



**HAL**  
open science

## Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder

Sven Cichon, Lutz Priebe, Franziska A. Degenhardt, Stefan Herms, Britta Haenisch, Manuel Mattheisen, Vanessa Nieratschker, Moritz Weingarten, Stephanie H. Witt, René Breuer, et al.

### ► To cite this version:

Sven Cichon, Lutz Priebe, Franziska A. Degenhardt, Stefan Herms, Britta Haenisch, et al.. Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder. *Molecular Psychiatry*, 2011, 10.1038/mp.2011.8 . hal-00616286

**HAL Id: hal-00616286**

**<https://hal.science/hal-00616286>**

Submitted on 22 Aug 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder**

L Priebe<sup>1,2</sup>, F Degenhardt<sup>1,2</sup>, S Herms<sup>1,2</sup>, B Haenisch<sup>1,2</sup>, M Mattheisen<sup>1,2,3</sup>, V Nieratschker<sup>4</sup>, M Weingarten<sup>1,2</sup>, S Witt<sup>4</sup>, R Breuer<sup>4</sup>, T Paul<sup>4</sup>, M Alblas<sup>1,2</sup>, S Moebus<sup>5</sup>, M Lathrop<sup>6</sup>, M Leboyer<sup>7,8,9</sup>, S Schreiber<sup>10</sup>, M Grigoriou-Serbanescu<sup>11</sup>, W Maier<sup>12</sup>, P Propping<sup>2</sup>, M Rietschel<sup>4</sup>, MM Nöthen<sup>1,2</sup>, S Cichon<sup>1,2,13,14</sup>, TW Mühleisen<sup>1,2,14</sup>

<sup>1</sup>*Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany*

<sup>2</sup>*Institute of Human Genetics, University of Bonn, Bonn, Germany*

<sup>3</sup>*Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany*

<sup>4</sup>*Central Institute of Mental Health, Division of Genetic Epidemiology in Psychiatry, Mannheim, Germany*

<sup>5</sup>*Institute for Medical Informatics, Biometry and Epidemiology, University Clinic Essen, Essen, Germany*

<sup>6</sup>*Centre National de Génotypage, Evry Cedex, France*

<sup>7</sup>*INSERM U-513, Faculté de Médecine, Créteil, France*

<sup>8</sup>*University of Paris, Faculty of Medicine, IRF10, Créteil, France.*

<sup>9</sup>*AP-HP, Albert Chenevier and Henri Mondor Hospitals, Department of Psychiatry, Créteil, France*

<sup>10</sup>*Institute of Clinical Molecular Biology, Christian Albrechts University, Kiel, Germany*

<sup>11</sup>*Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania*

<sup>12</sup>*Department of Psychiatry, University of Bonn, Bonn, Germany*

<sup>13</sup>*Institute of Neuroscience and Medicine (INM-1), Structural and Functional Organization of the Brain, Genomic Imaging, Research Center Juelich, Juelich, Germany*

<sup>14</sup>*These authors contributed equally to this work*

Correspondence to: Prof Dr S Cichon, Department of Genomics, Institute of Human Genetics, Life and Brain Center, University of Bonn, Sigmund-Freud-Str. 25, D-53127 Bonn, Germany.

E-mail: [sven.cichon@uni-bonn.de](mailto:sven.cichon@uni-bonn.de)

**Abstract:**

We used genome-wide SNP data to search for the presence of CNVs in 882 patients with bipolar disorder (BD) and 872 population-based controls. A total of 291 (33%) patients had an early age-at-onset  $\leq 21$  years ( $AO \leq 21y$ ). We systematically filtered for CNVs that cover at least 30 consecutive SNPs and which directly affect at least one RefSeq gene. We tested whether: (a) the genome-wide burden of these filtered CNVs differed between patients and controls, and (b) the frequency of specific CNVs differed between patients and controls. Genome-wide burden analyses revealed that the frequency and size of CNVs did not differ substantially between the total samples of BD patients and controls. However, separate analysis of patients with  $AO \leq 21y$  and  $AO > 21y$  showed that the frequency of microduplications was significantly higher ( $P=0.0004$ ) and the average size of singleton microdeletions was significantly larger ( $P=0.0056$ ) in patients with  $AO \leq 21y$  compared to controls. A search for specific BD-associated CNVs identified two common CNVs: (a) a 160 kb microduplication on 10q11 was overrepresented in  $AO \leq 21y$  patients (9.62%) compared to controls (3.67%;  $P=0.0005$ ), and (b) a 248 kb microduplication on 6q27 was overrepresented in the  $AO \leq 21y$  subgroup (5.84%) compared to controls (2.52%,  $P=0.0039$ ). These data suggest that CNVs have an influence on the development of early-onset, but not later-onset BD. Our study provides further support for previous hypotheses of an etiological difference between early-onset and later-onset BD.

Keywords: bipolar disorder; copy number variant; genome-wide burden; early age-at-onset; association

Running title: SNP-based search for CNVs in bipolar disorder

## Introduction:

Bipolar disorder (BD) is a common, severe mood disorder that is characterized by recurring episodes of extreme exaltation (mania) and depression. Mood symptoms are often accompanied by disturbances in thinking and behavior. BD has a lifetime risk in the general population of 0.5-1.5%.<sup>1</sup> Twin, family, and adoption studies have provided strong evidence for a genetic predisposition to BD.<sup>2,3</sup> Heritability has been estimated to be between 60% and 85%.<sup>4</sup>

Changes in the copy number of submicroscopic chromosomal segments, known as copy number variants (CNVs), are a major component of the difference between human genomes.<sup>5</sup> Based on their frequency, rare (<1%) and common (>1%) CNVs can be distinguished. Several recent studies have shown a strong influence of rare CNVs on the development of neuropsychiatric phenotypes such as autism<sup>6,7</sup> and schizophrenia.<sup>8-10</sup> To date, only a limited number of studies have been published for BD.<sup>11-15</sup> Lachman *et al.* investigated a mixed cohort of Caucasian patients and controls from the Czech Republic and the United States, and found that microdeletions and microduplications, affecting the gene *glycogen synthase kinase 3 beta (GSK3B)* were significantly increased in patients.<sup>11</sup> Zhang *et al.* investigated singleton microdeletions (i.e. those occurring only once in the total dataset of patients and controls) of >100 kb in their European American sample of 1 001 BD patients and 1 034 controls, and found that they were overrepresented in patients.<sup>12</sup> Recently, Yang *et al.* published a study of a three-generation Older Amish pedigree with segregating affective disorder.<sup>13</sup> They reported that a set of four CNVs on chromosomes 6q27, 9q21, 12p13, and 15q11 were enriched in affected family members, and that these altered the expression of neuronal genes. Grozeva *et al.*<sup>14</sup> screened a sample of 1 868 patients with BD and 2 938 controls for large (>100 kb) and rare (found in <1% of the population) CNVs. No specific CNV was associated with BD, and the authors found no increased genome-wide burden of CNVs in patients compared to controls. The Wellcome Trust Case Control Consortium (WTCCC<sup>15</sup>) investigated common CNVs (found in >5% of the population) of >0.5 kb in a sample of 2007 patients with BD and 3 000 controls, and found no association between CNVs and the disease.

