M. S. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genetics*, 37(5), 549–54.
- Heid, C. a, Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. *Genome Research*, 6(10), 986–994.
- Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J.P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106(23):9362-7.
- Hinds, D. A., Kloek, A. P., Jen, M., Chen, X., & Frazer, K. A.. (2006). Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nature Genetics*, 38(1), 82–5.
- Hurles, M. E., & Lupski, J. R. (2006). Recombination Hotspots in Nonallelic Homologous Recombination. *Genomic Disorders: The Genomic Basis of Disease*, V, 341–355.
- lafrate, a J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W., & Lee, C. (2004). Detection of large-scale variation in the human genome. *Nature Genetics*, 36(9), 949–51.

- Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietiläinen, O. P. H., Buizer-Voskamp, J. E., Strengman, E., Francks, C., Muglia, P., Gylfason, A., Gustafsson, O., Olason, P. I., Steinberg, S., Hansen, T., Jakobsen, K. D., Rasmussen, H. B., Giegling, I., Möller, H.-J., Hartmann, A., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Bramon, E., Kiemenev, L. A., Franke, B., Murray, R., Vassos, E., Touloupoulou, T., Mühleisen, T. W., Tosato, S., Ruggeri, M., Djurovic, S., Andreassen, O. A., Zhang, Z., Werge, T., Ophoff, R. a, Rietschel, M., Nöthen, M. M., Petursson, H., Stefansson, H., Peltonen, L., Collier, D., Stefansson, K., & Clair, D. M. S. (2009). Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Molecular Psychiatry*, 16(1), 17-25.
- Jacquemont, S., Reymond, A., Zufferey, F., Harewood, L., Walters, R. G., Kutalik, Z., Martinet, D., Shen, Y., Valsesia, A., Beckmann, N. D., Thorleifsson, G., Belfiore, M., Bouquillon, S., Campion, D., de Leeuw, N., de Vries, B. B. A., Esko, T., Fernandez, B. a., Fernández-Aranda, F., Fernández-Real, J. M., Gratacòs, M., Guilmatre, A., Hoyer, J., Jarvelin, M.-R., Frank Kooy, R., Kurg, A., Le Caignec, C., Männik, K., Platt, O. S., Sanlaville, D., Van Haelst, M. M., Villatoro Gomez, S., Walha, F., Wu, B., Yu, Y., Aboura, A., Addor, M.-C., Alembik, Y., Antonarakis, S. E., Arveiler, B., Barth, M., Bednarek, N., Béna, F., Bergmann, S., Beri, M., Bernardini, L., Blaumeiser, B., Bonneau, D., Bottani, A., Boute, O., Brunner, H. G., Cailley, D., Callier, P., Chiesa, J., Chrast, J., Coin, L., Coutton, C., Cuisset, J.-M., Cuvellier, J.-C., David, A., de Freminville, B., Delobel, B., Delrue, M.-A., Demeer, B., Descamps, D., Didelot, G., Dieterich, K., Disciglio, V., Doco-Fenzy, M., Drunat, S., Duban-Bedu, B., Dubourg, C., El-Sayed Moustafa, J. S., Elliott, P., Faas, B. H. W., Faivre, L., Faudet, A., Fellmann, F., Ferrarini, A., Fisher, R., Flori, E., Forer, L., Gaillard, D., Gerard, M., Gieger, C., Gimelli, S., Gimelli, G., Grabe, H. J., Guichet, A., Guillin, O., Hartikainen, A.-L., Heron, D., Hippolyte, L., Holder, M., Homuth, G., Isidor, B., Jaillard, S., Jaros, Z., Jiménez-Murcia, S., Joly Helas, G., et al. (2011). Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature*, 478(7367), 97-102 .
- Karayiorgou, M., Morris, M. a, Morrow, B., Shprintzen, R. J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P. S., & Lasseter, V. K. (1995). Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proceedings of the National Academy of Sciences*, 92(17), 7612–6.
- Kendler, K. S. (1993). A Pilot Swedish Twin Study of Affective Illness, Including Hospital- and Population-Ascertained Subsamples. *Archives of General Psychiatry*, 50(9), 699.
- Kirov, G., Rujescu, D., Ingason, A., Collier, D. A., O'Donovan, M. C., & Owen, M. J. (2009a). Neurexin 1 (NRXN1) deletions in schizophrenia. *Schizophrenia Bulletin*, 35(5), 851–4.
- Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K. K., Holmans, P., Craddock, N., Owen, M. J., & O'Donovan, M. C. (2009b). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Human Molecular Genetics*, 18(8), 1497–503.

- Konneker, T., Barnes, T., Furberg, H., Losh, M., Bulik, C. M., & Sullivan, P. F. (2008). A searchable database of genetic evidence for psychiatric disorders. *American Journal of Medical Genetics. Part B, Neuropsychiatric genetics*, 147B(6), 671–5.
- Korbel, J. O., Urban, A. E., Grubert, F., Du, J., Royce, T. E., Starr, P., Zhong, G., Emanuel, B. S., Weissman, S. M., Snyder, M., & Gerstein, M. B. (2007). Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 104(24), 10110–5.
- Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P. J. P., El Mokhtari, N. E., & Schreiber, S. (2006). PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genetics*, 9(1), 55–61.
- Kumar, R. A., KaraMohamed, S., Sudi, J., Conrad, D. F., Brune, C., Badner, J. a, Gilliam, T. C., Nowak, N. J., Cook, E. H., Dobyns, W. B., & Christian, S. L. (2008). Recurrent 16p11.2 microdeletions in autism. *Human Molecular Genetics*, 17(4), 628–38.
- Lam, K.-W. G., & Jeffreys, A. J. (2006). Processes of copy-number change in human DNA: the dynamics of α -globin gene deletion. *Proceedings of the National Academy of Sciences*, 103(24), 8921–7.
- Lee, J. A., Carvalho, C. M., & Lupski, J. R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, 131, 1235–1247.
- Levinson, D. F., Duan, J., Oh, S., Wang, K., Sanders, A. R., Shi, J., Zhang, N., Mowry, B. J., Olincy, A., Amin, F., Cloninger, C. R., Silverman, J. M., Buccola, N. G., Byerley, W. F., Black, D. W., Kendler, K. S., Freedman, R., Dudbridge, F., Pe'er, I., Hakonarson, H., Bergen, S. E., Fanous, A. H., Holmans, P. A., & Gejman, P. V. (2011). Copy Number Variants in Schizophrenia: Confirmation of Five Previous Findings and New Evidence for 3q29 Microdeletions and VIPR2 Duplications. *The American Journal of Psychiatry*, 168(3), 302-16.
- Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*, 373(9659), 234–9.
- Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *The Journal of Biological Chemistry*, 283(1), 1–5.
- Lupski, J. R., Oca-luna, R. M. D., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B. J., Saucedo-carclenas, O., Barker, D. F., Killian, J. M., Garcia, C. A., & Chakravarti, A. (1991). DNA Duplication Associated with Charcot-Marie-Tooth Disease Type 1A. *Cell*, 66, 219–232.
- Lupski, J. R., & Stankiewicz, P. (2005). Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genetics*, 1(6), e49.

