

De novo CNVs in bipolar affective disorder and schizophrenia

Lyudmila Georgieva¹, Elliott Rees¹, Jennifer L. Moran², Kimberly D. Chambert², Vihra Milanova³, Nicholas Craddock¹, Shaun Purcell^{4,5}, Pamela Sklar^{4,5}, Steven McCarroll², Peter Holmans¹, Michael C. O'Donovan¹, Michael J. Owen¹ and George Kirov^{1,*}

¹Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK, ²Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA, ³Department of Psychiatry, Medical University, Sofia, Bulgaria, ⁴Division of Psychiatric Genomics, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY, USA and ⁵Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

Received May 19, 2014; Revised and Accepted July 17, 2014

An increased rate of *de novo* copy number variants (CNVs) has been found in schizophrenia (SZ), autism and developmental delay. An increased rate has also been reported in bipolar affective disorder (BD). Here, in a larger BD sample, we aimed to replicate these findings and compare *de novo* CNVs between SZ and BD. We used Illumina microarrays to genotype 368 BD probands, 76 SZ probands and all their parents. Copy number variants were called by PennCNV and filtered for frequency (<1%) and size (>10 kb). Putative *de novo* CNVs were validated with the z-score algorithm, manual inspection of log R ratios (LRR) and qPCR probes. We found 15 *de novo* CNVs in BD (4.1% rate) and 6 in SZ (7.9% rate). Combining results with previous studies and using a cut-off of >100 kb, the rate of *de novo* CNVs in BD was intermediate between controls and SZ: 1.5% in controls, 2.2% in BD and 4.3% in SZ. Only the differences between SZ and BD and SZ and controls were significant. The median size of *de novo* CNVs in BD (448 kb) was also intermediate between SZ (613 kb) and controls (338 kb), but only the comparison between SZ and controls was significant. Only one *de novo* CNV in BD was in a confirmed SZ locus (16p11.2). Sporadic or early onset cases were not more likely to have *de novo* CNVs. We conclude that *de novo* CNVs play a smaller role in BD compared with SZ. Patients with a positive family history can also harbour *de novo* mutations.

INTRODUCTION

Bipolar affective disorder (BD) has a life-time risk of ~1% in the general population and a 10-fold increased risk in first-degree relatives (1). The heritability estimates range between 59 and 87% (2–4). It is a complex genetic disorder, with a high degree of genetic and phenotypic heterogeneity (5). Genome-wide association studies based on common genetic variants have identified a number of loci at compelling levels of statistical support (6–10). It has been estimated that about a third of the genetic variance in risk is contributed by common alleles that are tagged by SNPs on genotyping arrays (11).

Rare, moderate to highly penetrant copy number variants (CNVs) have been clearly established as risk factors for several neuropsychiatric disorders: schizophrenia (SZ), autism spectrum disorder and intellectual disability/developmental delay (ID/DD) (12–16). Therefore, CNVs may also account for some of the unexplained heritability of BD. Studies on BD have yielded conflicting results, with modest enrichments for CNVs in some studies but not in others (17–22). An enrichment of *de novo* CNVs in individuals with BD was first reported by Malhotra *et al.* (21) This team identified 10 CNVs in 185 probands, a rate of 5.4% (or 4.3% rate per person, as 2 probands had 2 CNVs each), compared with 4 CNVs in 426 controls

*To whom correspondence should be addressed at: Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff CF24 4HQ, UK. Tel: +44 2920688465; Fax: +44 2920687068; Email: kirov@cardiff.ac.uk

(a rate of 0.9%) suggesting that this class of variants is involved in the disorder, particularly in early-onset cases. More recently, Noor *et al.* (23) found 8 *de novo* CNVs among 215 BD probands, a rate of 3.7%. No new control group was tested in that study, but the rate was considered increased compared with control rates from previous studies of 1–2%.

We aimed to investigate the role of *de novo* CNVs in the aetiology of BD in the largest sample of BD parent-proband trios tested to date and compare them with *de novo* CNVs found in SZ patients.

