

IMMEDIATE COMMUNICATION

Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder

SE Bergen^{1,2}, CT O'Dushlaine^{1,2,3}, S Ripke^{2,3}, PH Lee^{1,2,3}, DM Ruderfer^{1,2,3}, S Akterin⁴, JL Moran², KD Chambert², RE Handsaker^{2,5}, L Backlund⁶, U Ösby⁶, S McCarroll^{2,5}, M Landen^{4,7}, EM Scolnick², PKE Magnusson⁴, P Lichtenstein⁴, CM Hultman^{4,10}, SM Purcell^{1,2,3,8,10}, P Sklar^{1,2,8,10} and PF Sullivan^{9,10}

Schizophrenia (SCZ) and bipolar disorder (BD) are highly heritable psychiatric disorders with overlapping susceptibility loci and symptomatology. We conducted a genome-wide association study (GWAS) of these disorders in a large Swedish sample. We report a new and independent case-control analysis of 1507 SCZ cases, 836 BD cases and 2093 controls. No single-nucleotide polymorphisms (SNPs) achieved significance in these new samples; however, combining new and previously reported SCZ samples (2111 SCZ and 2535 controls) revealed a genome-wide significant association in the major histocompatibility complex (MHC) region (rs886424, $P = 4.54 \times 10^{-8}$). Imputation using multiple reference panels and meta-analysis with the Psychiatric Genomics Consortium SCZ results underscored the broad, significant association in the MHC region in the full SCZ sample. We evaluated the role of copy number variants (CNVs) in these subjects. As in prior reports, deletions were enriched in SCZ, but not BD cases compared with controls. Singleton deletions were more frequent in both case groups compared with controls (SCZ: $P = 0.003$, BD: $P = 0.013$), whereas the largest CNVs (>500 kb) were significantly enriched only in SCZ cases ($P = 0.0035$). Two CNVs with previously reported SCZ associations were also overrepresented in this SCZ sample: 16p11.2 duplications ($P = 0.0035$) and 22q11 deletions ($P = 0.03$). These results reinforce prior reports of significant MHC and CNV associations in SCZ, but not BD.

Molecular Psychiatry (2012) **17**, 880–886; doi:10.1038/mp.2012.73; published online 12 June 2012

Keywords: schizophrenia; bipolar disorder; genetic; genome-wide association; major histocompatibility complex; copy number variation

INTRODUCTION

Schizophrenia (SCZ) and bipolar disorder (BD) are severe and chronic psychiatric disorders.¹ A persistent debate in psychiatry has been the extent to which SCZ and BD are distinct.² Although the classical definitions of these illnesses are relatively discrete,³ there are multiple shared features: age of onset in late teens/early adulthood, prevalence (~1%), high heritability (59–64% in the Swedish population) and some symptomatology (e.g., psychosis, chronicity and cognitive deficits).^{4–9} Genetic overlap between these disorders was validated in a large Swedish study in which the shared genetic contribution was estimated to be >60%.⁹

Genome-wide association studies (GWAS) have discovered robust and replicable risk loci for SCZ. Genetic markers in the major histocompatibility complex (MHC) region and *TCF4* have demonstrated replicated associations,^{10–14} and the Schizophrenia Psychiatric Genomics Consortium (PGC-SCZ) has recently identified other significant loci at 1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32–q24.33.

GWAS of BD have identified significant associations to *DGKH*,¹⁵ *ANKK1*¹⁶ and *NCAN*,¹⁷ and the largest BD GWAS, to date, discovered

significant loci in the calcium channel gene *CACNA1C* and *ODZ4*.¹⁸ In addition, a small number of loci were jointly analyzed by the PGC in SCZ and BD. These analyses demonstrated strong contributions from both disorders to the genome-wide significant results in *CACNA1C*, *ANKK1* and the *ITIH3–ITIH4* region, supporting the idea of shared genetic risk.^{14,18}

In contrast to common genetic variants (single-nucleotide polymorphisms (SNPs)), which have been implicated in these disorders, several large, rare copy number variants (CNVs) have been associated with SCZ, although the associations are less notable for BD. Deletions at 1q21.1, 2p21–16.3, 3q29, 15q11.2, 15q13.3 and 22q11, and duplications at 7q36.3, 16p11.2 and 16p13.11 confer substantial risk for SCZ in a small percentage of cases, but the risk is not specific to SCZ. These CNVs have also been observed at elevated rates in autism, epilepsy and mental retardation.^{19–22} Duplications at 16p11.2 have demonstrated risk for SCZ, autism, and to some extent, BD.^{23–25}