- Magri, C., Sacchetti, E., Traversa, M., Valsecchi, P., Gardella, R., Bonvicini, C., Minelli, A., Gennarelli, M., & Barlati, S. (2010). New copy number variations in schizophrenia. *PLoS ONE*, 5(10), e13422.
- Malhotra, D., McCarthy, S., Michaelson, J. J., Vacic, V., Burdick, K. E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E. S., Gill, M., Karayiorgou, M., Kelsoe, J. R., Krastoshevsky, O., Krause, V., Leibenluft, E., Levy, D. L., Makarov, V., Bhandari, A., Malhotra, A. K., McMahon, F. J., Nöthen, M. M., Potash, J. B., Rietschel, M., Schulze, T. G., & Sebat, J. (2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron*, 72(6), 951–63.
- Malhotra, D., & Sebat, J. (2012). CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. *Cell*, 148(6), 1223–1241.
- McCarthy, S. E., Makarov, V., Kirov, G., Addington, A. M., McClellan, J., Yoon, S., Perkins, D. O., Dickel, D. E., Kusenda, M., Krastoshevsky, O., Krause, V., Kumar, R. A., Grozeva, D., Malhotra, D., Walsh, T., Zackai, E. H., Kaplan, P., Ganesh, J., Krantz, I. D., Spinner, N. B., Roccanova, P., Bhandari, A., Pavon, K., Lakshmi, B., Leotta, A., Kendall, J., Lee, Y., Vacic, V., Gary, S., Iakoucheva, L. M., Crow, T. J., Christian, S. L., Lieberman, J. A., Stroup, T. S., Lehtimäki, T., Puura, K., Haldeman-Englert, C., Pearl, J., Goodell, M., Willour, V. L., DeRosse, P., Steele, J., Schulze, T. G., Nöthen, M. M., Cichon, S., Rietschel, M., Leibenluft, E., Kustanovich, V., Lajonchere, C. M., Sutcliffe, J. S., Skuse, D., Gill, M., Gallagher, L., Mendell, N. R., Wellcome Trust Case Control Consortium, Craddock, N., Owen, M. J., O'Donovan, M. C., Shaikh, T. H., Susser, E., DeLisi, L. E., Sullivan, P. F., Deutsch, C. K., Rapoport, J., Levy, D. L., King, M.-C., & Sebat, J. (2009). Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genetics*, 41(11), 1223–1227.
- McGuffin, P., Katz, R., Watkins, S., & Rutherford, J. (1996). A Hospital-Based Twin Register of the Heritability of DSM-IV Unipolar Depression. *Archives of General Psychiatry*, 53(2), 129.
- McGuffin, P., Rijdsdijk, F., Andrew, M., Sham, P., Katz, R., & Cardno, A. (2003). The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Archives of General Psychiatry*, 60(5), 497–502.
- McQuillin, A., Bass, N., Anjorin, A., Lawrence, J., Kandaswamy, R., Lydall, G., Moran, J., Sklar, P., Purcell, S., & Gurling, H. (2011). Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. *European Journal of Human Genetics*, 19(5), 588–92.
- Miller, S., Dykes, D., & Polesky, H. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16(3), 1215.

- Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., Abyzov, A., Yoon, S. C., Ye, K., Cheetham, R. K., Chinwalla, A., Conrad, D. F., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, F., Iakoucheva, L. M., Iqbal, Z., Kang, S., Kidd, J. M., Konkel, M. K., Korn, J., Khurana, E., Kural, D., Lam, H. Y. K., Leng, J., Li, R., Li, Y., Lin, C.-Y., Luo, R., Mu, X. J., Nemesh, J., Peckham, H. E., Rausch, T., Scally, A., Shi, X., Stromberg, M. P., Stütz, A. M., Urban, A. E., Walker, J. A., Wu, J., Zhang, Y., Zhang, Z. D., Batzer, M. a., Ding, L., Marth, G. T., McVean, G., Sebat, J., Snyder, M., Wang, J., Ye, K., Eichler, E. E., Gerstein, M. B., Hurler, M. E., Lee, C., McCarroll, S. a., & Korb, J. O. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*, 470(7332), 59–65.
- Moreno-De-Luca, D., Mulle, J. G., Kaminsky, E. B., Sanders, S. J., Myers, S. M., Adam, M. P., Pakula, A. T., Eisenhauer, N. J., Uhas, K., & Weik, L. (2010). Deletion 17q12 Is a Recurrent Copy Number Variant that Confers High Risk of Autism and Schizophrenia. *American Journal of Human Genetics*, 87(5), 618-30.
- Mulle, J. G., Dodd, A. F., McGrath, J. a, Wolyniec, P. S., Mitchell, A. a, Shetty, A. C., Sobreira, N. L., Valle, D., Rudd, M. K., Satten, G., Cutler, D. J., Pulver, A. E., & Warren, S. T. (2010). Microdeletions of 3q29 confer high risk for schizophrenia. *American Journal of Human Genetics*, 87(2), 229–36.
- Peiffer, D. A., Le, J. M., Steemers, F. J., Chang, W., Jenniges, T., Garcia, F., Haden, K., Li, J., Shaw, C. A., Belmont, J., Cheung, S. W., Shen, R. M., Barker, D. L., & Gunderson, K. L. (2006). High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Research*, 16(9), 1136–48.
- Pentao, L., Wise, C. A., Chinault, C., Patel, P. I., & Lupski, J. R. (1992). Genome-wide association analysis of copy number variation in recurrent depressive disorder. *Nature Genetics*, 2, 292–300.
- Pinkel, D., Straume, T., & Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, 83(9), 2934–8.
- Ponchel, F., Toomes, C., Bransfield, K., Leong, F. T., Douglas, S. H., Field, S. L., Bell, S. M., Combaret, V., Puisieux, A., Mighell, A. J., Robinson, P. A., Inglehearn, C. F., Isaacs, J. D., & Markham, A. F. (2003). Real-time PCR based on SYBR-Green I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions TCR-alpha. *BMC Biotechnology*, 13, 1–13.
- Priebe, L., Degenhardt, F. A., Herms, S., Haenisch, B., Mattheisen, M., Nieratschker, V., Weingarten, M., Witt, S., Breuer, R., Paul, T., Alblas, M., Moebus, S., Lathrop, M., Leboyer, M., Schreiber, S., Grigoriou-Serbanescu, M., Maier, W., Propping, P., Rietschel, M., Nöthen, M. M., Cichon, S., & Mühleisen, T. W. (2012). Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder. *Molecular Psychiatry*, 17(4), 421-32.