In the present study, we screened the genome-wide SNP data of 882 patients with BD and 872 population-based controls for predicted common and rare CNVs using more than 540 000 autosomal and X-chromosomal markers. All study participants were of German descent. We tested both the overall group of BD patients and a subgroup with an age-at-onset of  $\leq 21$  years since several studies have suggested that early-onset BD patients may represent a clinically and genetically more homogeneous subtype of BD. Clinical studies have demonstrated that early-onset BD is a more severe form of the disorder that is characterized by frequent psychotic features, more

mixed episodes, greater psychiatric co-morbidity, and poorer response to prophylactic lithium treatment.<sup>16-18</sup> Familial aggregation is more pronounced in relatives of early-onset BD patients than in relatives of later-onset BD patients.<sup>16-20</sup> Finally, the findings of several studies have suggested the existence of an intra-familial correlation for age-at-onset among bipolar siblings,<sup>21,22</sup> and a segregation analysis has shown that BD is transmitted differentially in early- and later-onset BD families.<sup>20</sup>

To control for the number of technical artifacts, we developed a stringent protocol for quality control (QC) and filtering. On the basis of these data, we conducted statistical tests for the genome-wide burden of CNVs and for all specific common and rare CNVs that were found to be associated with BD.

## **Materials and Methods:**

### *Sample description*

Unrelated patients with a clinical history of bipolar disorder (post QC: type I, n=767; type II, n=102; not other specified, n=13) were recruited at two centers: the Central Institute of Mental Health, Mannheim, and the Department of Psychiatry and Psychotherapy of the University of Bonn. The study was approved by the Institutional Review Boards, and all patients provided written informed consent prior to inclusion. DSM-IV life-time diagnoses of BD were assigned using a consensus best-estimate procedure that was based on all available information including the findings of a structured SCID-I interview.<sup>23</sup> The same set of instruments was used by both centers.<sup>24</sup> Age-at-onset (AO) was defined as the age at which the first DSM-IV-criteria episode of either depression or mania had occurred. Post QC, the mean age of patients at the time of recruitment was 44.03 years with a standard deviation (SD) of 13.41, the mean age-at-onset was 27.90 years (SD=11.28; median=24; mode=19). The AO distribution of the total sample deviated to the right of the Gaussian distribution (Kolmogorov-Smirnov  $Z=5.10$ ;  $P<0.001$ ; positive skewness=1.10). The male/female ratio was 0.47.

### *Determining the cut-off point for early age-at-onset*

Despite extensive debate over the past decade, no consensus has yet been reached concerning the cut-off point for the definition of early and late AO in BD. Authors who have applied the same expectation-maximization algorithm to different samples have described divergent cut-offs for the definition of the early AO. Some have reported that a three-AO-group distribution best fitted their data<sup>25,26</sup> and others a two-AO-group distribution.<sup>27</sup> This demonstrates that the results of an

admixture analysis are sample-dependent.

We therefore performed a commingling analysis in the present sample before selecting the cut-off point for the definition of early AO. We used the SEGREG-subroutine of the software S.A.G.E. (version 6.1, <http://darwin.cwru.edu/sage/>).<sup>28</sup> Commingling analysis reveals the distribution mixture of a trait through segregation analysis while allowing for ascertainment correction. Class D models are used as the regressive models in commingling analysis, which assume that the trait under investigation in the study probands is not conditional upon the trait in antecedent family members. The model that fitted the data best was selected on the basis of the smallest value of the Akaike Information Criterion.

Although the two-AO-group and the three-AO-group models had fitted our data equally well in a preliminary analysis of a larger sample,<sup>29</sup> the best model in the present sample was a two-AO-group distribution. Prior to QC, the mean AO in the early onset group was 20.67 years (SD=8.40) and the mean AO in the late onset group was 33.20 years (SD=10.30). To select patients with a clear early AO, we only selected patients with an AO that was lower or equal to the rounded mean of the early onset group, i.e. age 21 (AO $\leq$ 21y). In this AO $\leq$ 21y subgroup, the mean AO was 17.44 years (SD=2.59).

Prior to QC, our overall sample consisted of 957 patients and 880 controls. Post QC, this was reduced to 882 patients and 872 controls. A total of 291 patients remained in the AO $\leq$ 21y subgroup and for these patients the mean age-at-recruitment was 38.71 years (SD=12.88), the mean AO was 17.54 years (SD=2.52), and the male/female ratio was 0.43. In the later-onset subgroup with an AO $>$ 21 years (AO $>$ 21y, n=591), the mean age-at-recruitment was 46.84 years (SD=12.87), and the mean AO was 33.20 years (SD=10.30), and the male/female ratio was 0.49.

Controls were drawn from two population-based epidemiological studies: (a) Population-based Recruitment of Patients and Controls for the Analysis (PopGen, n=497)<sup>30</sup> from Schleswig-Holstein (Northern Germany); and (b) the Heinz Nixdorf Recall study (Risk Factors, Evaluation of Coronary Calcification, and Lifestyle; HNR, n=383)<sup>31</sup> from Essen, Bochum, and Mülheim a. d. Ruhr (Ruhr area). Post QC, the mean age at recruitment of the control group was 47.98 (SD=11.42), and the male/female ratio was 0.51. All patients and controls reported that they were of German descent.

#### *DNA extraction, genotyping, and quality control*

Venous blood samples were collected from all patients and controls. Lymphocyte DNA was extracted either by salting out with saturated sodium chloride solution<sup>32</sup> or by a Chemagic Magnetic Separation Module I (Chemagen, Baesweiler, Germany) according to the manufacturer's recommendations.

Individuals were genotyped using Illumina's HumanHap550v3 (HH550) or Human610-Quadv1 (H610Q) BeadArrays. The genotype data had been generated as part of a genome-wide association study of BD (Cichon, Mühleisen et al., unpublished data). The H610Q chip contains approximately 60 000 more probes and SNPs than the HH550 array. The majority of the excess content represents non-polymorphic CNV probes, which leads to an excess of CNV calls for individuals genotyped on the H610Q chip (data not shown). To avoid such a bias, we only analyzed SNPs that are present on both chips, i.e. a total of 541 524 SNPs (post QC).

Prior to computational CNV prediction, stringent QC criteria were applied to the genotype data at both the marker and the individual level. SNPs with a call rate of <98% were excluded. Individuals were excluded for the following reasons: (a) DNA call rate <97% (20 patients);(b) differences between X-chromosomally inferred and phenotypic sex (six patients); (c) DNA sample doublets identified by identity-by-state estimates (defined as  $IBS=2.0$ , two controls); (d) relatedness of individuals ( $1.6 \leq IBS < 2.0$ , no individual excluded); and (e) population outlier according to multi-dimensional scaling with HapMap phase 2 (one patient and three controls were excluded prior to the present CNV study).

#### *CNV detection*

CNVs were predicted using the program QuantiSNP (version 1.1, <http://www.well.ox.ac.uk/QuantiSNP>).<sup>33</sup> The algorithm implemented in QuantiSNP uses an Objective Bayes Hidden-Markov Model to estimate the copy number. To evaluate the presence of a CNV, QuantiSNP uses the normalized intensity data (i.e. log R ratio) and allele frequency data (i.e. B allele frequency) of each SNP. Both values were calculated by Illumina's BeadStudio Genotyping module (version 3.3.7, <http://www.illumina.com/pages.ilmn?ID=169>). Individuals were excluded if their SD from the log R ratio or their B allele frequency exceeded certain thresholds: log R ratio >0.36 or B allele frequency >0.12 (22 patients, six controls). We employed the normalization procedure for local GC content implemented in QuantiSNP to improve the accuracy of detection. The Log Bayes Factor (LBF) was computed for each CNV. This factor indicates the confidence of each predicted CNV, with higher values indicating higher statistical reliability.

To minimize the number of false-positive CNV calls, we only considered CNVs with a LBF  $\geq 30$ , that spanned a minimum of 30 consecutive SNPs, and which directly affect at least one RefSeq gene.<sup>34</sup> After applying these filters, we also excluded individuals with more than seven CNVs, since they were extreme outliers in terms of the number of CNV events (27 patients). Following quality control, 1 044 CNV calls from a total of 882 BD patients and 872 controls were statistically analyzed. A total of 291 of the patients (33%) had an AO  $\leq 21$  years, and 591 patients had an AO >21 years.