Few GWAS have examined SCZ and BD concurrently, and this represents the largest combined study of these disorders in a genetically homogeneous population to date. In this unique

¹Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; ²Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA; ³Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; ⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; ⁵Department of Genetics, Harvard Medical School, Boston, MA, USA; ⁶Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; ⁷Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden; ⁸Division of Psychiatric Genomics, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY, USA and ⁹Departments of Genetics, Psychiatry and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. Correspondence: Dr SE Bergen, 7 Cambridge Center, 6055H, Broad Institute, Cambridge, MA 02142, USA.
E-mail: sbergen@gmail.com

¹⁰These authors contributed equally to this work.

Received 20 December 2011; revised 2 April 2012; accepted 23 April 2012; published online 12 June 2012

sample, we conducted association analyses of common SNPs and explored the relationship of CNVs to BD and SCZ. Additional analyses assessing the aggregation of associated loci in biological pathways, explorations of common copy number polymorphisms (CNPs) in SCZ, and comparison of our SCZ association results with the PGC SCZ results through meta-analysis were also performed.

SUBJECTS AND METHODS

Subjects

Schizophrenia. Comprehensive analyses of most SCZ cases ($N=1507$) and controls ($N=2093$) are newly reported here, and some SCZ cases ($N=560$) and controls ($N=400$) from this study have been reported previously.^{10,34} Additionally, 44 SCZ cases and 42 controls previously removed from analyses owing to Finnish ancestry were reintegrated into the current analyses. All subjects were born in Sweden, and SCZ cases were identified via the Swedish Hospital Discharge Register containing all individuals hospitalized in Sweden since 1973. The Hospital Discharge Register has high agreement with psychiatric diagnoses.^{26–28} Diagnoses were established by the attending physician and confirmed in a subset of subjects by medical record review. Cases must have had at least two hospitalizations with an SCZ diagnosis. Control subjects, also selected through registers, were group-matched by age, sex and county of residence, and must not have been hospitalized with a psychiatric diagnosis. All subjects were at least 18 years old and gave written informed consent to participate. The study was approved by the Ethical Committee at Karolinska Institutet.

Bipolar disorder. The previously unreported BD subjects ($N=836$) were only collected in the most recent recruitment wave through three channels. Some BD cases ($N=256$) were identified using the Swedish Hospital Discharge Register with at least two hospitalizations with a BD diagnosis and confirmatory diagnostic review in a subset of subjects.²⁹ Additional subjects were recruited from the Affective Center at St Goran Hospital in Stockholm, Sweden, following physician's referral for BD ($N=216$). The diagnostic instrument used was a Swedish adaptation of the Affective Disorder Evaluation,³⁰ which includes the affective module of the Structured Clinical Interview for DSM-IV.³¹ A further 571 BD cases were recruited from the Stockholm County catchment area. Diagnoses were made according to the DSM-IV criteria, and cases were not reported previously. The control subjects used were the same as for the SCZ analyses described above. All ascertainment procedures were approved by the Regional Ethical Committees in Sweden.

Genotyping and quality control

Blood samples were obtained and DNA extracted from whole blood using standard methods at the Karolinska Institutet. Samples were genotyped in four waves (Supplementary Table 1). Most subjects (92.6%) were genotyped with Affymetrix 6.0 chips (Affymetrix, Santa Clara, CA, USA) with the remainder (7.4%) genotyped with the Affymetrix 5.0 chip. All genotyping was conducted at the Broad Institute of Harvard and MIT, and genotypes and CNVs were called using the Birdsuite algorithm.³² The quality control exclusionary measures for subjects were: genotype call rates <95%; ancestry outliers via multidimensional scaling; a randomly selected member of any pair of subjects with high relatedness ($\hat{r}>0.20$); and suspected sample error or contamination. SNPs were excluded for marked departure from Hardy–Weinberg equilibrium ($P<1\times 10^{-6}$), low minor allele frequencies (<1%), and non-random genotyping failure, inferred from the flanking haplotype background using the PLINK 'mishap' test ($P<1\times 10^{-10}$). Plate-based associations of $P<1\times 10^{-6}$ were taken as evidence of non-random plate failure, based on a comparison of allele frequency of each plate to all others and were removed on a plate-by-plate basis.