- Priebe, L., Degenhardt, F., Strohmaier, J., Breuer, R., Herms, S., Witt, S. H., Hoffmann, P., Kulbida, R., Mattheisen, M., Moebus, S., Meyer-Lindenberg, A., Walter, H., Mössner, R., Nenadic, I., Sauer, H., Rujescu, D., Maier, W., Rietschel, M., Nöthen, M. M., & Cichon, S. (2013). Copy Number Variants in German Patients with Schizophrenia. *PLoS One*, 8: e64035.
- Pruitt, K. D., Tatusova, T., Brown, G. R., & Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Research*, 40(Database issue), D130–5.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., Sklar, P., Debakker, P., & Daly, M. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 81(3), 559–575.
- Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, 77(2), 257–286.
- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., González, J. R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., Marshall, C. R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M. J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D. F., Estivill, X., Tyler-Smith, C., Carter, N. P., Aburatani, H., Lee, C., Jones, K. W., Scherer, S. W., & Hurles, M. E. (2006). Global variation in copy number in the human genome. *Nature*, 444(7118), 444–54.
- Rietschel, M., Mattheisen, M., Degenhardt, F., Mühleisen, T. W., Kirsch, P., Esslinger, C., Herms, S., Demontis, D., Steffens, M., Strohmaier, J., Haenisch, B., Breuer, R., Czerski, P. M., Giegling, I., Strengman, E., Schmael, C., Mors, O., Mortensen, P. B., Hougaard, D. M., Ørntoft, T., Kapelski, P., Priebe, L., Basmanav, F. F., Forstner, A. J., Hoffman, P., Meier, S., Nikitopoulos, J., Moebus, S., Alexander, M., Mössner, R., Wichmann, H.-E., Schreiber, S., Rivandeneira, F., Hofman, A., Uitterlinden, A. G., Wienker, T. F., Schumacher, J., Hauser, J., Maier, W., Cantor, R. M., Erk, S., Schulze, T. G., Craddock, N., Owen, M. J., O'Donovan, M. C., Børglum, A. D., Rujescu, D., Walter, H., Meyer-Lindenberg, A., Nöthen, M. M., Ophoff, R. A., & Cichon, S. (2012). Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. *Molecular Psychiatry*, 17(9), 906–17.
- Rucker, J. J. H., Breen, G., Pinto, D., Pedroso, I., Lewis, C. M., Cohen-Woods, S., Uher, R., Schosser, A., Rivera, M., Aitchison, K. J., Craddock, N., Owen, M. J., Jones, L., Jones, I., Korszun, A., Muglia, P., Barnes, M. R., Preisig, M., Mors, O., Gill, M., Maier, W., Rice, J., Rietschel, M., Holsboer, F., Farmer, A. E., Craig, I. W., Scherer, S. W., & McGuffin, P. (2011). Genome-wide association analysis of copy number variation in recurrent depressive disorder. *Molecular Psychiatry*, 18(2):183-9.
- Rucker, J. J. H., & McGuffin, P. (2012) Genomic structural variation in psychiatric disorders. *Development and Psychopathology*, 24(4):1335-44.

- Rujescu, D., Ingason, A., Cichon, S., Pietiläinen, O. P. H., Barnes, M. R., Toulopoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E., Murray, R., Ruggeri, M., Tosato, S., Bonetto, C., Steinberg, S., Sigurdsson, E., Sigmundsson, T., Petursson, H., Gylfason, A., Olason, P. I., Hardarsson, G., Jonsdottir, G. A., Gustafsson, O., Fossdal, R., Giegling, I., Möller, H.-J., Hartmann, A. M., Hoffmann, P., Crombie, C., Fraser, G., Walker, N., Lonngqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Djurovic, S., Melle, I., Andreassen, O. A., Hansen, T., Werge, T., Kiemenev, L. A., Franke, B., Veltman, J., Buizer-Voskamp, J. E., Sabatti, C., Ophoff, R. A., Rietschel, M., Nöthen, M. M., Stefansson, K., Peltonen, L., St Clair, D., Stefansson, H., & Collier, D. A. (2009). Disruption of the neurexin 1 gene is associated with schizophrenia. *Human Molecular Genetics*, 18(5), 988–96.
- Scherer, S. W., Lee, C., Birney, E., Altshuler, D. M., Eichler, E. E., Carter, N. P., Hurler, M. E., & Feuk, L. (2007). Challenges and standards in integrating surveys of structural variation. *Nature Genetics*, 39(7 Suppl), S7–15.
- Schmermund, A., Möhlenkamp, S., Stang, A., Grönemeyer, D., Seibel, R., Hirche, H., Mann, K., Siffert, W., Lauterbach, K., Siegrist, J., Jöckel, K.-H., & Erbel, R. (2002). Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: Rationale and design of the Heinz Nixdorf RECALL Study. *American Heart Journal*, 144(2), 212–218.
- Schouten, J. P., McElgunn, C. J., Waaijer, R., Zwiijnenburg, D., Diepvens, F., & Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research*, 30(12), e57.
- Sebat, J. (2007). Major changes in our DNA lead to major changes in our thinking. *Nature Genetics*, 39(7 Suppl), S3–5.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T. C., Trask, B., Patterson, N., Zetterberg, A., & Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science*, 305(5683), 525–8.
- Shaikh, T. H., Gai, X., Perin, J. C., Glessner, J. T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L. K., D'Arcy, M., Frackelton, E. C., Geiger, E. A., Haldeman-Englert, C., Imielinski, M., Kim, C. E., Medne, L., Annaiah, K., Bradfield, J. P., Dabaghyan, E., Eckert, A., Onyiah, C. C., Ostapenko, S., Otieno, F. G., Santa, E., Shaner, J. L., Skraban, R., Smith, R. M., Elia, J., Goldmuntz, E., Spinner, N. B., Zackai, E. H., Chiavacci, R. M., Grundmeier, R., Rappaport, E. F., Grant, S. F. A., White, P. S., & Hakonarson, H. (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Research*, 19(9), 1682–90.
- Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U., Pertz, L. M., Clark, R. a, Schwartz, S., Seagraves, R., Oseroff, V. V., Albertson, D. G., Pinkel, D., & Eichler, E. E. (2005). Segmental duplications and copy-number variation in the human genome. *American Journal of Human Genetics*, 77(1), 78–88.

- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, 29(1), 308–11.
- Shinawi, M., Liu, P., Kang, S.-H. L., Shen, J., Belmont, J. W., Scott, D. A., Probst, F. J., Craigen, W. J., Graham, B. H., Pursley, A., Clark, G., Lee, J., Proud, M., Stocco, A., Rodriguez, D. L., Kozel, B. a, Sparagana, S., Roeder, E. R., McGrew, S. G., Kurczynski, T. W., Allison, L. J., Amato, S., Savage, S., Patel, A., Stankiewicz, P., Beaudet, A. L., Cheung, S. W., & Lupski, J. R. (2010). Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *Journal of Medical Genetics*, 47(5), 332–41.
- Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Döhner, H., Cremer, T., & Lichter, P. (1997). Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes, Chromosomes & Cancer*, 20(4), 399–407.
- Stemers, F. J., Chang, W., Lee, G., Barker, D. L., Shen, R., & Gunderson, K. L. (2006). Whole-genome genotyping with the single-base extension assay. *Nature Methods*, 3(1), 2005–2007.
- Stemers, F. J., & Gunderson, K. L. (2007). Whole genome genotyping technologies on the BeadArray platform. *Biotechnology Journal*, 2(1), 41–9.
- Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O. P. H., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J. E., Hansen, T., Jakobsen, K. D., Muglia, P., Francks, C., Matthews, P. M., Gylfason, A., Halldorsson, B. V., Gudbjartsson, D., Thorgeirsson, T. E., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Bjornsson, A., Mattiasdottir, S., Blondal, T., Haraldsson, M., Magnusdottir, B. B., Giegling, I., Möller, H.-J., Hartmann, A., Shianna, K. V., Ge, D., Need, A. C., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Paunio, T., Touloupoulou, T., Bramon, E., Di Forti, M., Murray, R., Ruggeri, M., Vassos, E., Tosato, S., Walshe, M., Li, T., Vasilescu, C., Mühleisen, T. W., Wang, A. G., Ullum, H., Djurovic, S., Melle, I., Olesen, J., Kiemenev, L. A., Franke, B., Sabatti, C., Freimer, N. B., Gulcher, J. R., Thorsteinsdottir, U., Kong, A., Andreassen, O. A., Ophoff, R. A., Georgi, A., Rietschel, M., Werge, T., Petursson, H., Goldstein, D. B., Nöthen, M. M., Peltonen, L., Collier, D. A., St Clair, D., & Stefansson, K. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature*, 455(7210), 232–6.
- Sullivan, P. F. (2010). The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron*, 68(2), 182–6.
- Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a Complex Trait: evidence from a meta-analysis of twin studies. *Archives of General Psychiatry*, 60(12), 1187-92.
- Sullivan, P. F., Neale, M. C., Ph, D., & Kendler, K. S. (2000). Genetic Epidemiology of Major Depression: Review and Meta-Analysis. *American Journal of Psychiatry*, 157(10), 1552–1562.

- 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: 537-51.
- The 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), 56-65.
- The International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164), 851-61.
- The International Schizophrenia Consortium. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, 455(7210), 237-241.
- The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), 661-78.
- The Wellcome Trust Case Control Consortium. (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, 464(7289), 713-20.
- Turner, D. J., Miretti, M., Rajan, D., Fiegler, H., Carter, N. P., Blayney, M. L., Beck, S., & Hurles, M. E. (2008). Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nature Genetics*, 40(1), 90-5.
- Tuzun, E., Sharp, A. J., Bailey, J. a, Kaul, R., Morrison, V. A., Pertz, L. M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., Olson, M. V., & Eichler, E. E. (2005). Fine-scale structural variation of the human genome. *Nature Genetics*, 37(7), 727-32.
- Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H.-H., Peoples, A., Makarov, V., Yoon, S., Bhandari, A., Corominas, R., Iakoucheva, L. M., Krastoshevsky, O., Krause, V., Larach-Walters, V., Welsh, D. K., Craig, D., Kelsoe, J. R., Gershon, E. S., Leal, S. M., Aquila, M. D., Morris, D. W., Gill, M., Corvin, A., Insel, P. A., McClellan, J., King, M.-C., Karayiorgou, M., Levy, D. L., DeLisi, L. E., & Sebat, J. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature*, 471(7339), 499-503.
- Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., Nord, A. S., Kusenda, M., Malhotra, D., Bhandari, A., Stray, S. M., Rippey, C. F., Roccanova, P., Makarov, V., Lakshmi, B., Findling, R. L., Sikich, L., Stromberg, T., Merriman, B., Gogtay, N., Butler, P., Eckstrand, K., Noory, L., Gochman, P., Long, R., Chen, Z., Davis, S., Baker, C., Eichler, E. E., Meltzer, P. S., Nelson, S. F., Singleton, A. B., Lee, M. K., Rapoport, J. L., King, M.-C., & Sebat, J. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, 320(5875), 539-43.

- Walters, R. G., Jacquemont, S., Valsesia, A., de Smith, A. J., Martinet, D., Andersson, J., Falchi, M., Chen, F., Andrieux, J., Lobbens, S., Delobel, B., Stutzmann, F., El-Sayed Moustafa, J. S., Chèvre, J.-C., Lecoeur, C., Vatin, V., Bouquillon, S., Buxton, J. L., Boute, O., Holder-Espinasse, M., Cuisset, J.-M., Lemaitre, M.-P., Ambresin, A.-E., Brioschi, A., Gaillard, M., Giusti, V., Fellmann, F., Ferrarini, A., Hadjikhani, N., Champion, D., Guilmatre, A., Goldenberg, A., Calmels, N., Mandel, J.-L., Le Caignec, C., David, A., Isidor, B., Cordier, M.-P., Dupuis-Girod, S., Labalme, a, Sanlaville, D., Béri-Dexheimer, M., Jonveaux, P., Leheup, B., Ounap, K., Bochukova, E. G., Henning, E., Keogh, J., Ellis, R. J., Macdermot, K. D., van Haelst, M. M., Vincent-Delorme, C., Plessis, G., Touraine, R., Philippe, A., Malan, V., Mathieu-Dramard, M., Chiesa, J., Blaumeiser, B., Kooy, R. F., Caiazzo, R., Pigeyre, M., Balkau, B., Sladek, R., Bergmann, S., Mooser, V., Waterworth, D., Reymond, A., Vollenweider, P., Waeber, G., Kurg, A., Palta, P., Esko, T., Metspalu, A., Nelis, M., Elliott, P., Hartikainen, a-L., McCarthy, M. I., Peltonen, L., Carlsson, L., Jacobson, P., Sjöström, L., Huang, N., Hurler, M. E., O’Rahilly, S., Farooqi, I. S., Männik, K., Jarvelin, M.-R., Pattou, F., Meyre, D., Walley, A. J., Coin, L. J. M., Blakemore, A. I. F., Froguel, P., & Beckmann, J. S. (2010). A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*, 463(7281), 671–5.
- Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F. A., Hakonarson, H., & Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research*, 17(11), 1665–74.
- Weiss, L. A., Shen, Y., Korn, J. M., Arking, D. E., Miller, D. T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M. A. R., Green, T., Platt, O. S., Ruderfer, D. M., Walsh, C.A., Altshuler, D., Chakravarti, A., Tanzi, R. E., Stefansson, K., Santangelo, S. L., Gusella, J. F., Sklar, P., Wu, B.-L., & Daly, M. (2008). Association between microdeletion and microduplication at 16p11. 2 and autism. *New England Journal of Medicine*, 358(7), 667–673.
- Wichmann, H.-E., Gieger, C., & Illig, T. (2005). KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*, 67 Suppl 1(S 01), S26–30.
- Winchester, L., Yau, C., & Ragoussis, J. (2009). Comparing CNV detection methods for SNP arrays. *Briefings in functional genomics & proteomics*, 8(5), 353–66.
- 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*, 3, 118.
- Xu, B., Roos, J. L., Levy, S., van Rensburg, E. J., Gogos, J. A., & Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature Genetics*, 40(7), 880–5.
- Yang, S., Wang, K., Gregory, B., Berrettini, W., Wang, L.-S., Hakonarson, H., & Bucan, M. (2009). Genomic landscape of a three-generation pedigree segregating affective disorder. *PloS ONE*, 4(2), e4474.