To confirm our QuantiSNP-based CNV results, we additionally screened our dataset with PennCNV.<sup>35</sup> Both PennCNV and QuantiSNP apply a Hidden-Markov Model to estimate the copy number of an individual. They also take the log R ratio and B allele frequency of each SNP into account, and correct for GC content. The CNV data of both algorithms are available upon request.

#### *Statistical analysis of CNV burden and specific CNVs*

All association tests for genome-wide CNV burden and association of specific CNVs were performed using PLINK (version 1.06, <http://pngu.mgh.harvard.edu/purcell/plink/>).<sup>36</sup> We conducted the burden tests for CNV frequency (PROP, RATE) as well as for CNV length (TOTKB, AVGKB), i.e. the total number of CNVs in patients vs. controls (RATE); proportion of individuals with one or more CNVs in patients vs. controls (PROP); total length spanned by CNVs per individual in the patient group vs. the control group (TOTKB), and average size of CNVs per individual in the patient group vs. the control group (AVGKB). All *P*-values were generated using 50 000 permutations.

We defined three major comparison groups for the genome-wide burden analyses:

1. All patients vs. all controls
2. AO $\leq$ 21y patients vs. controls
3. AO $>$ 21y patients vs. controls

We tested each of these three comparison groups for six different categories of CNVs:

1. All CNVs
2. Microduplications
3. Microdeletions
4. All singleton CNVs
5. Singleton microduplications
6. Singleton microdeletions

We performed a total of four different burden tests (RATE, PROP, TOTKB, and AVGKB) in 18 test groups, i.e. three major comparison groups for the six different categories of CNVs as described above, resulting in a total of 72 tests for association between CNV burden and BD. To account for all 72 tests, we also applied Bonferroni's method, although this procedure may be too conservative given that the tests were not independent of each other.

In addition, we monitored the distribution of CNVs in chromosomal regions 1q21, 2p16, 7q34-36, 15q11, 15q13, 16p11, 17p12, and 22q11 which have previously been reported to be associated with



a variety of neuropsychiatric disorders (Table 2). Since the borders of these regions are known, we relaxed our filter criteria, i.e. we included all CNVs with  $LBF \geq 10$  which were visually inspected by two independent investigators regardless of the number of affected SNPs. Association tests for these CNVs were performed using Fisher's exact test.

#### *Verification of specific CNVs*

All CNVs identified by QuantiSNP and PennCNV that had been found to be associated with neuropsychiatric disorders in previous studies, or which were located within chromosomal regions associated with BD, were visually inspected using Illumina's GenomeStudio.

The specific CNVs that were found to be associated with BD in the present study (6q27 and 10q11) were verified by quantitative real-time PCR (qPCR) using TaqMan Copy Number Assays (Applied Biosystem, Foster City, CA, USA). We confirmed each CNV carrier by qPCR and also tested non-CNV carriers (as defined by QuantiSNP and PennCNV) to detect possible CNV carriers who had not been identified by QuantiSNP and PennCNV. The status of all CNV carriers was confirmed by qPCR, and no CNV carriers were detected among the putative non-CNV carriers tested. Copy numbers were calculated using the  $\Delta\Delta C_t$  method implemented in the CopyCaller Software (v1.0, <http://www.appliedbiosystems.com/support/software/copycaller/>).

#### *Analysis of pathways and biological processes*

We analyzed whether genes affected by CNVs were enriched in certain pathways or biological processes using the web-based the program Ingenuity Pathways Analysis platform (IPA, version 8.0, <http://www.ingenuity.com>). IPA is based on functional annotation and molecular interactions. Gene lists were assembled using RefSeq genes that are affected by the CNVs identified in the burden analysis (microduplications, singleton deletions). Lists were uploaded into IPA and investigated using the "core analysis" function and default settings. In the functional analysis, biological functions were grouped into different categories from the Ingenuity Knowledge Base. To calculate the statistical significance of pathways and biological processes assigned to gene sets, *P*-values of the Fisher's exact test were corrected by the Benjamini-Hochberg method.

### **Results:**

We systematically analyzed our samples for significant genome-wide differences in the distribution of all CNVs between patients and controls (genome-wide burden tests) as well as for a significant overrepresentation of specific CNVs in patients or controls.

### *General description of the CNV dataset*

Following the baseline QC of SNP data, QuantiSNP identified a total of 124 146 putative CNVs in the initial sample of 957 patients and 880 controls (an average of 67.6 CNVs per individual). Following the application of all QC filters, 1 044 potential CNVs remained in the filtered sample of 882 BD patients and 872 controls (an average of 0.59 CNVs per individual).

We examined the distribution of the number of CNVs per individual and identified a total of 27 extreme outliers in terms of CNV observations, with more than seven CNVs being detected in each individual. The samples of a total of 24 of these individuals were clustered in the outer rows of the same 96-well plate, and thus these findings are likely to represent plate effects. These individuals were excluded from the downstream analyses.

### *Association analysis for genome-wide CNV burden*

Overall, 10 of the 72 genome-wide burden tests revealed nominally significant differences in CNV burden between patients and controls. In the following, we provide a detailed description of the most important findings, as outlined in Table 1.

*All patients vs. controls.* In the total sample of 882 BD patients and 872 controls, the genome-wide burden of singleton microdeletions showed nominally significant association for two out of the four tests performed. The average total length of all singleton microdeletions per individual was 487.6 kb in patients compared to 265.1 kb in controls (TOTKB:  $P=0.014$ ). The average size per singleton microdeletion was 472.6 kb in patients and 249.3 kb in controls (AVGKB:  $P=0.014$ ). The PROP- and RATE-tests generated no significant P-values.

*Patients with an early AO vs. controls.* When comparing the 291  $AO \leq 21y$  patients with all controls, we again found that singleton microdeletions in patients were, on average, larger (661.7 kb in  $AO \leq 21y$  patients vs. 249.3kb in controls, AVGKB:  $P=0.0056$ ) and spanned longer chromosomal regions per individual than in controls (679.6 kb in  $AO \leq 21y$  patients vs. 261.2 kb in controls, TOTKB:  $P=0.0084$ ). Furthermore, we observed that the total proportion of individuals with at least one microduplication (44.3% in  $AO \leq 21y$  patients vs. 33.1% in controls, PROP:  $P=0.00040$ ), at least one CNV (52.9% in  $AO \leq 21y$  patients vs. 42.3% in controls, PROP:  $P=0.00092$ ), or at least one singleton microduplication (17.2% in  $AO \leq 21y$  patients vs. 11.9% in controls, PROP:  $P=0.017$ ) was significantly higher in this BD subgroup. The PROP test  $P$ -value for microduplications, which was the most significant of all of the burden analyses, withstood correction for multiple testing with the Bonferroni method which accounted for the number of all tests performed in this study ( $n=72$ ,  $P_{\text{adjusted}}=0.029$ ). However, there were no significant differences in the proportion of individuals who carried either microdeletions or singleton microdeletions.

*Patients with AO>21y vs. controls.* In the third test group, we analyzed the burden of CNVs in the AO>21y subgroup (n=591) vs. all controls. These tests revealed no significant differences in CNV burden between patients and controls.

#### *Association analyses of specific CNVs*

We identified two common microduplications that were significantly overrepresented in patients compared to controls: (a) a 248 kb microduplication on chromosome 6q27, and (b) a 160 kb microduplication on chromosome 10q11 (Figure 1).