RESULTS

We genotyped 368 BD and 76 SZ probands and all their parents [after quality control (QC) filtering]. We successfully validated 21 *de novo* CNVs: 13 deletions and 8 duplications (Table 1). Of those, 15 were found in BD probands, a rate of 4.1% (2.4% for CNVs >100 kb) and 6 in SZ probands, a rate of 7.9% (6.4% for CNVs >100 kb). These differences were not significant, but the current sample sizes were clearly too small. One BD proband had two *de novo* CNVs, so the *de novo* can also be expressed as affecting 14 of 368 (3.8%) probands. We then analysed the current data together with those from four previous *de novo* CNV studies in BD and/or SZ (Table 2). In order to be conservative in the analysis, we restricted the data to large CNVs (>100 kb), as smaller ones might not have been called with some of the arrays used in the previous studies. Using the >100-kb cut-off, the *de novo* rates were 1.5, 2.2 and 4.3% in controls, BD and SZ, respectively (Fig. 1 and Table 2). The differences between SZ and BD and between SZ and controls were significant ($P = 0.015$ and 4.3×10^{-7} , respectively); however, the increase in the *de novo* rate in BD

over controls was not significant ($P = 0.21$, Table 2). For completion, we also present the results for CNVs of >10 kb (Table 2 and Fig. 1) but consider these comparisons less reliable, owing to the stronger potential bias caused by different array coverage. Using the >10-kb cut-off, the overall rate of *de novo* CNVs in BD was higher than the rate in controls (4.3 versus 2.0%, $P = 0.00065$) and lower than in SZ (4.3 versus 5.9%, $P = 0.14$).

The median size of *de novo* CNVs was 189 kb in BD and 640 kb in SZ, but this difference was not significant (Mann–Whitney U -test, $P = 0.12$). When results were combined with those from previous studies, and a threshold of >100 kb was applied as mentioned above, CNVs in BD were intermediate in size between controls and SZ with medians of 338, 448 and 613 kb in controls, BD and SZ, respectively. Only the SZ/controls difference was significant ($P = 0.001$). The trend for SZ patients to have larger *de novo* CNVs is also clear from the Kaplan–Meier survival graph in Supplementary Material, Figure S1.

Unexpectedly, the rate of *de novo* CNVs was higher in familial cases compared with sporadic ones, both in BD (10.0 versus 3.1%, $P = 0.039$) and in SZ (10.8 versus 5.1%, $P = 0.42$). When we combined our results with the previous ones, the rates of *de novo* CNVs were not significantly different between familial and sporadic cases: for SZ: 5.2 versus 5.9%, $P = 0.87$; and for BD: 6.4 versus 3.3%, $P = 0.16$ (Supplementary Material, Table S2). There was a non-significant trend for parents of *de novo* BD carriers to have been older at the time of birth of their offspring: average paternal age = 31.3 years (SD = 7.4) versus 28.8 (SD = 5.5), $P = 0.1$, and average maternal age = 27.8 (SD = 6.6) versus 25.6 (SD = 5.2), $P = 0.13$. A small trend in the same direction has also been reported in a larger CNV study on probands with intellectual disability (26).

Table 1. *De novo* CNVs detected in this study

Diagnosis	Cytoband	Chr	Start	End	Sample_ID	CNV type	CNV size (bp)	Family_type	FH	Genes
BD	2q31.1	2	175365518	175813657	6038-3	del	448 139	UK BD Sibling	Sibling	<i>WIPF1, CHRNA1, CHN1</i>
BD	5q15	5	92414916	95724985	1702-1	dupl	3 310 069	BG BD Trio	No	18 genes
BD	6p22.3	6	23671880	23724469	6008-1	dupl	52 589	UK BD Trio	No	0
BD	9p21.3	9	23765046	23953634	6194-1	del	188 588	UK BD Trio	No	<i>ELAVL2</i>
BD	10p14	10	6968905	6985007	1421-1	dupl	16 102	BG BD Trio	Father	0
BD	10q21.1	10	56517409	56677171	602-1	del	159 762	UK BD Trio	No	<i>PCDH15</i>
BD	11q14.1	11	84139937	84345829	6320-1	del	205 892	UK BD Trio	No	<i>DLG2</i>
BD	16p13.3	16	6891681	6922307	6198-1	del	30 626	UK BD Trio	No	<i>RBFOX1</i> , intronic
BD	16p11.2 distal	16	28825605	29043450	3128-1	dupl	217 845	BG BD Trio	No	12 genes
BD	16p11.2	16	29595483	30198151	6023-9	dupl	602 668	UK BD Trio	Father	31 genes
BD	17q23.1	17	57696973	57779678	1002-1	del	82 705	BG BD Trio	No	<i>CLTC, PTRH2</i>
BD	18q12.1	18	28277082	28375949	6194-1	dupl	98 867	UK BD Trio	No	0
BD	19q12-q13.12	19	30861683	36685690	2283-1	dupl	5 824 007	BG BD Trio	No	99 genes
BD	20p12.1	20	14771194	14853050	4052-1	del	81 856	BG BD Sibling	sibling	<i>MACROD2</i> , exonic
BD	22q11.21	22	21069073	21608479	5003-1	del	539 406	BG BD Trio	Mother	14 genes
SZ	1q41	1	222152402	223209450	UK526-8	del	1 057 048	UK SZ Sibling	Sibling	6 genes
SZ	2p16.3	2	50865334	51492973	4097-6	del	627 639	BG mixed	Sibling	<i>NRXN1</i> , exonic
SZ	15q24.3	15	77282884	77340328	UK516-4	dupl	57 444	UK SZ Sibling	Sibling	<i>PSTPIP1, TSPAN3</i>
SZ	22q11.21	22	18886915	21463730	UK1238-4	del	2 576 815	UK SZ Trio	No	VCFS region
SZ	22q11.22	22	22698552	23070912	UK602-4	del	372 360	UK SZ Sibling	No	5 genes
SZ	22q11.22 distal	22	22998337	23651318	UK662-4	del	652 981	UK SZ Sibling	Sibling	5 genes