- Zhang, D., Cheng, L., Qian, Y., Alliey-Rodriguez, N., Kelsoe, J. R., Greenwood, T., Nievergelt, C., Barrett, T. B., McKinney, R., Schork, N., Smith, E. N., Bloss, C., Nurnberger, J., Edenberg, H. J., Foroud, T., Sheftner, W., Lawson, W. B., Nwulia, E. A., Hipolito, M., Coryell, W., Rice, J., Byerley, W., McMahon, F., Schulze, T. G., Berrettini, W., Potash, J. B., Belmonte, P. L., Zandi, P. P., McInnis, M. G., Zöllner, S., Craig, D., Szelinger, S., Koller, D., Christian, S. L., Liu, C., & Gershon, E. S. (2009). Singleton deletions throughout the genome increase risk of bipolar disorder. *Molecular Psychiatry*, 14(4), 376–80.
- Zhang, J., Feuk, L., Duggan, G. E., Khaja, R., & Scherer, S. W. (2006). Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenetic and Genome Research*, 115(3-4), 205–14.

10 Eigene Publikationen

- Abou Jamra R., Wohlfart S., Zweier M., Uebe S., **Priebe L.**, Ekici A., Giesebrecht S., Abboud A., Al Khateeb M. A., Fakher M., Hamdan S., Ismael A., Muhammad S., Nöthen M. M., Schumacher J., & Reis A. (2011). Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. *European Journal of Human Genetics*, 19: 161-6.
- Brockschmidt F. F., Heilmann S., Ellis J. A., Eigelshoven S., Hanneken S., Herold C., Moebus S., Alblas M. A., Lippke B., Kluck N., **Priebe L.**, Degenhardt F. A., Jamra R. A., Meesters C., Jöckel K. H., Erbel R., Harrap S., Schumacher J., Fröhlich H., Kruse R., Hillmer A. M., Becker T., & Nöthen M. M. (2011). Susceptibility variants on chromosome 7p21.1 suggest HDAC9 as a new candidate gene for male-pattern baldness. *British Journal of Dermatology*, 165: 1293-302.
- Cichon S., Mühleisen T. W., Degenhardt F. A., Mattheisen M., Miró X., Strohmaier J., Steffens M., Meesters C., Herms S., Weingarten M., **Priebe L.**, Haenisch B., Alexander M., Vollmer J., Breuer R., Schmäler C., Tessmann P., Moebus S., Wichmann H. E., Schreiber S., Müller-Myhsok B., Lucae S., Jamain S., Leboyer M., Bellivier F., Etain B., Henry C., Kahn J. P., Heath S., Bipolar Disorder Genome Study (BiGS) Consortium, Hamshere M., O'Donovan M. C., Owen M. J., Craddock N., Schwarz M., Vedder H., Kammerer-Ciernioch J., Reif A., Sasse J., Bauer M., Hautzinger M., Wright A., Mitchell P. B., Schofield P. R., Montgomery G. W., Medland S. E., Gordon S. D., Martin N. G., Gustafsson O., Andreassen O., Djurovic S., Sigurdsson E., Steinberg S., Stefansson H., Stefansson K., Kapur-Pojkic L., Oruc L., Rivas F., Mayoral F., Chuchalin A., Babadjanova G., Tiganov A. S., Pantelejeva G., Abramova L. I., Grigoriu-Serbanescu M., Diaconu C. C., Czerski P. M., Hauser J., Zimmer A., Lathrop M., Schulze T. G., Wienker T. F., Schumacher J., Maier W., Propping P., Rietschel M., & Nöthen M. M. (2011). Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *American Journal of Human Genetics*, 88: 372-81.
- Degenhardt F., Priebe L., Strohmaier J., Herms S., Hoffmann P., Mattheisen M., Mössner R., Nenadic I., Sauer H., Rujescu D., Maier W., Schulze T. G., Rietschel M., Nöthen M. M., & Cichon S. (2013). No evidence for an involvement of copy number variation in ABCA13 in schizophrenia, bipolar disorder, or major depressive disorder. *Psychiatric Genetics*, 23: 45-6.
- Degenhardt F.*, **Priebe L.***, Herms S., Mattheisen M., Mühleisen T. W., Meier S., Moebus S., Strohmaier J., Groß M., Breuer R., Lange C., Hoffmann P., Meyer-Lindenberg A., Heinz A., Walter H., Lucae S., Wolf C., Müller-Myhsok B., Holsboer F., Maier W., Rietschel M., Nöthen M. M., & Cichon S. (2012). Association between copy number variants in 16p11.2 and major depressive disorder in a German case-control sample. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 159B: 263-73.
- Grigoriu-Serbanescu M., Diaconu C. C., Herms S., Bleotu C., Vollmer J., Mühleisen T. W., Prelipceanu D., **Priebe L.**, Mihailescu R., Georgescu M. J., Sima D., Grimberg M., Nöthen M. M., & Cichon S. (2008). Investigation of the tryptophan hydroxylase 2 gene in bipolar I disorder in the Romanian population. *Psychiatric Genetics*, 18: 240-7.