The 10q11 microduplication (Figure 1a) was observed in 53 patients (6.01%) and in 32 controls (3.67%,  $P=0.035$ , OR [95%-CI] =1.53 [1.05-2.72]). Of these 53 patients, 28 belong to the AO?21y subgroup (9.62% in AO≤21y patients vs. 3.67% in controls,  $P=0.00052$ , OR [95%-CI] =2.79 [1.59-4.89]). Following genome-wide correction using permutation, this P-value remained significant (n=50 000;  $P_{\text{adjusted}}=0.032$ ). In view of the genetic marker resolution provided by our approach, all observed microduplications at the 10q11 locus appear to have the same breakpoints (length approximated 160 kb, chr10:47.01-47.17 Mb, NCBI build 36), and carry 30 consecutive SNPs. This CNV covers the complete gene *anthrax toxin receptor-like gene (ANTXRL)*.

The microduplication on chromosome 6q27 (Figure 1b) was detected in 17 patients from the AO≤21y subgroup (5.84%) and in 22 controls (2.52%,  $P=0.0039$ , OR [95%-CI] =2.40 [1.18-4.80]). There were no significant differences in distribution when all BD patients were compared to controls. In CNV carriers, slight differences in CNV size were observed, with a shared overlap of around 248 kb (chr6:168.09-168.33 Mb, NCBI build 36) that was due to 110 adjacent SNPs. Three genes are affected by this microduplication: (a) *kinesin family member 25 (KIF25)*, (b) *FERM domain containing 1 (FRMD1)*, and (c) parts of the 3' terminus of mixed-lineage leukemia translocated to 4 (*MLLT4*). The gene *dapper, antagonist of beta-catenin, homolog 2 (DACT2)* lies around 115 kb downstream from this common microduplication. DACT2 participates in the WNT signaling pathway<sup>37,38</sup> which is known to regulate neurogenesis and neuroprotection.

In a follow-up analysis, we tested whether patients carrying either the common CNVs on 6q27 or the CNV on 10q11 showed differences in sex distribution or family history of psychiatric disorder compared to patients who were non-carriers. No significant associations were observed for either phenotypic item (data not shown).

#### *Specific CNVs at loci previously associated with psychiatric disorder*

We tested whether CNVs in one of six genomic regions that have previously been reported to be associated with neuropsychiatric disorders (1q21, 2p16, 7q34-36, 15q11, 15q13, 16p11, 17p12 and 22q11) were overrepresented in our BD patients. CNVs in these regions were not significantly

overrepresented (Table 2). However, the power of our sample to find significant association with these rare CNVs was low.

#### *Pathways and biological processes impacted by CNVs in early-onset patients*

To further characterize the two top association findings of our burden analyses in  $AO \leq 21$ y patients (Table 1), we used IPA to explore possible functional relationships between genes covered by either microduplications or singleton microdeletions. In the analysis of the 46 genes hit by singleton microdeletions, the significant first five top hits for biological processes (i.e. "drug metabolism", "lipid metabolism", "molecular transport", "small molecular biochemistry" and "endocrine system development and function") were enriched by the presence of the genes *SLCO1B1* and *SLCO1B3*, which are both located on 12p12. These were covered by a singleton microdeletion in one  $AO \leq 21$ y patient. These categories were therefore omitted from data interpretation. Further top process categories that were significantly overrepresented due to several input genes being hit by singleton microdeletions included "endocrine system disorder", "genetic disorder", "metabolic disease", "immunological disease," and "infectious disease" (Table 3). The same functional categories were found to be enriched when analyzing the 287 genes affected by microduplications in  $AO \leq 21$ y patients (Table 4) although there was no overlap in affected genes between the two gene lists, except for *MAD1L1*.

No significant result was obtained for canonical pathways in IPA after correction for multiple testing (data not shown). This was probably due to the limited number of genes introduced into the analysis.

#### **Discussion:**

In the present study, we investigated a large sample of patients and controls of German descent for the presence of CNVs that may be involved in the development of BD. Our analyses included tests to monitor the genome-wide burden of CNVs (frequency and size) between patients and controls as well as tests to detect specific common and rare CNVs that were significantly overrepresented in either patients or controls. Since clinical and formal genetic studies have suggested that BD with an  $AO \leq 21$ y may be genetically distinct from BD with an  $AO > 21$  years,<sup>16-22</sup> we also analyzed these subgroups separately. The basis for the statistical analyses was provided by high-quality CNV prediction (QuantiSNP) using the SNP intensity data of 1 754 individuals and verification (PennCNV, visual inspection, and TaqMan for specific CNVs). Parameters were specified to ensure that only relatively large CNVs (detected by  $\geq 30$  consecutive SNPs) with high statistical confidence

(QuantiSNP: log Bayes factor  $\geq 30$ ; PennCNV: confidence value  $\geq 30$ ) passed QC. Using such stringent criteria is always a trade-off between sensitivity and quality of data. We have chosen our criteria to reduce type I error and are aware that this may lead to an increase of type II error at the same time.

Our genome-wide burden tests (RATE and PROP) provided no evidence that the overall number of singleton CNVs (both microdeletions and microduplications) overlapping with RefSeq genes was enriched in the total sample of BD patients compared with controls. The TOTKB and AVGKB tests demonstrated that the length of singleton microdeletions differed significantly between patients and controls (Table 1). Singleton microdeletions in patients were approximately twice the size (490 kb) of those in controls (260 kb; AVGKB  $P=0.014$ ). Separate analyses of the  $AO \leq 21y$  subgroup and the  $AO > 21y$  subgroup demonstrated that this effect was mainly attributable to the  $AO \leq 21y$  subgroup, in which the average size of microdeletions was approximately 680 kb (AVGKB  $P=0.0056$ ), and to a lesser extent to the  $AO > 21y$  subgroup, in which the average size of singleton microdeletions (348 kb) was not significantly larger than in controls (AVGKB  $P=0.17$ ; data not shown). The most significant finding of our burden analyses was that the total number of patients with at least one microduplication (both common as well as singleton microduplications) was significantly higher in the  $AO \leq 21y$  group, but not in the  $AO > 21y$  subgroup (PROP,  $P=0.0004$ ). The burden results suggest that both a higher CNV load and a larger CNV size are associated with BD in patients with an  $AO \leq 21y$ . The effect in the early-onset group is was attributable to longer singleton deletions as well as to a higher frequency of microduplications (common and rare). Analysis of either the overall sample or the  $AO > 21y$  sample alone produced only marginal evidence that the genome-wide burden of CNVs plays a role in disease development.

These results show a similar trend to those of Zhang *et al.*<sup>12</sup> who reported a significant overrepresentation of singleton deletions of more than 100 kb in their BD cases (16.2%) compared to controls (12.3%;  $P=0.007$ ). Interestingly, they also found that this effect was more pronounced in an early AO form of BD (age of mania onset  $\leq 18$  years). This is consistent with our observation of the presence of longer singleton deletions in early-onset BD patients. However, Zhang *et al.*<sup>12</sup> found no evidence of a higher frequency of microduplications in their early-onset mania subsample. Unfortunately, a more exact comparison of our data with that of Zhang *et al.*<sup>12</sup> is limited by the methodological differences between the two studies: Zhang *et al.*<sup>12</sup> used Affymetrix 6.0 arrays whereas we used Illumina HH550/H610Q arrays; Zhang *et al.*<sup>12</sup> performed CNV detection with Birdsuite whereas we used QuantiSNP and PennCNV; Zhang *et al.*<sup>12</sup> took all CNVs above a certain size threshold into account whereas we filtered for CNVs that overlapped with at least 30 consecutive SNPs and RefSeq genes.