Following quality control, the analysis data set consisted of 745 006 SNPs genotyped in 1507 SCZ cases, 836 BD cases and 2093 controls. GWAS results for these samples have not been previously reported. Additional SCZ cases ($N=604$) and controls ($N=442$) from earlier collection waves of

this study^{11,34} bring the total sample size to 2947 combined cases and 2535 controls.

Statistical methods

Imputation was conducted using BEAGLE.³³ First, as with the PGC-SCZ data, we imputed our genotypes against autosomal genotype data from HapMap3³⁴ and retained only SNP dosages imputed with very high confidence (Info >0.8). Second, we imputed MHC Class 1 and 2 SNP and amino acid variants.³⁵ Following standard definitions for human leukocyte antigen (HLA) allele sequences from the EMBL-EBI Immunogenetics HLA Database, we imputed against the 2767 unrelated founder individuals of European descent, with high-resolution MHC genotyping data for 263 HLA alleles, 3852 SNPs and 372 amino acid positions. The genomic window imputed was 29.3–33.9 MB on chromosome 6.

Association analyses

All association analyses were conducted using logistic regression in PLINK.³⁶ Multidimensional scaling was performed on the entire data set, and each collection wave was analyzed separately using the first four multidimensional scaling components as covariates, to control for population substructure. Results were then combined by meta-analysis in PLINK. Four main association analyses were conducted using the HapMap3 imputed data for: BD and SCZ separately (new subjects), SCZ (full sample), and all SCZ and BD cases combined (Supplementary Table 1). We used a genome-wide significance threshold of $P<5\times 10^{-8}$.³⁷ Imputed HLA alleles or amino acid dosages were tested for association using a logistic regression model with covariates to correct for population substructure and genotyping batch. For amino acid positions with >2 alleles, we used the omnibus test in the conditional haplotype analysis module in PLINK. Conditional analyses in the MHC region were performed by conditioning on the genotyped SNP with the lowest P -value in the full SCZ results using the condition function in PLINK, which includes the additive effect of the SNP in the association model.

Meta-analysis with PGC-SCZ

Some Swedish SCZ subjects ($N=558$) and controls ($N=396$) were among the 17 samples included in the PGC-SCZ report,¹⁴ and we combined the individual non-Swedish PGC sample results with all Swedish SCZ results using the meta-analysis function in PLINK. This yielded P -values and odds ratios (ORs) for ~1.1 million markers in the combined sample of 11 271 cases and 14 601 controls. This is the largest SCZ GWAS meta-analysis yet reported.

Pathway analysis

Analyses using INRICH software³⁸ and the set screen test implemented in PLINK³⁹ were conducted to assess the joint action of common SNPs within pathways. For these tests, genomic regions encompassing the top results as well as surrounding SNPs in linkage disequilibrium were compared against gene pathways. Full methods are in the Supplementary Material.

Polygenic scoring

We carried out polygenic scoring analyses^{10,14} using the PGC-SCZ stage 1 samples. After excluding Swedish samples, the training data set consisted of 9160 cases and 12 066 controls, and 82 437 SNPs. The testing data set consisted of 2111 SCZ cases or 836 BD cases, and 2535 controls, and quantitative scores were computed for each subject based on the pT (P -value threshold), the proportion of SNPs with P -values < pT in the training data set. For each SNP set defined by pT, we calculated the proportion of variance explained (R^2) by subtracting the Nagelkerke's R^2 attributable to ancestry covariates alone from the R^2 for polygenic scores plus covariates. Full methods are described in the Supplementary Material.

CNV quality control and analyses

Intensity data from both SNP and structural variation probes were used to identify CNVs based on a hidden Markov model.³² Only subjects who

passed SNP quality control filters were considered for CNV analyses. We excluded CNVs with >0.01 frequency, common CNV regions identified in the HapMap samples,⁴⁰ and any CNVs spanning large genomic gaps (e.g., centromeres). Analyses were restricted to autosomal CNVs >100 kb, and subjects were removed for having either CNVs totaling >10 Mb or >20 total CNVs. This resulted in 4147 CNVs in 1505 SCZ cases, 834 BD cases and 2087 controls remaining for analysis.

The number of events in cases and controls, the rate per subject, the proportion of cases/controls to have at least one event, the total distance spanned per person and the average event size per person were evaluated for each CNV burden test across a range of frequencies (1, 2–6, or all), and sizes (100–200 kb, 200–500 kb and >500 kb) for duplications and deletions, separately and combined in both case groups. Statistical tests were one-sided, presuming an increased CNV burden in cases, and significance was evaluated using 10 000 permutations. In addition to the number of CNVs, the proportion of genes intersecting these regions in cases versus controls was tested. Large CNVs with previously reported associations with SCZ or BD were specifically examined, and novel loci were explored. For specific CNVs, only deletions and duplications spanning the full CNV region were reported.