- Grimminger T., Pernhorst K., Surges R., Niehusmann P., **Priebe L.**, von Lehe M., Hoffmann P., Cichon S., Schoch S. & Becker A. J. (2013). Levetiracetam resistance: Synaptic signatures & corresponding promoter SNPs in epileptic hippocampi. *Neurobiology of Disease*, 60: 115-25.
- Guha S., Rees E., Darvasi A., Ivanov D., Ikeda M., Bergen S. E., Magnusson P. K., Cormican P., Morris D., Gill M., Cichon S., Rosenfeld J. A., Lee A., Gregersen P. K., Kane J. M., Malhotra A. K., Rietschel M., Nöthen M. M., Degenhardt F., **Priebe L.**, Breuer R., Strohmaier J., Ruderfer D. M., Moran J. L., Chambert K. D., Sanders A. R., Shi J., Kendler K., Riley B., O'Neill T., Walsh D., Malhotra D., Corvin A., Purcell S., Sklar P., Iwata N., Hultman C. M., Sullivan P. F., Sebat J., McCarthy S., Gejman P. V., Levinson D. F., Owen M. J., O'Donovan M. C., Lencz T., & Kirov G. (2013); Molecular Genetics of Schizophrenia Consortium; Wellcome Trust Case Control Consortium 2. Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA Psychiatry*, 70: 253-60.
- Hass J., Walton E., Kirsten H., Liu J., **Priebe L.**, Wolf C., Karbalai N., Gollub R., White T., Roessner V., Müller K. U., Paus T., Smolka M. N., Schumann G., IMAGEN Consortium, Scholz M., Cichon S., Calhoun V. & Ehrlich S.. (2013). A Genome-Wide Association Study Suggests Novel Loci Associated with a Schizophrenia-Related Brain-Based Phenotype. *PLoS One*, 8: e64872.
- Håvik B., Le Hellard S., Rietschel M., Lybæk H., Djurovic S., Mattheisen M., Mühleisen T. W., Degenhardt F., **Priebe L.**, Maier W., Breuer R., Schulze T. G., Agartz I., Melle I., Hansen T., Bramham C. R., Nöthen M. M., Stevens B., Werge T., Andreassen O. A., Cichon S., & Steen V. M. (2011). The complement control-related genes CSMD1 and CSMD2 associate to schizophrenia. *Biological Psychiatry*, 70: 35-42.
- Lavebratt C., Olsson S., Backlund L., Frisé L., Sellgren C., **Priebe L.**, Nikamo P., Träskman-Bendz L., Cichon S., Vawter M. P., Osby U., Engberg G., Landén M., Erhardt S., & Schalling M. (2013). The KMO allele encoding Arg(452) is associated with psychotic features in bipolar disorder type 1, and with increased CSF KYNA level and reduced KMO expression. *Molecular Psychiatry*, 2013 Mar 5 (elektronische Vorveröffentlichung).
- Mühleisen T. W., Mattheisen M., Strohmaier J., Degenhardt F., **Priebe L.**, Schultz C. C., Breuer R., Meier S., Hoffmann P., Genetic Risk and Outcome in Psychosis (GROUP Investigators), Rivandeneira F., Hofman A., Uitterlinden A. G., Moebus S., Gieger C., Emeny R., Ladwig K. H., Wichmann H. E., Schwarz M., Kammerer-Ciernioch J., Schlösser R. G., Nenadic I., Sauer H., Mössner R., Maier W., Rujescu D., Lange C., Ophoff R. A., Schulze T. G., Rietschel M., Nöthen M. M., & Cichon S. (2012). Association between schizophrenia and common variation in neurocan (NCAN), a genetic risk factor for bipolar disorder. *Schizophrenia Research*, 138, 69-73.
- Mühleisen T. W., Basmanav F. B., Forstner A. J., Mattheisen M., **Priebe L.**, Herms S., Breuer R., Moebus S., Nenadic I., Sauer H., Mössner R., Maier W., Rujescu D., Ludwig M., Rietschel M., Nöthen M. M., & Cichon S. (2011). Resequencing and follow-up of neurexin 1 (NRXN1) in schizophrenia patients. *Schizophrenia Research*, 127, 35-40.

- Pernhorst K., van Loo K. M., von Lehe M., **Priebe L.**, Cichon S., Herms S., Hoffmann P., Helmstaedter C., Sander T., Schoch S., & Becker A. J. (2013). Rs6295 promoter variants of the serotonin type 1A receptor are differentially activated by c-Jun in vitro and correlate to transcript levels in human epileptic brain tissue. *Brain Research*, 1499: 136-44.
- Priebe L.***, Degenhardt F.*, Strohmaier J., Breuer R., Herms S., Witt S. H., Hoffmann P., Kulbida R., Mattheisen M., Moebus S., Meyer-Lindenberg A., Walter H., Mössner R., Nenadic I., Sauer H., Rujescu D., Maier W., Rietschel M., Nöthen M. M., & Cichon S. (2013). Copy Number Variants in German Patients with Schizophrenia. *PLoS One*, 8: e64035.
- Priebe L.**, Degenhardt F. A., Herms S., Haenisch B., Mattheisen M., Nieratschker V., Weingarten M., Witt S., Breuer R., Paul T., Alblas M., Moebus S., Lathrop M., Leboyer M., Schreiber S., Grigoriou-Serbanescu M., Maier W., Propping P., Rietschel M., Nöthen M. M., Cichon S., & Mühleisen T.W. (2012). Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder. *Molecular Psychiatry*, 17, 421-32.
- Rietschel M., Mattheisen M., Degenhardt F., Genetic Risk and Outcome in Psychosis (GROUP Investigators), Mühleisen T. W., Kirsch P., Esslinger C., Herms S., Demontis D., Steffens M., Strohmaier J., Haenisch B., Breuer R., Czerski P. M., Giegling I., Strengman E., Schmael C., Mors O., Mortensen P. B., Hougaard D. M., Orntoft T., Kapelski P., **Priebe L.**, Basmanav F. B., Forstner A. J., Hoffmann P., Meier S., Nikitopoulos J., Moebus S., Alexander M., Mössner R., Wichmann H. E., Schreiber S., Rivandeneira F., Hofman A., Uitterlinden A. G., Wienker T. F., Schumacher J., Hauser J., Maier W., Cantor R. M., Erk S., Schulze T. G. SGENE-plus Consortium, Craddock N., Owen M. J., O'Donovan M. C., Børglum A. D., Rujescu D., Walter H., Meyer-Lindenberg A., Nöthen M. M., Ophoff R. A., & Cichon S. (2011). Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. *Molecular Psychiatry*, 17: 906-17.
- Schramm C., Draaken M., Tewes G., Bartels E., Schmiedeke E., Märzheuser S., Grasshoff-Derr S., Hosie S., Holland-Cunz S., **Priebe L.**, Kreiss-Nachtsheim M., Hoffmann P., Aretz S., Nöthen M. M., Reutter H., & Ludwig M. (2011). Autosomal-dominant non-syndromic anal atresia: sequencing of candidate genes, array-based molecular karyotyping, and review of the literature. *European Journal of Pediatrics*, 170: 741-6.
- Schramm C., Draaken M., Bartels E., Boemers T. M., Schmiedeke E., Grasshoff-Derr S., Märzheuser S., Hosie S., Holland-Cunz S., Baudisch F., **Priebe L.**, Hoffmann P., Zink A. M., Engels H., Brockschmidt F. F., Aretz S., Nöthen M. M., Ludwig M., & Reutter H. (2011). De novo duplication of 18p11.21-18q12.1 in a female with anorectal malformation. *American Journal of Medical Genetics Part A*, 155A: 445-9.