Another recent study by Yang *et al.*<sup>13</sup> found no evidence for any significant association between the average number and size of CNVs and affective disorders (including BD and major depression). However, their study investigated a single three-generation Old Order Amish family, and their main focus was upon the identification of CNVs that co-segregated with disease status across generations. The limited number of affected (n=19) and unaffected family members (n=32) and their genetic relatedness, as well as the broader phenotype definition used may have resulted in limited power to investigate the genome-wide burden aspects of CNVs in BD.

Two recent genome-wide studies of CNVs in BD have been published (Grozeva *et al.*<sup>14</sup>, WTCCC<sup>15</sup>), which investigated common and rare CNVs. They tested for association between BD and specific CNVs and CNVs at previously reported loci, as well as for genome-wide burden. Neither of these two studies found evidence for the involvement of CNVs in disease development. The majority of patients and controls investigated by Grozeva *et al.*<sup>14</sup> were also analyzed by the WTCCC<sup>15</sup>, and thus the results of the two studies are not entirely independent of each other. Major methodological differences exist between these two studies and our study, which hamper any direct comparison of the results. Firstly, different arrays were used to screen for CNVs (Grozeva *et al.*<sup>14</sup>: Affymetrix' Genechip Human Mapping 500K Array Set; WTCCC<sup>15</sup>: Agilent Comparative Genomic Hybridization arrays). Secondly, Grozeva *et al.*<sup>14</sup> searched for rare CNVs (MAF <1%) only, whereas the WTCCC<sup>15</sup> investigated common CNVs (MAF >5%) only. The present study did not have any restrictions with regard to CNV frequency, and the frequencies of the two specific BD-associated CNVs on 10q11 (3.68% in controls) and 6q27 (2.53% in controls) are within a frequency range that was not investigated by Grozeva *et al.*<sup>14</sup> or the WTCCC.<sup>15</sup> Another important difference is the separate analyses of AO subgroups in the present study, which suggest that CNVs play a role in early onset, but not in later onset BD. Since the aforementioned studies did not take AO into account, it is unclear whether such an effect was present in their BD samples. Nonetheless, all studies, including the present study, are in agreement that CNVs do not appear to influence BD when AO is not taken into account.

In a subsequent step of our genome-wide burden analysis, we performed an IPA to investigate whether the genes affected by microduplications (n=287) or singleton microdeletions (n=46) in AO $\leq$ 21y patients were significantly enriched for biological processes or pathways. When removing single gene-based enriched categories, the top five significantly overrepresented biological function categories were the same for both the microduplications and the microdeletions gene lists, i.e. the disorder and disease processes "endocrine system disorder", "genetic disorder", "metabolic disease", "immunological disease," and "infectious disease". Follow-up studies in independent samples are clearly necessary to confirm that biological functions within these disease categories are involved in

the pathophysiology of (early-onset) BD. One interesting gene with prior suggestive evidence for association with BD, *mitotic arrest deficient-like 1 (MAD1L1)*, 7p22.3), was affected by a singleton microdeletion as well as a singleton microduplication in early-onset BD patients. Support for the involvement of this locus was provided by a recent genome-wide association study (GWAS) of BD, in which two *MAD1L1* SNPs (rs11764590 and rs10278591;  $r^2=0.7$ ) were the second and third best results in the meta-analysis step ( $P=1.28 \times 10^{-7}$  and  $P=1.81 \times 10^{-5}$ , respectively, unpublished data). *MAD1L1* is a mitotic spindle assembly checkpoint protein. Homozygous knockout of *MAD1L1* in mice confers embryonic lethality, indicating that *MAD1L1* plays an essential role during embryonic development.<sup>39</sup>

Two genes affected by microduplications in early AO patients, *epidermal growth factor receptor (EGFR)*, 7p11.2, one patient) and *nucleoredoxin (NXN)*, 17p13.3, one patient, two controls), were identified as susceptibility genes for BD by two recent GWAS.<sup>40,41</sup> *EGFR* rs17172438 ( $P=3.26 \times 10^{-5}$ , OR=1.32) and *EGFR* rs729969 ( $P=3.30 \times 10^{-5}$ , OR=1.36) were ranked among the top 20 findings in the GWAS by Sklar *et al.*<sup>40</sup> *EGFR* and its ligands are cell signaling molecules that mediate diverse downstream cellular functions, including cell proliferation and differentiation. In a further GWAS, Baum *et al.*<sup>41</sup> found that *NXN* rs2360111 ( $P=0.0003$ , OR=1.23) were associated with BD. The *NXN* gene encodes the protein nucleoredoxin which is involved in the inhibition of the WNT signaling pathway.<sup>42</sup>

Another aim of the present study was to identify specific BD-associated CNVs. We found that the frequency of two relatively common CNVs on 6q27 and 10q11 differed significantly between patients and controls. Common microduplications located on chromosome 10q11 showed association in the overall sample ( $P=0.035$ , OR=1.53). Again, this effect was much stronger in the AO $\leq$ 21y subgroup ( $P=0.00052$ , OR=2.79). The other microduplication on chromosome 6q27 was overrepresented in the AO $\leq$ ?21y subgroup ( $P=0.0039$ , OR=2.40), but not in the overall sample.

Interestingly, the 6q27 CNV region was implicated in BD in a three-generation Old Order Amish pedigree.<sup>13</sup> The 6q27 region was one of four regions that were enriched in affected pedigree members and which had an effect on the expression of genes within or near the rearrangement. The identification of the 6q27 region in two independent studies is clearly of interest, although further independent studies are required to support this finding. One limitation which prevents stronger conclusions being drawn from the Yang *et al.* study is that there was only modest co-segregation of the 6q27 CNV with BD. Thus the possibility that that this relatively common variant segregated within the family independently of disease cannot be excluded.

In summary, the present study found evidence for a significant association between BD and microdeletions and microduplications. Although the frequency of microdeletions was not

significantly higher in patients compared to controls, the size of singleton microdeletions was significantly larger. Our data also suggest that a higher frequency of microduplications is implicated in disease development. We found both the genome-wide burden of microduplications as well as common specific microduplications on 6q27 and 10q11 to be enriched in BD patients. A further important finding of our study is that CNVs were strongly associated in patients with an AO  $\leq 21$  years, but not in patients with an AO  $> 21$  years. In the overall BD sample, only a very weak association with CNVs was detected. Our results support for findings from previous clinical and formal genetic studies that early and later onset BD may represent genetically distinct forms of the disease. Future studies of CNVs in BD should therefore take the AO of their patient samples into account.



**Acknowledgements:**

We thank all of the patients who participated in this study. We also thank all of the probands from the community-based cohort of PopGen as well as that of the Heinz Nixdorf Recall (HNR) study, which was established with the support of the Heinz Nixdorf Foundation. The present study was supported by the German Federal Ministry of Education and Research (BMBF) within the context of the National Genome Research Network plus (NGFNplus) and the MooDS-Net (grant 01GS08144 to S.C. and M.M.N., grant 01GS08147 to M.R.). M.M.N. also received support from the Alfried Krupp von Bohlen und Halbach-Stiftung.

Some of the results of this study were obtained using the program package S.A.G.E., which is supported by a U.S. Public Health Service Resource Grant (RR03655) from the National Center for Research Resources.

We are grateful to J. Sebat (Department of Psychiatry, University of California, San Diego) for his critical review of the manuscript.

**Conflict of Interest statement:**

The authors declare they have no competing financial interests, as defined by Nature Publishing Group, or any other interests that might be perceived to influence the results and discussion reported in this paper.