Copy number polymorphisms

Common CNP deletions 50–260 000 bp in length and present at $>1\%$ frequency were identified in the 1000 Genomes data.⁴¹ These CNPs were in linkage disequilibrium with 40 767 SNPs present in the HapMap3-imputed results for our full SCZ sample and were cross-referenced with these results. When multiple SNPs tagged a CNP, the SNP with the greatest R^2 and lowest P -value was retained.

RESULTS

SNP association results

We performed genome-wide association analyses of SCZ and BD, and both disorders considered together using a common set of controls in a large sample from Sweden. Our basic analyses were new independent SCZ samples, all SCZ samples (new and previously reported), BD samples (all new), and SCZ and BD combined. The genomic control, λ (the median observed χ^2 -statistic divided by expectation under the null), showed little evidence for inflation of the test statistics. For SCZ (new and full samples) and SCZ–BD combined $\lambda = 1.04$ and for BD, $\lambda = 1.05$ (Supplementary Figure 1). Regions of association with $P < 1 \times 10^{-5}$ for the four main analyses are reported in Table 1. The quantile–quantile and Manhattan plots for all analyses can be found in Supplementary Figures S1 and S2.

In the new and previously unreported sample of 1507 SCZ cases and 2093 controls, none of 1 105 620 imputed SNPs were genome-wide significant (Table 1). The most highly associated SNP was in an intron of *GALNT13* (rs12998068, $P = 1.50 \times 10^{-6}$; OR = 1.40). Other noteworthy results include rs11150863 ($P = 3.49 \times 10^{-6}$; OR = 0.78) in an intron of *RPTOR* and many SNPs in the MHC region (top SNP: rs886424, 8.80×10^{-6} ; OR = 0.71).

We next combined the new and previously reported SCZ samples ascertained as part of the same study (2111 SCZ cases and 2535 controls). One SNP in the MHC region attained genome-wide significance (rs886424, $P = 4.54 \times 10^{-8}$; OR = 0.68). This marker was directly genotyped, and conditioning on this SNP revealed no independent signals in the 10-Mb surrounding region with $P < 1 \times 10^{-4}$ (Supplementary Figure 3). Some evidence of heterogeneity in ORs for this region across collection waves was observed (detailed in Supplementary Text). After the MHC, the next strongest association signal was from a SNP in *RGS7* (rs984402, $P = 3.43 \times 10^{-7}$; OR = 0.79). Association testing of 836 BD cases and 2093 controls yielded no genome-wide significant findings. The SNP with the lowest P -value, rs17746001 ($P = 3.22 \times 10^{-7}$; OR = 2.03), was in a gene-poor region of chromosome 4.

We next tested for association using SCZ and BD as a combined phenotype in 2949 cases (2111 SCZ and 836 BD) and 2535 controls. The smallest P -value in the combined analysis was also detected for rs17746001 ($P = 2.83 \times 10^{-7}$; OR = 1.75) and it was strongly represented in the SCZ results ($P = 2.23 \times 10^{-5}$; OR = 1.64). This was closely followed by a large cluster of MHC associations (top SNP: rs2524005, $P = 4.95 \times 10^{-7}$; OR = 0.76).

MHC imputation results

The highly polymorphic MHC region presents challenges to standard imputation methods. An imputation approach specifically tailored for this region was used to refine the signal at this locus through prediction of classical HLA alleles, and their constituent single-nucleotide variants and corresponding amino acids separately. The MHC-specific imputation in the full SCZ sample revealed an additional SNP with genome-wide significance (rs1264353, $P = 3.89 \times 10^{-8}$). Two classic HLA alleles were also among the top results (HLA_B*0801, $P = 2.36 \times 10^{-6}$, HLA_A*0101, $P = 2.83 \times 10^{-6}$; Supplementary Figure 4).