*Geteilte Erstautorenschaft

11 Anhang

11.1 Perl-Skripte

11.1.1 Formatierung des Finalreports und Erstellung einer Stapelverarbeitungsdatei

```

print STDERR "Skript $0 $ARGV[0]\n";
print STDERR "Stefan Herms, 2008-02-25-17130\n";
print STDERR "editiert von L.Priebe, 15.09.2010 06.09.2012\n";
print STDERR "\n";
print STDERR "\n";
$infile = $ARGV[0];
$outputpath = ""; #Pfad der Output-Dateien
$inputpath = ""; #Pfad der Input-Dateien
$gkdir = ""; #Pfad der GC-Data
$quasnp = ""; #Pfad in dem QuantiSNP installiert ist.
$setting = 2000000;
$emitters = 25;
$id = int(rand(1)*10000);
open(REIN,"gender.dat");
$_ = <REIN>;
$gendercount = 0;
while(<REIN>) {
    chomp();
    @we = split(/\t/);
    $gender{$we[1]} = $we[2]; #Spalte anpassen (perl f%oingt bei 0 an zu
z%ohlen)
    $gendercount++;
}
close(REIN);
print STDERR "Es wurden $gendercount Geschlechtsinformationen
eingelesen!\n";
open(REIN,$infile);
open(CMD,">quasnp".$id.".bat");
for ($t = 0; $t <10; $t++) {$_ = <REIN>;} #Auskommentieren falls SNP-Subset
chomp();
@header = split(/\t/);
# SNPID SAMPLE CHR POS logR b-allele
$outfile = ">test";
while (<REIN>) {
    chomp();
    @we = split(/\t/);
    if ($outfile ne ">>" . $we[1]) {
        $outfile = ">>" . $we[1];
        close(RAUS);
        print STDERR "Datei geoeffnet: $outfile \t $we[1] \n";
        open(RAUS,$outfile);
    }
}

```

```

if ($filedone{$outfile} ne "done") { # OK QuantiSNP2 scheint keine
Headerzeile mehr zu benötigen!
    #print $header[2] . "\t" . $header[3] . "\t" . $header[4] . "\t" . $header[5] .
"\n";
    #print RAUS $header[2] . "\t" . $header[3] . "\t" . $header[4] . "\t" .
$header[5] . "\n";
    $filedone{$outfile} = "done";
    print CMD $quasnp. "quantisnp2.exe --outdir " . $outputpath. " --sampleid "
.$we[1]. " --gender " . $gender{$we[1]}. " --emitters " . $emitters. " --lsetting "
.$lsetting. " --gkdir " . $gkdir. " --input-files " . $inputpath . $we[1] . " --config " .
$quasnp . "params.dat --levels " . $quasnp . "levels-hd.dat \n";
    #print RMC "del " . $we[1] . "\n";
}
}
if (!/NaN/) {
    print RAUS $we[0] . "\t" . $we[2] . "\t" . $we[3] . "\t" . $we[4] . "\t" . $we[5] .
"\n";
} # While REIN
close(REIN);
close(CMD);
close(RMC);

```

11.2 Shell-Skripte

11.2.1 Chromosomale Zusammenfassung der QuantiSNP Ergebnisse

```

for ((i=1; i<=22; i++))
do
awk -F"\t" -v chr=$i '$2==chr {print $0}' quaSNP_UCO_all.txt
>quaSNP_UCO_chr"$i".tab
cat head.txt quaSNP_UCO_chr"$i".tab >quaSNP_UCO_chr"$i".txt
rm quaSNP_UCO_chr"$i"_emi50.tab
done

```

11.3 QuantiSNP Kommandozeile

```

quantisnp2.exe --outdir --sampleid --gender --emitters --lsetting --gkdir --input-
files --config params.dat --levels levels-hd.dat

```

11.4 PennCNV Kommandozeile

```

perl detect_cnv.pl --test --pfb lib/hhall.hg18.pfb --hmm lib/hhall.hmm
--confi --listfile listfile --out --log

```

11.5 Primersequenzen für SYBR-Green qRT-PCR

11.5.1 MDD – Primer in 7p21.3

Primerpaar1

```

forward primer: ACCCTGTTTTTCTCCCATC

```

reverse primer: GTGCAGGTGAAGGAGTAGCC

Primerpaar2

forward primer: TGCTGACCTTCCACATTCAA

reverse primer: TGCTGTACAACCAGGCAGTT

Primerpaar3

forward primer: TTCACAAGGGATGTGCAAAC

reverse primer: CCACGCTACTCCTTTGAAGC

11.5.2 MDD – Primer in 15q26.3

Primerpaar1

forward primer: CAAGGACGACAGCAAAGACA

reverse primer: GCTTCCCTGCAAGTTGAGTC

Primerpaar2

forward primer: ATAGCAGCAAACCCCTCCT

reverse primer: TCTCCAGACTGGGCTTCTGT

Primerpaar3

forward primer: TGGGCTCCTTAGAGAACAGC

reverse primer: CCAGGTGCATCATACACGAC

Primerpaar4

forward primer: TAAACAGACGGCTGCATGTC

reverse primer: AGTATCAGCAGCCAGGGAGA

11.5.3 MDD – Primer in 18p11.32

Primerpaar1

forward primer: TGGCAAATGCTATGTGATCC

reverse primer: CAGCAAGCCACCATCAACT

Primerpaar2

forward primer: CGGCTTTGGTGATGAAGTTT

reverse primer: TGCTGAAACCTGACAAATGG

Primerpaar3

forward primer: CTTGCCATTTGCAGTGAAGA

reverse primer: TGGGCAAATTTTCTCTTGG

11.5.4 Referenzprimer

BNC1

forward primer: GCAGATGTCACACTGGAAGC

reverse primer: TCAGTGCTTTGTCCAACAGG

CFTR

forward primer: GGAGATGCTCCTGTCTCCTG

reverse primer: GGGAGTCTTTTGCACAATGG

RPP38

forward primer: CTGCCATGATCACCTCACAC

reverse primer: GAACGCCAAGGCTAGAACAC

11.6 Ergebnisse burden Tests

11.6.1 Bipolare Störung

Tab. 11.1: Übersicht der burden Tests bei Bipolarer Störung. Es werden nur die P-Werte kleiner als 0,1 dargestellt, die Gesamtergebnisse befinden sich im Anhang. Die fett markierten P-Werte werden im Text genauer beschrieben. * Dieser P-Wert blieb nach Korrektur für multiples Testen signifikant (korrigierter P-Wert: 0,0288). Modifiziert nach Priebe et al., 2012.