All 21 *de novo* CNVs discovered and validated are sorted by diagnosis and genomic location. Cytoband, chromosome, start and end are listed according to UCSC build 37, hg19. BG, Bulgaria, UK, United Kingdom.

Table 2. Comparison of rates and sizes of CNVs with previous *de novo* CNV studies

	<i>N</i> trios	<i>N</i> CNVs (%) > 10 kb	Median size in kb > 10 kb	<i>N</i> CNVs (%) > 100 kb	Median size in kb > 100 kb
Current study					
BD	368	15 (4.1%)	189	9 (2.4%)	448
SZ	76	6 (7.9%)	640	5 (6.4%)	653
Malhotra <i>et al.</i> (21)					
BD	185	10 (5.4%)	137	5 (2.7%)	611
SZ	177	9 (5.1%)	348	6 (3.4%)	824
CON	426	4 (0.9%)	41	1 (0.2%)	1425
Kirov <i>et al.</i> (24)					
SZ	662	34 (5.1%)	321	25 (3.8%)	574
CON	2623	59 (2.2%)	259	46 (1.8%)	320
Xu <i>et al.</i> (25)					
SZ	200	17 (8.5%)	260	12 (7.9%)	489
CON	159	2 (1.3%)	2804	2 (1.3%)	2804
Noor <i>et al.</i> (23)					
BD	215	8 (3.7%)	76	3 (1.4%)	418
Totals/ <i>P</i> -value					
BD	768	33 (4.3%)	133	17 (2.2%)	448
SZ	1115	66 (5.9%)	356	48 (4.3%)	613
CON	3208	65 (2.0%)	259	49 (1.5%)	338
<i>P</i> -value, BD versus CON		0.00065	0.14	0.21	0.74
<i>P</i> -value, SZ versus CON		1.4×10^{-9}	0.54	4.3×10^{-7}	0.001
<i>P</i> -value, BD versus SZ		0.14	0.079	0.015	0.44

All rates refer to the number of CNVs in the sample (rather than the number of carriers of CNVs). CNVs on the X-chromosome are excluded. Significant results are shown in bold.

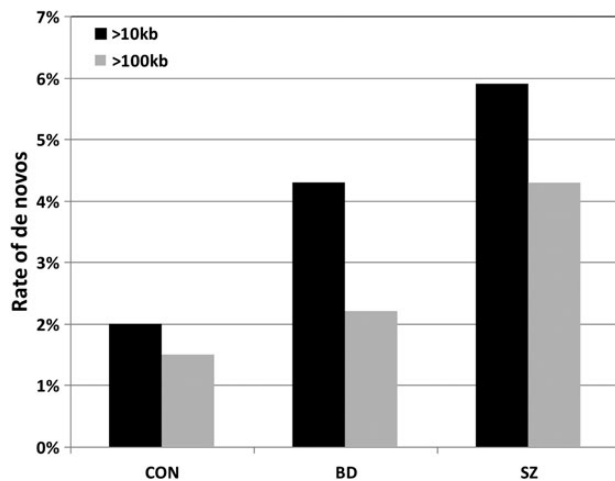


Figure 1. Comparison of the *de novo* rates for CNVs >10 and >100 kb in controls (CON), BD and SZ, based on the studies listed in Table 2.

To provide a further comparison of CNVs between BD and SZ, we assessed how many CNVs (transmitted or *de novo*) were found in 15 CNV regions previously implicated in SZ (12) in the BD and SZ probands in the current study (i.e. in samples not used in the discovery of these associations). The results are presented in Table 3. The overall rate of these CNVs in BD is significantly lower (1.35%) than that in the SZ sample (9%) (Fisher Exact test, $P = 0.0007$), and on six occasions, they were not transmitted to BD probands from carrier parents. In contrast, there were no non-transmitted CNVs from this list in the SZ sample. No person had two CNVs from this list.