## References:

- 1 Craddock N, Jones I. Genetics of bipolar disorder. *J Med Genet* 1999; **36**: 585-594.
- 2 Berrettini W. Progress and pitfalls: bipolar molecular linkage studies. *J Affect Disord* 1998; **50**: 287-297.
- 3 Kendler KS, Pedersen NL, Johnson L, Neale MC, Mathé AA. A pilot Swedish twin study of affective illness, including hospital- and population-ascertained subsamples. *Arch Gen Psychiatry* 1993; **50**: 699-700.
- 4 Burmeister M, McInnis MG, Zöllner S. Psychiatric genetics: progress amid controversy. *Nat Rev Genet* 2008; **9**: 527-540.
- 5 Sebat J. Major changes in our DNA lead to major changes in our thinking. *Nat. Genet* 2007; **39**: S3-S5.
- 6 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T *et al.* Strong association of de novo copy number mutations with autism. *Science* 2007; **316**: 445-449.
- 7 Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008; **358**: 667-675.
- 8 Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* 2008; **455**: 232-236.
- 9 The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**: 237-241.
- 10 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**: 539-543.
- 11 Lachman HM, Pedrosa E, Petruolo OA, Cockerham M, Papolos A, Novak T *et al.* Increase in GSK3beta gene copy number variation in bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144B**: 259-265.
- 12 Zhang D, Cheng L, Qian Y, Alliey-Rodriguez N, Kelsoe JR, Greenwood T *et al.* Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol Psychiatry* 2009; **14**: 376-380.
- 13 Yang S, Wang K, Gregory B, Berrettini W, Wang L, Hakonarson H *et al.* Genomic landscape of a three-generation pedigree segregating affective disorder. *PLoS One*, 2009; **4**: e4474.

- 14 Grozeva D, Kirov G, Ivanov D, Jones IR, Jones L, Green EK et al. Rare Copy Number Variants A Point of Rarity in Genetic Risk for Bipolar Disorder and Schizophrenia. *Arch Gen Psychiatry*. 2010; **67**:318-327.
- 15 The Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, 2010; **464**:713-720.
- 16 Weissman MM, Gershon ES, Kidd KK, Prusoff BA, Leckman JF, Dibble E *et al*. Psychiatric disorders in the relatives of probands with affective disorders. The Yale University-National Institute of Mental Health Collaborative Study. *Arch Gen Psychiatr* 1984; **41**: 13–21.
- 17 Schurhoff F, Bellivier F, Jouvent R, Mouren-Simeoni MC, Bouvard M, Allilaire JF *et al*. Early and late onset bipolar disorders: two different forms of manic-depressive illness? *J Affect Disord* 2000; **58**: 215–221.
- 18 Strober M, Morrell W, Burroughs J, Lampert C, Danforth H, Freeman R. A family study of bipolar I disorder in adolescence. Early onset of symptoms linked to increased familial loading and lithium resistance. *J Affect Disord* 1988; **15**: 255–268.
- 19 Faraone SV, Glatt SJ, Tsuang MT. The genetics of pediatric-onset bipolar disorder. *Biol Psychiatr* 2003; **53**: 970–977.
- 20 Grigoriou-Serbanescu M, Martinez M, Nothen MM, Grinberg M, Sima D, Propping P *et al*. Different familial transmission patterns in bipolar I disorder with onset before and after age 25. *Am J Med Genet* 2001; **105**: 765–773.
- 21 Baron M, Risch N, Mendlewicz J. Age at onset in bipolar-related major affective illness: clinical and genetic implications. *J Psychiatr Res* 1982; **17**: 5–20.
- 22 Leboyer M, Bellivier F, McKeon P, Albus M, Borrman M, Perez-Diaz F *et al*. Age at onset and gender resemblance in bipolar siblings. *Psychiatr Res* 1998; **81**: 125–131.
- 23 Spitzer RL, Williams JBW, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID) I: History, Rationale, and Description. *Arch Gen Psychiatry* 1992; **49**: 624-629.
- 24 Fangerau H, Ohlraun S, Granath RO, Nöthen MM, Rietschel M, Schulze TG. Computer-assisted phenotype characterization for genetic research in psychiatry. *Hum Hered* 2004; **58**: 122-30.
- 25 [Bellivier F](#), [Golmard JL](#), [Rietschel M](#), [Schulze TG](#), [Malafosse A](#), [Preisig M](#), et al. Age at onset in bipolar I affective disorder: further evidence for three subgroups. *Am J Psychiatry* 2003; **160**: 999-1001.
- 26 Kennedy N, Everitt B, [Boydell J](#), Van Os J, Jones PB, Murray RM. Incidence and distribution of first-episode mania by age: results from a 35-year study. *Psychol Med*. 2005; **35**: 855-863.
- 27 Hamshere ML, [Gordon-Smith K](#), Forty L, Jones L, Caesar S, Fraser C, et al. Age-at-onset in

bipolar-I disorder: mixture analysis of 1369 cases identifies three distinct clinical sub-groups. *J Affect Disord.* 2009; **116**: 23-29.

28 S.A.G.E. 6.1.0 [2010]. Statistical Analysis for Genetic Epidemiology <http://darwin.cwru.edu/sage/>.

29 Grigoriu-Serbanescu M, Rietschel M., Paul T, Schulze TG, Noethen MM, Cichon S, *et al.* Two or three age-of-onset groups in bipolar I disorder? Findings of commingling analysis in Romanian and German bipolar I patients. *Eur Psychiatry* 2010; **25**, Suppl 1, 1428.

30 Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 2006; **9**: 55-61.

31 Schmermund A, Möhlenkamp S, Stang A, Grönemeyer D, Seibel R, Hirche H *et al.* 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. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am Heart J* 2002; **144**: 212-218.

32 Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1215.

33 Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P *et al.* QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 2007; **35**: 2013-2025.

34 Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2005; **33**: D501-D504.

35 Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 2007; **17**: 1665-1674.

36 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analysis. *Am J Hum Genet*; 2007 **81**: 559-575.

37 Machon O, Backman M, Machonova O, Kozmik Z, Vacik T, Andersen L *et al.* A dynamic gradient of Wnt signaling controls initiation of neurogenesis in the mammalian cortex and cellular specification in the hippocampus. *Dev Biol* 2007; **311**: 223-237.

38 Toledo EM, Colombres M, Inestrosa NC. Wnt signaling in neuroprotection and stem cell

differentiation. *Prog Neurobiol* 2008; **86**: 281-296.

39 Iwanaga Y, Chi YH, Miyazato A, Sheleg S, Haller K, Peloponese JM Jr, *et al.* Heterozygous deletion of mitotic arrest-deficient protein 1 (MAD1) increases the incidence of tumors in mice. *Cancer Res* 2007, **67**: 160-166.

40 Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, *et al.* Whole-genome association study of bipolar disorder. *Mol Psychiatry* 2008, **13**, 558-569.

41 Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, *et al.* A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry* 2008, **13**: 197-207.

42 Funato Y, Michiue T, Asashima M, Miki H. The thioredoxin-related redox-regulating protein nucleoredoxin inhibits Wnt- $\beta$ -catenin signalling through dishevelled. *Nat Cell Biol* 2006; **8**: 501–508.

43 Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med.* 2008; **359**: 1685-1699.

44 Kim HG, Kishikawa S, Higgins AW, Seong IS, Donovan DJ, Shen Y, *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet.* 2008; **82**: 199-207.

45 Rujescu D, Ingason A, Cichon S, Pietiläinen OP, Barnes MR, Toulopoulou T *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* 2009; **18**: 988-996.

46 Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH, *et al.* CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol Psychiatry.* 2008, **13**: 261-266.