Meta-analysis with PGC-SCZ

With the PGC samples, we conducted a meta-analysis of 11 271 SCZ cases and 14 601 controls (Table 2). The MHC region manifests the most significant association, and was driven by strong signals in both PGC-SCZ and the Swedish subjects (rs17693963, $P = 3.08 \times 10^{-11}$, OR = 1.24). This was followed by a chromosome 10 locus containing *CNNM2*, *C10orf32* and other genes (rs11191580, $P = 1.73 \times 10^{-9}$, OR = 1.23), and a chromosome 7 locus encompassing *MAD1L1* (rs12666575, $P = 1.75 \times 10^{-9}$, OR = 0.89), which had not attained genome-wide significance in the PGC-SCZ samples alone. Other regions of convergent and genome-wide association include the *NT5C2* locus on chromosome 10 (rs1926034, $P = 1.37 \times 10^{-8}$, OR = 0.89) and a chromosome 5 region lacking in annotated genes (rs7709645, $P = 3.80 \times 10^{-8}$, OR = 1.11), which was not previously noted in the results from either sample.

Pathway analysis results

Top PGC-SCZ results did show evidence of enrichment in the SCZ full-sample results ($P = 0.044$). The 301 targets of *MIR137* were not overrepresented in the top results for the SCZ new ($P = 0.97$) or full samples ($P = 0.94$). In the synaptic gene tests, the endocytosis-related set was significantly enriched in the full SCZ ($P = 0.046$) and combined SCZ–BD ($P = 0.027$) results. Calcium channel genes did not show enrichment in the BD or combined SCZ–BD results ($P = 1$). No single Gene Ontology term was significant in the SCZ–BD results following multiple testing correction; however, more genes contained within Gene Ontology categories demonstrated nominal associations of $P < 0.05$ than expected by chance ($P = 0.0055$; Supplementary Table 2).

Polygenic scoring results

Using the PGC-SCZ sample minus the Swedish subjects as the discovery set, we demonstrated enrichment of putatively associated SCZ score alleles in both the SCZ and BD target case groups, compared with controls for all discovery pT. In the full Swedish SCZ sample, we were able to predict 3.6% of variance explained ($P < 2 \times 10^{-16}$ for all pT). A smaller, but still significant, percent of variance explained (1.4%) was evident for the BD sample, supporting evidence for shared genetic risk between these two disorders ($P < 5.92 \times 10^{-4}$ for all pT). These results corroborate prior reports indicating a polygenic contribution of multiple common SNPs to risk for both SCZ and BD^{10,14} (Supplementary Table 3).

Table 1. Regions significant at $P < 1 \times 10^{-5}$ for each of the four main association analyses and genes located in these regions