One previous study on BD found an increased rate of singleton deletions in subjects with an onset of illness before the age of 18

years (17). Another one (21) found an increased rate of *de novo* CNVs in early-onset BD cases. In the present study, the mean age at onset among BD probands with and without *de novos* was practically identical (22.6 versus 22.8 years), with a similar distribution of ages (Fig. 2).

Gene pathway analyses in the combined datasets did not reveal an enrichment of BD *de novo* CNV hits relative to control *de novo* CNVs after controlling for multiple testing (Supplementary Material).

DISCUSSION

We have conducted the largest analysis of *de novo* CNVs in BD to date (Table 2). We analysed BD and SZ families together, in order to compare the *de novo* rates with the same methods and arrays.

Frequency of *de novo* CNVs in BD

De novo CNVs were found in 4.1% of BD probands. This rate is increased compared with controls from previous studies, but lower than the 7.9% in SZ in the current sample. To obtain a more meaningful comparison, we included in our analyses data from previous large studies on *de novo* CNVs in SZ, BD and controls. In order to minimise possible bias caused by different array resolutions used in the different studies, we analysed these differences for CNVs > 100 kb (as these are more likely to be detected by all arrays). Table 2 and Figure 1 show the rates in these phenotypes in the combined data. The rate in BD probands was intermediate between those in controls and SZ (1.5 versus 2.2 versus 4.3%), although the rates in BD were not significantly different from controls. Despite the weak statistical support, both comparisons (with cut-offs of 10 and 100 kb) show

Table 3. Transmission status of CNVs at loci implicated in SZ [according to our review of the literature (12)]

Locus	Position in Mb	BD (N = 371)	SZ (N = 78)
1q21.1 del	chr1: 14 657–14 739	1 T from BD F	
1q21.1 dup	chr1: 14 657–14 739	1 T from healthy M	
<i>NRXN1</i> del	chr2: 5015–5126		1 <i>de novo</i>
3q29 del	chr3: 19 573–19 734		
WBS dup	chr7: 7274–7414		
<i>VIPR2</i> dup	chr7: 15 882–15 894		
15q11.2 del	chr15: 2280–2309	3 NT from healthy parents	5 T (one M is SZ)
Angelman/Prader–Willi dup	chr15: 2482–2843		
15q13.3 del	chr15: 3113–3248		
16p13.11 dup	chr16: 1551–1630	1 T 2 NT from healthy parents	
16p11.2 distal del	chr16: 2882–2905		
16p11.2 dup	chr16: 2964–3020	1 <i>de novo</i> 1 NT from healthy M	
17p12 del	chr17: 1416–1543	1 T from healthy F	
17q12 del	chr17: 3481–3620		
22q11.2 del	chr22: 1902–2026		1 <i>de novo</i>
Total in probands		5 (1.35%); 6 NT	7 (9%); 0 NT

The ‘Total in probands’ include transmissions + *de novos*.
T/NT, transmitted/not transmitted from a parent; M, mother; F, father.

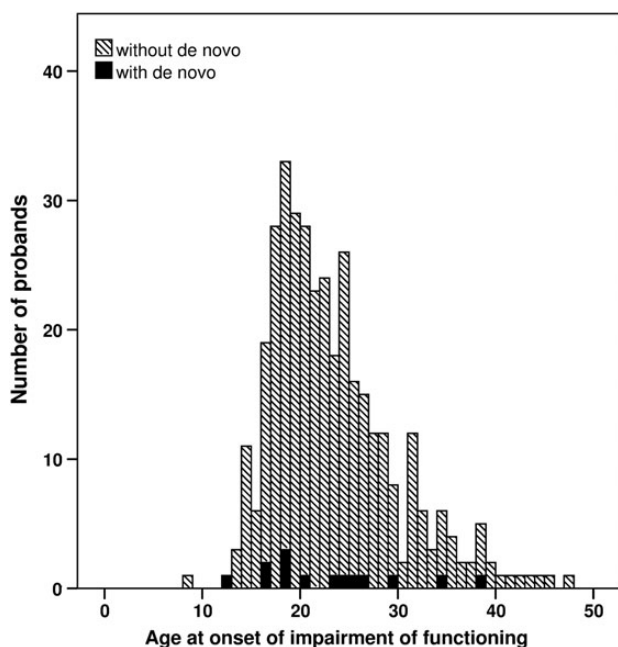


Figure 2. Age at onset among BD probands with *de novo* CNVs (black columns) and without *de novos* (dashed columns).

similar trends, with the rates of *de novo* CNVs in BD being intermediate between controls and SZ (Fig. 1).