47 Sutcliffe JS, Han MK, Amin T, Kesterson RA, Nurmi EL, *et al.* Partial duplication of the APBA2 gene in chromosome 15q13 corresponds to duplicon structures. *BMC Genomics* 2003; **4**: 15.

48 Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, *et al.* A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet.* 2008; **40**: 322-328.

49 Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P, *et al.* Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet.* 2009, **18**: 1497-1503.

50 Niklasson L, Rasmussen P, Oskarsdóttir S, Gillberg C. Autism, ADHD, mental retardation and behavior problems in 100 individuals with 22q11 deletion syndrome. *Res Dev Disabil.* 2009; **30**:

763-773.

51 Hashimoto R, Okada T, Kato T, Kosuga A, Tatsumi M, Kamijima K, *et al.* The breakpoint cluster region gene on chromosome 22q11 is associated with bipolar disorder. *Biol Psychiatry* 2005, **57**: 1097-1102.

## Tables:

**Table 1** Association results for genome-wide burden of CNVs in BD.

Comparison group	Test			
	CNV frequency		CNV length	
	RATE <i>P</i>	PROP <i>P</i>	TOTKB <i>P</i>	AVGKB <i>P</i>
<b>All pat. (n=882) vs. con. (n=872)</b>				
All CNVs	-	-	-	-
Microduplications	-	-	-	-
Microdeletions	-	-	-	-
Singleton CNVs	-	-	-	-
Singleton microduplications	0.063	0.053	-	-
Singleton microdeletions	-	-	0.014	0.014
<b>AO<math>\leq</math>21y pat. (n=291) vs. con. (n=872)</b>				
All CNVs	0.087	0.00092	-	0.092
Microduplications	0.024	0.00040 <sup>a</sup>	-	-
Microdeletions	-	-	-	-
Singleton CNVs	0.045	0.087	0.081	0.094
Singleton microduplications	0.010	0.017	-	-
Singleton microdeletions	-	-	0.0084	0.0056
<b>AO<math>&gt;</math>21y pat. (n=591) vs. con. (n=872)</b>				
All CNVs	-	-	-	-
Microduplications	-	-	-	-
Microdeletions	-	-	-	-
Singleton CNVs	-	-	-	-
Singleton microduplications	-	-	-	-
Singleton microdeletions	-	-	-	-

Abbreviations: pat, patients; con, controls; *P*, *P*-value of which only values <0.1 are shown; RATE, number of CNVs per individual; PROP, proportion of individuals with at least one CNV; TOTKB, total length spanned by CNVs; AVGKB, average size spanned by CNVs; AO $\leq$ 21y, early age-at-onset patients; AO $>$ 21y, patients with an age-at-onset later than 21 years.

<sup>a</sup>Withstood Bonferroni correction for 72 tests,  $P_{\text{adjusted}}=0.029$ .



**Table 2** Association results for specific CNVs in BD at loci with prior evidence for association with psychiatric disorders.

Previous studies				Present study			
Locus	Type (n)	Associated disorder	Reference	CNV (frequency)			
				Type (n)	In 882 patients	In 872 controls	<i>P</i> (OR, 95%CI)
1q21	Microdel	ASD, MR, SCZ	8,43	Microdel (1)	0 CNV (0%)	1 Microdel (0.11%)	0.49 (0.99 [0.01-77.56])
2p16	Microdel	ASD, SCZ	44,45	Microdel (8)	5 Microdel (0.56%)	3 Microdel (0.34%)	0.73 (1.65 [0.32-10.67])
7q34-36	Microdel	SCZ	46	Micordel (15), Microdup (4)	11 Microdel (1.25%), 1 Microdup (0.11%)	4 Microdel (0.45%), 3 Microdup (0.34%)	0.36 (1.70 [0.62-5.13])
15q11	Microdel	SCZ	8	Microdel (4), Microdup (8)	1 Microdel (0.11%), 6 Microdup (0.56%)	3 Microdel (0.34%), 2 Microdup (0.57%)	0.77 (1.39 [0.38-5.56])
15q13	Microdel	ASD, EP, MR, SCZ	8,47,48	Microdel (35), Microdup (31)	15 Microdel (1.25%), 19 Microdup (1.70%)	20 Microdel (2.29%), 12 Microdup (1.38%)	0.90 (1.05 [0.62-1.78])
16p11	Microdel, Microdup	ASD	7	Microdel (0), Microdup (3)	0 Microdel (0%), 3 Microdup (0.23%)	0 CNV (0%)	0.25 (2.97 [0.24, 156.29])
17p12	Microdel	SCZ	49	Microdel (1), Microdup (8)	1 Microdel (0.11%), 4 Microdup (0.45%)	0 Microdel (0%), 4 Microdup (0.45%)	1 (1.24 [0.27-6.26])
22q11	Microdel	ASD, BD, MR, SCZ	9,50,51	Microdel (16), Microdup (36)	10 Microdel (1.13%), 18 Microdup (2.04%)	6 Microdel (0.69%), 18 Microdup (2.18%)	0.67 (1.16 [0.64-2.11])

Abbreviations: n, number of observed CNVs; Microdel, microdeletions of variable size; Microdup, microduplications of variables size, ASD, autism spectrum disorder; BD, bipolar disorder; EP, epilepsy; MR, mental retardation; SCZ, schizophrenia; *P*, *P*-value from Fisher's exact test; CI, confidence interval.

**Table 3** Functional categories<sup>a</sup> significantly enriched by several genes affected by singleton microdeletions in the AO≤21y subgroup.

Category	Genes in CNV	<i>P</i> -value <sup>b</sup>	Adjusted <i>P</i> -value <sup>c</sup>
Endocrine System Disorders	<i>ARHGAP26, CHRNA9, CNTN4, EEF1D, ERBB2IP, KLRG1, LINGO2, MAD1L1, MSRA, NAALADL2, NKAIN2, NLGN1, PDGFRL, PRKG2, SLCO1B1, SPATA16, UNC13C</i>	1.38 x 10 <sup>-4</sup> - 4.36 x 10 <sup>-3</sup>	2.09 x 10 <sup>-3</sup> - 1.99 x 10 <sup>-2</sup>
Metabolic Disease	<i>ARHGAP26, CHRNA9, CNTN4, EEF1D, ERBB2IP, KLRG1, LINGO2, MAD1L1, MSRA, NAALADL2, NKAIN2, NLGN1, PDGFRL, PRKG2, SLCO1B1, SPATA16, UNC13C</i>	1.38 x 10 <sup>-4</sup> - 4.36 x 10 <sup>-3</sup>	2.09 x 10 <sup>-3</sup> - 1.99 x 10 <sup>-2</sup>
Genetic Disorder	<i>ADAMTS19, ARHGAP26, BMP3, C4ORF22, CCDC102B, CHRNA9, CNTN4, ERBB2IP, GCNT2, LINGO2, MAD1L1, MAK, MBTPS1, MSRA, MTMR7, NAALADL2, NKAIN2, NLGN1, PDGFRL, PRKG2, RASGEF1B, RBM47, SLC7A2, SLCO1B1, SLCO1B3, SPATA16, TRAPPC9, UNC13C</i>	1.38 x 10 <sup>-4</sup> - 4.42 x 10 <sup>-2</sup>	2.09 x 10 <sup>-3</sup> - 9.02 x 10 <sup>-2</sup>
Immunological Disease	<i>C4ORF22, CCDC102B, CNTN4, EEF1D, ERBB2IP, KLRG1, LINGO2, MAD1L1, MSRA, NAALADL2, NKAIN2, NLGN1, SLCO1B1, UNC13C</i>	1.54 x 10 <sup>-4</sup> - 9.54 x 10 <sup>-3</sup>	2.09 x 10 <sup>-3</sup> - 3.02 x 10 <sup>-2</sup>
Infectious Disease	<i>CCDC102B, CNTN4, MAD1L1, NAALADL2, UNC13C</i>	1.54 x 10 <sup>-4</sup> <sup>d</sup>	2.09 x 10 <sup>-3</sup> <sup>d</sup>

<sup>a</sup>Ingenuity Pathway Analysis was performed using the "core analysis" function and default settings.