CHR	SNP	Allele	OR	P-values	N	Position	KB	Genes
<i>SCZ: new subjects</i>								
2	rs12998068	A	1.40	1.50×10^{-6}	26	154931117–155038981	107.9	<i>GALNT13</i>
17	rs11150863	A	0.78	3.49×10^{-6}	45	76071303–76340856	269.6	<i>NPTX1, KIAA1303</i>
21	rs2827775	T	0.80	4.73×10^{-6}	31	23208407–23347097	138.7	No genes
5	rs10515677	T	0.62	7.47×10^{-6}	9	151887279–152400647	513.4	No genes
6	rs886424	T	0.71	8.80×10^{-6}	151	30404232–31351863	947.6	MHC, many genes
6	rs9346146	T	1.24	9.78×10^{-6}	18	69024600–69115229	90.6	No genes
16	rs821470	A	1.28	9.83×10^{-6}	6	55608655–55623057	14.4	<i>NLRC5</i>
22	rs4820902	T	0.78	9.98×10^{-6}	4	29468869–29501274	32.4	<i>OSBP2</i>
<i>SCZ: full sample</i>								
6	rs886424	T	0.68	4.54×10^{-8}	158	30404232–31380240	976.0	MHC, many genes
1	rs984402	T	0.79	3.43×10^{-7}	20	239156776–239228271	71.5	<i>RGS7</i>
6	rs2524005	A	0.74	3.81×10^{-7}	214	29519411–30387109	867.7	MHC, many genes
4	rs4580657	A	0.61	1.61×10^{-6}	3	102744068–102787481	43.4	No genes
6	rs3117099	A	0.79	4.12×10^{-6}	138	31986412–32863268	876.9	MHC, many genes
13	rs1409497	A	1.39	5.17×10^{-6}	22	22741893–22887738	145.8	<i>SGCG, SACS</i>
6	rs3094228	T	1.31	5.96×10^{-6}	58	31420305–31915519	495.2	MHC, many genes
5	rs4262150	T	0.81	6.34×10^{-6}	74	151771819–152489194	717.4	<i>NMUR2</i>
9	rs7872515	A	0.73	6.65×10^{-6}	23	93572203–93991056	418.9	<i>SPTLC1, ROR2</i>
6	rs652888	A	1.29	7.28×10^{-6}	43	31573896–32447054	873.2	MHC, many genes
6	rs13213152	A	1.42	8.43×10^{-6}	174	27960336–28946608	986.3	MHC, many genes
6	rs2235698	A	0.80	8.59×10^{-6}	11	29525796–29663757	138.0	MHC, many genes
9	rs9411508	A	0.49	8.59×10^{-6}	1	135177441–135177441	0.0	<i>SURF6</i>
2	rs12998068	A	1.32	8.60×10^{-6}	24	154931117–155038981	107.9	<i>GALNT13</i>
6	rs2171250	A	1.30	8.68×10^{-6}	5	153766741–153789744	23.0	No genes
6	rs406511	C	1.41	9.07×10^{-6}	39	28999155–29585800	586.6	MHC, many genes
<i>BD</i>								
4	rs17746001	T	2.03	3.22×10^{-7}	7	180892619–180921485	28.9	No genes
12	rs17040430	T	1.89	9.22×10^{-7}	2	107243140–107266950	23.8	<i>CMKLR1</i>
1	rs271738	A	1.32	4.81×10^{-6}	28	232723219–232828069	104.9	<i>IRF2BP2</i>
8	rs289585	T	0.69	5.18×10^{-6}	5	13236908–13304907	68.0	<i>DLC1</i>
9	rs3011267	A	1.41	9.82×10^{-6}	4	134871014–134890520	19.5	<i>GTF3C5, GFI1B</i>
<i>SCZ and BD combined</i>								
4	rs17746001	T	1.75	2.83×10^{-7}	11	180852651–180921485	68.8	No genes
6	rs2524005	A	0.76	4.95×10^{-7}	247	29519411–30473427	954.0	MHC, many genes
9	rs7872515	A	0.72	5.79×10^{-7}	33	93575453–93991056	415.6	<i>SPTLC1, ROR2</i>
6	rs886424	T	0.73	9.25×10^{-7}	119	30623022–31380240	757.2	MHC, many genes
3	rs2054399	A	0.82	1.29×10^{-6}	29	179688504–179854731	166.2	<i>KCNMB2</i>
16	rs821470	A	1.25	2.56×10^{-6}	6	55608655–55623057	14.4	<i>NLRC5</i>
6	rs3117099	A	0.80	2.70×10^{-6}	139	31986412–32863268	876.9	MHC, many genes
8	rs1484642	C	1.43	3.33×10^{-6}	24	9931186–10039190	108.0	<i>MSRA</i>
10	rs181500	T	1.24	4.84×10^{-6}	4	119169072–119174309	5.2	No genes
5	rs4262150	T	0.82	7.19×10^{-6}	61	151771819–152489194	717.4	<i>NMUR2</i>
5	rs253959	T	1.21	8.09×10^{-6}	96	115440636–115779788	339.2	<i>COMMD10</i>
2	rs2058710	A	1.24	9.39×10^{-6}	10	206766767–207166124	399.4	<i>ZDBF2, GPR1, ADAM23</i>

Abbreviations: SCZ, schizophrenia; BD, bipolar disorder; N, number of SNPs in the reported region; CHR, chromosome; SNP, single-nucleotide polymorphism; OR, odds ratio; MHC, major histocompatibility complex.

Position = hg18 coordinates. Genes in these regions or the 20-kb flanking regions were identified using the UCSC Genome Browser.

CNV analysis results

We investigated the role of CNVs > 100 kb in our new SCZ and BD samples (Table 3). Overall, deletions were enriched to a greater extent in the SCZ sample than in the BD subjects compared with controls. Deletions observed only once were more common in both case groups compared with controls (SCZ: $P = 0.003$; BD: $P = 0.013$), whereas the largest size category CNVs > 500 kb, were only significantly enriched in SCZ ($P = 0.00048$). However, it is of note that the gene count in deletions > 500 kb was significantly enriched in both SCZ and BD. Duplications did not show significant overrepresentation for either case group for any size or frequency examined.