Potential role of specific CNVs in BD

The trend we observed for an increased rate of *de novo* CNVs in BD compared with controls is consistent with a small proportion of these loci playing a role in the pathogenesis of the disorder. The more likely candidate loci are the following: the

deletion at *DLG2*, a gene implicated in SZ (24) and BD (23); the duplication at 16p11.2, as it is also implicated in SZ and BD (13,21) [our proband with *de novo* duplication was among the cases used in the original case–control study that found an association with BD (13)]; the duplication of the ‘distal 16p11.2’ locus, as it is an ID locus, while the reciprocal deletion is both an SZ and ID locus (15,27); the deletion at *PCDH15*, as mutations in this gene can cause deafness and Usher syndrome Type IF (<http://omim.org/entry/602083>), a disorder with a possibly increased rate of psychosis and behavioural problems (28,29).

CNVs in BD might be less pathogenic than those found in SZ

Overall *de novo* CNVs tended to be smaller in BD (median of 448 kb) than in SZ (median of 613 kb) in the combined datasets (Table 2). The lack of significance might be due to the small sample sizes, as the distribution of CNV sizes suggests a trend for CNVs in SZ to be larger (Supplementary Material, Fig. S1). Previous case–control studies in BD also report that the rate of very large (>1 Mb) and rare (<1%) CNVs in BD is similar or even lower than that in controls (18,22). Only two deletions and six duplications in BD probands in the current study were >1 Mb (rates of 0.54 and 1.62%, respectively, including transmitted CNVs). These rates are lower than those in previous controls analysed by us with the same methods: among 11 255 controls in our recent study, we reported rates of 0.65 and 1.95%, respectively (30). A smaller proportion of CNVs in BD probands were also found at 15 loci that have been shown to be pathogenic for SZ and other neurodevelopmental disorders, either as *de novo*, or inherited (Table 3). The cumulative rate of 1.35% of these CNVs in BD patients is close to the 0.96% rate we reported among 11 255 controls and lower than the 2.49% among 6882 SZ patients in our recent study (30). The strongest difference between the two disorders is for 15q11.2 deletions, which were not transmitted from three unaffected

parents to their BD offspring, whereas such deletions were transmitted from three parents (one affected with SZ) to five SZ offspring (including two affected SZ sib-pairs). In our previous study on BD, we also found a particularly low rate of 15q11.2 deletions among 1697 cases (0.18%) (18), which is even lower than the 0.28% in population controls (12). All these observations suggest that very large and rare CNVs, and those shown to increase risk for SZ, ID and DD, play at best a very modest role in BD.

Cases with a positive family history also have an increased rate of *de novo* CNVs

It has generally been assumed that sporadic cases are more likely to carry *de novo* CNVs than familial cases. To test whether *de novo* CNVs are more common in sporadic cases, we stratified the sample by history of BD/SZ/psychotic disorder in first-degree relatives and found that the *de novo* CNV rate is not significantly different between familial and sporadic cases; in fact, it was higher in familial cases in the current study (Supplementary Material, Table S2). The first study of *de novo* CNVs in SZ reported that they occur more frequently in sporadic cases (25), but a subsequent study (21) found that the rates of *de novo* CNVs in BD and SZ cohorts were similar in sporadic and familial cases. In our previous study on SZ (24), we considered the family history as positive only if it was present in parents (reasoning that siblings can have independent *de novo* CNVs) and found a slightly higher rate in sporadic cases. Combining all these studies and re-coding our previous data (24) to include cases with affected siblings as well, we find that the rate of *de novo* CNVs is similar in familial and sporadic cases: 5.2 versus 5.9% for SZ and 6.4 versus 3.3% for BD, respectively (the increased rate in familial BD cases is not significant, $P = 0.16$).

Two examples of *de novo* CNVs in family history-positive cases are particularly striking. The 16p11.2 duplication has a high penetrance of 34% for any neurodevelopmental disorder (31) but was found in the daughter of a father who also suffers with a severe form of BD (being *de novo*, that mutation is not found in the father). The exonic *NRXN1* deletion [penetrance of 32% for any disorder (31)] was found in a SZ proband from a multiply affected family: the proband's sister had schizoaffective disorder, she was married to a BD patient and her daughter had BD. No other family member had a pathogenic CNV.