<sup>b</sup>Single test *P*-value range within high level function (Fisher's exact test).

<sup>c</sup>Multiple test-corrected *P*-value range using Benjamini-Hochberg correction.

<sup>d</sup>The genes were assigned to only one subcategory yielding only one *P*-value.

**Table 4** Functional categories<sup>a</sup> significantly enriched by several genes affected by microduplications in the AO<sub>≤</sub>21y subgroup.

Category	Genes in CNV	P-value <sup>a</sup>	Adjusted P-value <sup>b</sup>
Immunological Disease	<i>ACACB, ALKBH1, ATP10A, CD209, CDC16, CDH12, CDH13, CLEC4M, CSMD1, CYFIP1, DEPDC6, DGKB, DOC2A, FAT1, GPC5, GPR158, HCP5, KDM4C, LYN, MAD1L1, MAMDC2, MICA, MTHFD1L, NEGRI, NXN, NXP1, PARK2, PDE8A, PDE8B, RBP3, RSPO4, TAOK2</i>	1.01 x 10 <sup>-6</sup> - 2.65 x 10 <sup>-2</sup>	1.34 x 10 <sup>-3</sup> - 1.38 x 10 <sup>-1</sup>
Infectious Disease	<i>ACACB, CD209, CD36, CDH12, CHRNA7, CLEC4G, CLEC4M, CSMD1, CTSLI, CYFIP1, DEPDC6, DGKB, EGFR, F2, FAT1, GPC5, HCP5, KDM4C, LILRA3, LILRB5, LYN, MAD1L1, MPHOSPH6, NXN, PARK2, PDE8A, RSPO4, SEC61G, SNW1, STXBP2, TOP3B, TRIM58, XAB2, ZNF254, ZNF354A</i>	1.01 x 10 <sup>-6</sup> - 2.65 x 10 <sup>-2</sup>	1.34 x 10 <sup>-3</sup> - 1.38 x 10 <sup>-1</sup>
Endocrine System Disorders	<i>ABCA4, ACACB, ALKBH1, AMBRA1, APBA1, ATP10A, BTBD3, C14ORF166B, CD36, CDC16, CDH12, CDH13, CHRNA7, CLK4, COL23A1, COLEC12, CSMD1, CTSLI, CYP2E1, DAPK1, DEPDC6, DGKB, DOC2A, FAT1, GPC5, GPR115, GPR158, HCP5, IMMP2L, KCNT2, KDM4C, KIF25, MAD1L1, MAMDC2, MICA, MLLT4, MTHFD1L, NEGRI, NXP1, PARK2, PDE8A, PDE8B, PTGER3, PTPRT, RCAN1, RETN, RSPO4, SYT10, TAF2, TAOK2, TNR, ZNF350, ZNF613, ZNF615, ZNF649</i>	1.06 x 10 <sup>-5</sup> - 2.65 x 10 <sup>-2</sup>	7.04 x 10 <sup>-3</sup> - 1.38 x 10 <sup>-1</sup>
Genetic Disorder	<i>ABCA4, ABCB10, ALDOA, ALG10, AMBRA1, APBA1, ATXN7L1, BTBD3, C14ORF156, C14ORF166B, C3ORF20, CD209, CD36, CDH12, CDH13, CFH, CHRNA7, CLK4, COL23A1, COLEC12, CRYZ, CSMD1, CTH, CTSLI, CUL2, CYFIP1, CYP2E1, DAPK1, DEPDC6, DGKB, DOCK4, DOCK8, DSCC1, E2F1, EGFR, F2, FAT1, FOLH1, GDAP1, GLDC, GPC5, GPR115, GPR158, GPRC5C, GRIP2, HYDIN, IMMP2L, JPH1, KCNE1, KCNE2, KCTD13, KDM4C, KIF25, LILRA3, LILRB5, LYN, MAD1L1, MAMDC2, MLLT4, MPHOSPH6, MTHFD1L, MTNR1A, NEGRI, NIPAI, NLRP12, NXN, NXP1, PARK2, PDE8A, PDE8B, PDPN, PLEKHG1, PPP1R9B, PRKCG, PSMF1, PTGER3, PTPRT, RBP3, RCAN1, RETN, RNASE1, RSPO4, SEC61G, SLC22A23, SNX25, SPTLC3, SYT10, TAF2, TAOK2, THOC1, TNR, TSPO, TTYH2, TUBGCP5, UPF3A, WDR41, YWHAE, ZNF254, ZNF331, ZNF350, ZNF557, ZNF613, ZNF614, ZNF615, ZNF649, ZNF828</i>	1.06 x 10 <sup>-5</sup> - 3.94 x 10 <sup>-2</sup>	7.04 x 10 <sup>-3</sup> - 1.59 x 10 <sup>-1</sup>

Metabolic Disease	<i>ABCA4, ACACB, ALKBH1, AMBRA1, APBA1, ATP10A, BTBD3, C14ORF166B, CD36, CDC16, CDH12, CDH13, CHRNA7, CLK4, COL23A1, COLEC12, CSMD1, CTH, CTS1, CYP2E1, DAPK1, DEPDC6, DGKB, DOC2A, F2, FAT1, GLDC, GPC5, GPR115, GPR158, HCP5, IMM2L, KCNT2, KDM4C, KIF25, MAD1L1, MAMDC2, MAZ, MICA, MLLT4, MTHFD1L, NEG1, NXP1, PARK2, PDE8A, PDE8B, PTGER3, PTPRT, RCAN1, RETN, RSPO4, SYT10, TAF2, TAOK2, TNF, ZNF350, ZNF613, ZNF615, ZNF649</i>	1.06 x 10 <sup>-5</sup> - 3.94 x 10 <sup>-2</sup>	7.04 x 10 <sup>-3</sup> - 1.59 x 10 <sup>-1</sup>
-------------------	--	---	---

<sup>a</sup> Ingenuity Pathway Analysis was performed using the "core analysis" function and default settings.

<sup>b</sup> Single test *P*-value range within high level function (Fisher's Exact Test).

<sup>c</sup> Multiple test-corrected *P*-value range using Benjamini-Hochberg correction.

## Figures:

**Figure 1** Common specific microduplications. **(a)** On 10q11, each microduplication had the same putative 5' and 3' breakpoints in patients and controls, resulting in a length of around 160 kb. They were observed in 53 of 882 BD patients (blue bars), including 28 from the  $AO \leq 21y$  subgroup. A total of 32 of 872 controls also carried this duplication (green bars). This CNV completely covers the gene *ANTXR1*. **(b)** On 6q27, the microduplications were characterized by six different possible breakpoints in early-onset BD patients ( $AO \leq 21y$ , blue bars) and controls (green bars), leading to a shared sequence of nearly 250 kb (grey bars). These duplications were observed in 17 of 291  $AO \leq 21y$  patients and in 22 of 872 controls. The 6q27 CNVs overlapped with the 3' terminus of *MLLT4* as well as with the complete sequences of *KIF25* and *FRMD1*. *DACT2*, a WNT signaling pathway gene, is located around 115 kb downstream of this CNV. CNVs in patients and controls were sorted according to their length. Schematic drawings were generated using UCSC Genome browser (NCBI Build 36.1).