Many large CNVs previously implicated in SCZ, BD, and other psychiatric and neurological conditions were also observed in this

sample (Supplementary Table 4). SCZ subjects possessed significantly more duplications at the 16p11.2 locus (Supplementary Figure 5; nine cases, one control; $P = 0.0035$) and four SCZ cases had deletions at 22q11, and one control showed an atypical partial deletion (Supplementary Figure 6; $P = 0.033$). A novel, nominal association was also observed for duplications in SCZ and BD combined at 9q34.3 (31 cases, 11 controls; $P = 0.0014$, genome-wide $P = 0.052$; Supplementary Figure 7). Including the previously reported subjects (six SCZ duplications and one control) strengthened the association ($P = 0.0008$, genome-wide $P = 0.023$). However, no enrichment for duplications at this locus was observed in cases in another SCZ sample (International Schizophrenia Consortium, $P = 0.37$)⁴² or a combined SCZ–BD sample (Genetic Association Information Network, $P = 0.41$).^{43,11}

Table 2. Results of the meta-analysis of PGC-SCZ minus the Swedish samples and all Swedish SCZ results with $P < 1 \times 10^{-6}$

CHR	SNP	PGC-SCZ		Swedish sample		Meta-analysis		N	Position	KB	Genes in range
		P-values	OR	P-values	OR	P-values	OR				
6	rs17693963	2.43×10^{-8}	1.22	1.78×10^{-4}	1.33	3.08×10^{-11}	1.24	239	27050006–28808331	1758	MHC
10	rs11191580	2.85×10^{-9}	1.25	0.104	1.13	1.73×10^{-9}	1.23	42	104277349–104929205	652	NT5C2, CNNM2, and others
7	rs12666575	2.47×10^{-7}	0.89	0.00161	0.87	1.75×10^{-9}	0.89	51	1825144–2312641	487	MAD1L1, SNX8, NUDT1, FTSJ2
10	rs1926034	8.17×10^{-8}	0.89	0.0521	0.92	1.37×10^{-8}	0.89	52	104490649–105378681	888	NT5C2, CNNM2, and others
5	rs7709645	3.65×10^{-6}	1.10	0.00217	1.14	3.80×10^{-8}	1.11	40	60522272–60879681	357	No genes
8	rs7004633	1.08×10^{-8}	0.85	0.514	0.96	6.44×10^{-8}	0.87	95	89257570–89864562	607	MMP16
2	rs13025591	1.62×10^{-6}	0.90	0.0159	0.90	7.74×10^{-8}	0.90	30	236434814–236498899	64	CENTG2
18	rs17597926	1.27×10^{-6}	1.35	0.0677	1.31	1.01×10^{-7}	1.36	17	51046416–51356936	311	TCF4
12	rs4765905	1.76×10^{-6}	1.11	0.0314	1.10	1.60×10^{-7}	1.11	37	2162951–2310725	148	CACNA1C
15	rs12592967	1.51×10^{-5}	1.10	0.0042	1.13	2.69×10^{-7}	1.10	33	59616968–59694773	78	No genes
11	rs10894294	8.61×10^{-6}	0.91	0.0101	0.89	2.81×10^{-7}	0.91	23	130299565–130400186	101	No genes
1	rs1198588	2.58×10^{-6}	0.88	0.0621	0.91	4.71×10^{-7}	0.89	58	98088873–98325420	237	hsa-mir-137, DPYD
1	rs12071951	3.84×10^{-5}	0.89	0.00285	0.85	5.26×10^{-7}	0.88	23	96883528–97057084	174	PTBP2
11	rs11220082	4.64×10^{-6}	1.10	0.0479	1.09	6.66×10^{-7}	1.10	19	124792166–124951853	160	PKNOX2, FEZ1, EI24
10	rs16915157	3.72×10^{-6}	1.13	0.0648	1.10	6.84×10^{-7}	1.13	19	61885077–62016644	132	No genes
8	rs10503256	1.86×10^{-7}	1.12	0.596	1.02	8.61×10^{-7}	1.10	31	4165199–4228540	63	CSMD1

Abbreviations: SCZ, schizophrenia; BD, bipolar disorder; N, number of SNPs within each associated region; CHR, chromosome; SNP, single-nucleotide polymorphism; OR, odds ratio; MHC, major histocompatibility complex; PGC-SCZ, Schizophrenia Psychiatric GWAS Consortium; LD, linkage disequilibrium; GWAS, genome-wide association studies.

Only the top MHC clumped region was retained to prevent this large region of high LD from dominating the results. Position = hg18 coordinates.

Copy number polymorphisms

CNPs represent an additional source of genetic variation that is rarely investigated, but potentially etiologically important. The top SNP in the full sample SCZ analyses, rs886424, (4.54×10^{-8}), tags two CNPs (chr6: 30 994 409–30 994 549, $R^2 = 0.91$; chr6:30 338 498–30 338 570, $R^2 = 0.80$), whereas another SNP in this region (rs3129953; $P = 1.20 \times 10^{-5}$) is in perfect linkage disequilibrium with a third SNP (chr6: 32 667 947–32 668 383, $R^2 = 1$). These results indicate that at least three sub-kilobase deletions segregate with SCZ-associated alleles in the MHC (Supplementary Table 5).