In conclusion, this study confirms previous suggestions that very large and rare CNVs, especially those implicated in neurodevelopmental disorders (such as most of the CNVs listed in Table 3), play a lesser role in BD compared with SZ. However, we did observe a non-significant trend for the rate of *de novo* CNVs to be higher in BD than in controls, suggesting that larger and more powerful studies might reveal a significant excess. In addition, several of the loci impacted by *de novo* CNVs have been previously implicated in neuropsychiatric disorders, which enhance their credibility as candidates for BD. These include 16p11.2, *DLG2* and *PCDH15*. Finally, we also observed an excess of *de novo* mutations in familial cases of major psychiatric disorders. With hindsight this should not be surprising, as disorders of complex genetic inheritance are not due to single gene defects, but to an accumulation of a number of susceptibility factors.

MATERIALS AND METHODS

Participants

The total sample (after QC) filtering consists of 449 probands: 368 with BD (256 from Bulgaria and 112 from the UK) and 76 with SZ (15 from Bulgaria and 61 from the UK). Three hundred and eighty-one probands were from parent-offspring trios (342 with BD and 39 with SZ), 42 were from families with 2 affected siblings (16 with BD and 26 with SZ) and 21 (10 BD and 11 SZ) were from families with more complex structures, including families with a mixture of diagnoses (Supplementary Material, Table S1). Probands affected with schizoaffective disorder were excluded from this study. Probands with a history of psychosis in a sibling or parent (50 with BD and 37 with SZ) were included, as none of the risk CNVs identified to date is sufficiently penetrant to fully explain the disorder in carriers (31), and therefore, we wanted to test whether familial cases can also have *de novo* CNVs. The proportion of affected sibling pairs with SZ from the UK is very high because part of this cohort was recruited as affected sib-pairs for linkage analysis, whereas all BD trios were recruited specifically for studying parent-offspring trios.

The recruitment of families in Bulgaria has been described before (24). Each proband had a history of hospitalisation and was interviewed with an abbreviated version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (32). Consensus best-estimate diagnoses were made according to DSM-IV criteria by two researchers. This recruitment also included SZ trios, which have been genotyped with Affymetrix arrays and reported previously (24), apart from some families with different diagnoses that are reported here. In the UK, the BD patients were recruited and interviewed in person by GK, using the same rating instruments. Consensus best-estimate diagnoses were made by two researchers (G.K. and N.C.), based on the interview and hospital notes. The SZ families from the UK were recruited as part of sib-pair and case-control collections. The main purpose for the inclusion of SZ probands in the current study is to compare in an unbiased way (using identical methods), the *de novo* CNV rate between BD and SZ, and also to enlarge the sample of family history-positive cases, where fewer data are available from previous studies. Ethics committee approval for the study was obtained from the relevant research ethics committees and all individuals provided written informed consent for participation. A small proportion of the probands from the UK have been included as cases in previous case-control studies: 55 BD probands are in the Grozeva *et al.* study (18) and 29 SZ probands in the Kirov *et al.* study (33); however, they were not evaluated for *de novo* CNVs. Comparisons with *de novo* CNVs from healthy control populations were made with probands from three previous studies (21,24,25).

Genotyping of blood-derived DNA from all samples was performed at the Stanley Centre for Psychiatric Research at the Broad Institute of MIT, USA on two arrays: HumanOmni Express-12v1 (referred further for short as 'OmniExpress array'), containing 730 525 probes, and any poorly performing samples were re-genotyped on HumanOmniExpressExome-8v1 ('Combo array'), containing 951 117 probes. The Combo array contains SNPs from both the Omni Express array and the Illumina HumanExome-12v1_A ('Exome array'); however, for the analysis of Combo array data, we only used the probes present on the OmniExpress array ($N = 699\ 865$).

CNV calling and QC

Raw intensity data were processed using Illumina Genome Studio software (v2011.1). SNPs were clustered using the current samples, and LRR and B-allele frequencies were generated for CNV detection. PennCNV (34) was used to call CNVs following the standard protocol and adjusting for GC content. Sample-level QC was performed using the QC metrics generated by PennCNV. These include: LRR standard deviation, B-allele frequency drift, wave factor and total number of CNVs called per person. Samples were excluded if for any one of these metrics they constituted an outlier in their source dataset (details not presented). All poorly performing samples were re-genotyped on Combo arrays. If one family member making up a trio was excluded, then we excluded the whole trio, thus excluding 33 families.