Vergleichsgruppe	Test			
	CNV Frequenz		CNV Länge	
	RATE-Test	PROP-Test	TOTKB-Test	AVGKB-Test
Alle Patienten (882) gegen alle Kontrollen (872)				
Alle CNVs	0,6298	0,3924	0,3912	0,1889
Nur Duplikationen	0,5426	0,1635	0,5441	0,2530
Nur Deletionen	0,6751	0,9488	0,1427	0,2860
Alle <i>singleton</i> CNVs	0,3155	0,3560	0,5458	0,5072
Nur <i>singleton</i> Duplikationen	0,0633	0,0534	0,6289	0,6817
Nur <i>singleton</i> Deletionen	0,7028	0,7101	0,0141	0,0139
Patienten mit frühem Erkrankungsalter (291) gegen alle Kontrollen (872)				
Alle CNVs	0,0875	0,0009	0,4477	0,0919
Nur Duplikationen	0,0238	0,0004	0,5265	0,2245
Nur Deletionen	0,7854	0,6860	0,2170	0,1750
Alle <i>singleton</i> CNVs	0,0453	0,0874	0,0812	0,0939
Nur <i>singleton</i> Duplikationen	0,0102	0,0166	0,4281	0,5150
Nur <i>singleton</i> Deletionen	0,5403	0,5206	0,0084	0,0056
Patienten mit spätem Erkrankungsalter (591) gegen alle Kontrollen (872)				
Alle CNVs	0,9265	0,9494	0,4083	0,4516
Nur Duplikationen	0,9519	0,8250	0,5893	0,3785
Nur Deletionen	0,5854	0,9767	0,1568	0,4503
Alle <i>singleton</i> CNVs	0,7641	0,7692	0,5401	0,5875
Nur <i>singleton</i> Duplikationen	0,3606	0,2224	0,7901	0,7865
Nur <i>singleton</i> Deletionen	0,8720	0,9001	0,1687	0,1827

11.6.2 Majore Depression

Tab. 11.2: Gesamtübersicht der burden Tests bei Majorer Depression. Keiner der durchgeführten Tests war mit der Erkrankung assoziiert. Modifiziert nach Degenhardt und Priebe et al., 2012.

Vergleichsgruppe	Test			
	CNV Frequenz		CNV Länge	
	RATE-Test	PROP-Test	TOTKB-Test	AVGKB-Test
Patienten (575) gegen Kontrollen (1.618)				
Alle CNVs	0,56	0,53	0,64	0,57
Nur Duplikationen	0,83	0,76	0,40	0,22
Nur Deletionen	0,23	0,38	0,14	0,14
Alle <i>singleton</i> CNVs	0,51	0,25	0,23	0,51
Nur <i>singleton</i> Duplikationen	0,64	0,46	0,79	0,56
Nur <i>singleton</i> Deletionen	0,68	0,76	0,17	0,15

11.6.3 Schizophrenie

Tab. 11.3: Gesamtübersicht der *burden* Tests bei Schizophrenie. Keiner der durchgeführten Tests war mit der Erkrankung assoziiert.

Vergleichsgruppe	Test	
	RATE-Test	PROP-Test
Patienten (1.637) gegen Kontrollen (1.627)		
Alle CNVs	0,77	0,53
Nur Duplikationen	0,08	0,20
Nur Deletionen	0,07	0,09
Alle <i>singleton</i> CNVs	0,34	0,78
Nur <i>singleton</i> Duplikationen	0,08	0,22
Nur <i>singleton</i> Deletionen	0,47	0,57

Danksagung

Ich danke Herrn Prof. Dr. Markus M. Nöthen, der es mir ermöglichte, diese Arbeit am Institut für Humangenetik der Universität Bonn durchzuführen.

Herrn Prof. Dr. Sven Cichon, Institut für Humangenetik der Universität Bonn, danke ich sehr für die gute Betreuung und die Übernahme des 1. Referats.

Herrn Prof. Schultze, LIMES Institut der Universität Bonn, danke ich für die Übernahme des Korreferates. Ebenso danke ich Herrn Prof. Höfeld und Herrn Prof. Becker für die Komplettierung meiner Prüfungskommission.

Mein besonderer Dank gilt Franziska Degenhardt für die hervorragende Zusammenarbeit über die Jahre und Ausdauer bei unseren gemeinsamen Projekten.

Mein Dank gilt außerdem meinen Kollegen des Instituts für Humangenetik. Insbesondere möchte ich Rami Abou Jamra, Jessica Becker, Anne Böhmer, Felix Brockschmidt, Markus Draaken, Stefanie Heilmann, Timo Hess, Tillmann Kleppi, Florian Siedek und Alexander Zink danken.

Meinen Eltern und Geschwistern, insbesondere meiner Mutter, danke ich sehr dafür, dass sie mich während meines Studiums in jedweder Hinsicht unterstützt haben.

Lebenslauf

Name: Lutz Priebe
Geburtsdatum: 15.07.1982
Geburtsort: Troisdorf
Staatsangehörigkeit: Deutsch
Familienstand: Ledig, keine Kinder
Schulausbildung: 1989-1993 Alfred-Delp Grundschule, Troisdorf
1993-2002 Gymnasium zum Altenforst, Troisdorf
Juni 2002 Allgemeine Hochschulreife
Hochschulstudium: 2002-2009 Studium der Biologie an der Rheinischen Friedrich-Wilhelms Universität Bonn
Oktober 2004 Diplom-Vorprüfung
Januar 2008 Diplomprüfung
2008-2009 Diplomarbeit am Institut für Humangenetik des Universitätsklinikums Bonn unter Betreuung von PD Dr. Sven Cichon mit dem Titel: „Systematische Identifikation von Kopienzahlvarianten (Copy Number Variants – CNVs) im Genom des Menschen: Untersuchung ihres Einflusses auf die Entstehung der bipolaren Störung“
Februar 2009 Erhalt des Diploms
2009-2013 Doktorarbeit am Institut für Humangenetik des Universitätsklinikums Bonn unter Betreuung von Prof. Dr. Sven Cichon mit dem Titel: „Identifikation von Kopienzahlvarianten und deren Assoziation mit neuropsychiatrischen Erkrankungen“

Lutz Priebe

Troisdorf, den 07.02.2014

Erklärung

Hiermit versichere ich an Eides statt, dass

1. die vorgelegte Arbeit – abgesehen von den ausdrücklich bezeichneten Hilfsmitteln – persönlich, selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde,
2. die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte unter Angabe der Quelle kenntlich gemacht sind,
3. die vorgelegte Arbeit oder ähnliche Arbeiten nicht bereits anderweitig als Dissertation eingereicht worden ist bzw. sind, sowie eine Erklärung über frühere Promotionsversuche und deren Resultate,
4. für die Erstellung der vorgelegten Arbeit und/oder die Gelegenheit zur Promotion keine fremde Hilfe, insbesondere keine entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberatern / -vermittlern oder anderen Personen) in Anspruch genommen wurde.

Lutz Priebe

Troisdorf, den 07.02.2014