All individual CNVs also went through QC filtering. First, raw CNVs in the same sample were joined together if the distance separating them was <50% of their combined length. CNVs were then excluded if they were either <10 kb, covered by <10 probes, overlapped with low copy repeats by >50% of their length (using PLINK) (35) or had a probe density of >20 kb per probe. The remaining CNVs from each dataset were then analysed together and CNV loci with a frequency of >1% were excluded using PLINK. The putative *de novo* CNVs were validated with the median z-score outlier method (24), software freely available at http://x004.psychm.uwcm.ac.uk/~dobril/z_scores_cnvs. The z-score histograms of CNVs with marginal z-scores were manually inspected. For all putative *de novo* CNVs, the LRR and B-allele frequencies were also visually inspected using Illumina Genome Studio v2011.1 software. Validation of all remaining putative *de novo* CNVs was performed using real-time PCR based on SYBR-Green I fluorescence with at least three primer sets per CNV. All samples were amplified using Sensimix kit (Bioline, UK), and data were analysed using Rotor-Gene Q series software. Each primer set was compared with a primer set outside the CNV which served as 'control' and data were normalized using delta Ct (cycle threshold) values.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the participants and clinicians who took part in the study.

Conflict of Interest statement. None declared.

FUNDING

The work at Cardiff University was funded by Medical Research Council (MRC) Centre (G0800509) and Program Grants (G0801418) and an MRC PhD Studentship to E.R. Funding for the recruitment of trios in Bulgaria was provided by the Janssen Research Foundation in 1999–2004, Ref: 045856. Funding for the recruitment of BD samples in the UK was provided by the Wellcome Trust as a Training Fellowship to G.K.

in 1996–1999, Ref: 045856. The samples were genotyped at the Broad Institute, USA, funded by a philanthropic gift to the Stanley Center for Psychiatric Research. Funding to pay the Open Access publication charges for this article was provided by Research Councils UK.

REFERENCES

- Craddock, N. and Jones, I. (1999) Genetics of bipolar disorder. *J. Med. Genet.*, **36**, 585–594.
- Lichtenstein, P., Yip, B.H., Björk, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F. and Hultman, C.M. (2009) Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*, **373**, 234–239.
- Smoller, J.W. and Finn, C.T. (2003) Family, twin, and adoption studies of bipolar disorder. *Am. J. Med. Genet. C Semin. Med. Genet.*, **123C**, 48–58.
- McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R. and Cardno, A. (2003) The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch. Gen. Psychiatry*, **60**, 497–502.
- Potash, J.B., Toolan, J., Steele, J., Miller, E.B., Pearl, J., Zandi, P.P., Schulze, T.G., Kassem, L., Simpson, S.G., Lopez, V. *et al.* (2007) The bipolar disorder phenome database: a resource for genetic studies. *Am. J. Psychiatry*, **164**, 1229–1237.
- Ferreira, M.A., O'Donovan, M.C., Meng, Y.A., Jones, I.R., Ruderfer, D.M., Jones, L., Fan, J., Kirov, G., Perlis, R.H., Green, E.K. *et al.* (2008) Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat. Genet.*, **40**, 1056–1058.
- Cichon, S., Mühleisen, T.W., Degenhardt, F.A., Mattheisen, M., Miró, X., Strohmaier, J., Steffens, M., Meesters, C., Herms, S., Weingarten, M. *et al.* (2011) Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am. J. Hum. Genet.*, **88**, 372–381.
- Group, P.G.C.B.D.W. (2011) Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.*, **43**, 977–983.
- Green, E.K., Grozeva, D., Forty, L., Gordon-Smith, K., Russell, E., Farmer, A., Hamshere, M., Jones, I.R., Jones, L., McGuffin, P. *et al.* (2013) Association at SYNE1 in both bipolar disorder and recurrent major depression. *Mol. Psychiatry*, **18**, 614–617.
- Green, E.K., Hamshere, M., Forty, L., Gordon-Smith, K., Fraser, C., Russell, E., Grozeva, D., Kirov, G., Holmans, P., Moran, J.L. *et al.* (2013) Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. *Mol. Psychiatry*, **18**, 1302–1307.
- Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S. *et al.* (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.*, **45**, 984–994.
- Rees, E., Walters, J.T., Georgieva, L., Isles, A.R., Chambert, K.D., Richards, A.L., Mahoney-Davies, G., Legge, S.E., Moran, J.L., McCarroll, S.A. *et al.* (2014) Analysis of copy number variations at 15 schizophrenia-associated loci. *Br. J. Psychiatry*, **204**, 108–114.
- McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J., Yoon, S., Perkins, D.O., Dickel, D.E., Kusenda, M., Krastovshvsky, O. *et al.* (2009) Microduplications of 16p11.2 are associated with schizophrenia. *Nat. Genet.*, **41**, 1223–1227.
- Malhotra, D. and Sebat, J. (2012) CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*, **148**, 1223–1241.
- Girirajan, S., Rosenfeld, J.A., Coe, B.P., Parikh, S., Friedman, N., Goldstein, A., Filipink, R.A., McConnell, J.S., Angle, B., Meschino, W.S. *et al.* (2012) Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N. Engl. J. Med.*, **367**, 1321–1331.
- Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A. *et al.* (2011) Multiple recurrent *de novo* CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*, **70**, 863–885.
- Zhang, D., Cheng, L., Qian, Y., Alliey-Rodriguez, N., Kelsoe, J.R., Greenwood, T., Nievergelt, C., Barrett, T.B., McKinney, R., Schork, N. *et al.* (2009) Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol. Psychiatry*, **14**, 376–380.

18. 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. *et al.* (2010) Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch. Gen. Psychiatry*, **67**, 318–327.
19. McQuillin, A., Bass, N., Anjorin, A., Lawrence, J., Kandaswamy, R., Lydall, G., Moran, J., Sklar, P., Purcell, S. and Gurling, H. (2011) Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. *Eur. J. Hum. Genet.*, **19**, 588–592.
20. Priebe, L., Degenhardt, F.A., Herms, S., Haenisch, B., Mattheisen, M., Nieratschker, V., Weingarten, M., Witt, S., Breuer, R., Paul, T. *et al.* (2012) Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder. *Mol. Psychiatry*, **17**, 421–432.
21. Malhotra, D., McCarthy, S., Michaelson, J.J., Vacic, V., Burdick, K.E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E.S. *et al.* (2011) High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron*, **72**, 951–963.
22. Grozeva, D., Kirov, G., Conrad, D.F., Barnes, C.P., Hurles, M., Owen, M.J., O'Donovan, M.C. and Craddock, N. (2013) Reduced burden of very large and rare CNVs in bipolar affective disorder. *Bipolar Disord.*, **15**, 893–898.
23. Noor, A., Lionel, A.C., Cohen-Woods, S., Moghimi, N., Rucker, J., Fennell, A., Thiruvahindrapuram, B., Kaufman, L., Degagne, B., Wei, J. *et al.* (2014) Copy number variant study of bipolar disorder in Canadian and UK populations implicates synaptic genes. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **165B**, 303–313.
24. Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L. *et al.* (2012) De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry*, **17**, 142–153.
25. Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A. and Karayiorgou, M. (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat. Genet.*, **40**, 880–885.
26. Hehir-Kwa, J.Y., Rodríguez-Santiago, B., Vissers, L.E., de Leeuw, N., Pfundt, R., Buitelaar, J.K., Pérez-Jurado, L.A. and Veltman, J.A. (2011) De novo copy number variants associated with intellectual disability have a paternal origin and age bias. *J. Med. Genet.*, **48**, 776–778.
27. Guha, S., Rees, E., Darvasi, A., Ivanov, D., Ikeda, M., Bergen, S.E., Magnusson, P.K., Cormican, P., Morris, D., Gill, M. *et al.* (2013) Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA Psychiatry*, **70**, 253–260.
28. Domanico, D., Fragiotta, S., Trabucco, P., Nebbioso, M. and Vingolo, E.M. (2012) Genetic analysis for two Italian siblings with usher syndrome and schizophrenia. *Case Rep. Ophthalmol. Med.*, **2012**, 380863.
29. Rao, N.P., Danivas, V., Venkatasubramanian, G., Behere, R.V. and Gangadhar, B.N. (2010) Comorbid bipolar disorder and Usher syndrome. *Prim. Care Companion J. Clin. Psychiatry*, **12**.
30. Rees, E., Walters, J.T., Chambert, K.D., O'Dushlaine, C., Szatkiwicz, J., Richards, A.L., Georgieva, L., Mahoney-Davies, G., Legge, S.E., Moran, J.L. *et al.* (2014) CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1. *Hum. Mol. Genet.*, **23**, 1669–1676.
31. Kirov, G., Rees, E., Walters, J.T., Escott-Price, V., Georgieva, L., Richards, A.L., Chambert, K.D., Davies, G., Legge, S.E., Moran, J.L. *et al.* (2014) The penetrance of copy number variations for schizophrenia and developmental delay. *Biol. Psychiatry*, **75**, 378–385.
32. Wing, J.K., Babor, T., Brugha, T., Burke, J., Cooper, J.E., Giel, R., Jablenski, A., Regier, D. and Sartorius, N. (1990) SCAN. Schedules for clinical assessment in neuropsychiatry. *Arch. Gen. Psychiatry*, **47**, 589–593.
33. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K.K., Holmans, P., Craddock, N., Owen, M.J., O'Donovan, M.C., Consortium, I.S. *et al.* (2009) Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum. Mol. Genet.*, **18**, 1497–1503.
34. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H. and Bucan, M. (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.*, **17**, 1665–1674.
35. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.