DISCUSSION

In a large Swedish sample, we found a genome-wide significant association in the MHC region for SCZ. This signal was attenuated by addition of subjects with BD. Our results are consistent with a growing body of research supporting MHC involvement in SCZ, but not in BD.^{10–12,14,44} This is the second most gene-dense region in the genome, and the association encompasses a multi-megabase region. In an effort to localize this signal, we imputed the MHC region using methods specifically tailored to this highly polymorphic region, and revealed nominal associations with HLA-B*0801 and HLA-A*0101 previously observed in the partially overlapping International Schizophrenia Consortium sample.¹⁰ Meta-analysis with the PGC-SCZ results also implies a role for the MHC in SCZ, and additional investigations of CNPs uncovered further potentially relevant genetic variation in this region. Association of the MHC region in this genetically similar Swedish SCZ sample provides additional evidence that a false positive due to population stratification is unlikely to underlie this signal.

Although the MHC region contained the only genome-wide significant finding, several other loci attained nominal significance and merit further investigation. For the SCZ analysis in the full sample, the regulator of G-protein signaling 7 (*RGS7*) gene, functioning in inhibition of signal transduction and synaptic vesicle exocytosis, demonstrated the second strongest association. The top SNP for the new SCZ sample analysis falls in the neuronally expressed gene *GALNT13*, identified as one of the genes driving the significance of the 'glycan structures biosynthesis 1' pathway in a prior report.⁴⁵ The same SNP was most significant in both the BD and the combined SCZ–BD analyses; however, this region lacks known genes, and association is not

supported by the PGC–BD results (rs17746001, $P = 0.083$).¹⁸ The larger number of SCZ subjects confers greater power for detection of associations compared with the smaller BD sample, and the lack of genome-wide significant results for the latter should be considered in this context.

Despite the smaller sample and inherent decrease in power for BD, we generally observed similar (and higher) rates of CNVs in SCZ and BD compared with controls; however, statistical significance was more often attained in the much larger SCZ sample. Direct comparison of the two case groups revealed a significant increase for large CNVs (> 500 kb) in the SCZ group ($P = 0.007$), but not for any other size or frequency classes. This is compatible with the growing literature implicating increased burden of large, rare CNVs in SCZ;^{46,42} however, the role of CNVs in BD has been much less clear. We detected a significant excess of singleton deletions in BD as previously reported,⁴⁷ but our results are also consistent with a prior observation of lower CNV rates in BD for other size and frequency categories,⁴⁸ or no CNV enrichment in BD.⁴⁹

The most significant specific CNV we observed, 16p11.2, showed more duplication events in SCZ than controls. This is not a novel finding in SCZ,^{23,25} nor is this locus only implicated in SCZ. Most notably, 16p11.2 duplications are one of the few CNVs with some evidence of association with BD,²³ although only one subject in our BD sample had a duplication at this locus. CNVs at this locus have also shown repeated associations with autism as well as occurrences in cases of developmental delay and mental retardation, syringomyelia, and congenital kidney and urinary abnormalities.^{23,24,50–52} The novel, nominally associated 160-kb duplications in the combined case group at 9q34.3 encompass one uncharacterized open-reading frame and are 60 kb from the brain-expressed gene *KIAA0649*, which may regulate cell proliferation.

The relative homogeneity of this Swedish sample is one of its main strengths; however, replication of these findings in other (particularly non-European) samples is important. It is becoming increasingly clear that both rare CNVs and common SNPs contribute to the etiology of SCZ, although for BD, common genetic variation appears more salient. The mechanisms of action by which the genetic associations reported here and elsewhere confer risk for SCZ and BD are yet unknown. Establishing the genetic and environmental background predisposing carriers of

Table 3. CNV burden and gene count results for only the new samples SCZ ($N = 1505$) versus controls ($N = 2087$), and BD ($N = 834$) versus controls. (a) By CNV frequency and type; (b) by CNV size and type

CNV type	Frequency	CNV burden (number)			CNV burden (gene count)			
		P-values	Total	Case/control ratio	Controls	P-values	Case/control ratio	Controls
(a) SCZ								
All	1	0.15	998	1.08	0.269	0.0022	1.48	0.560
All	2–6	0.45	763	1.01	0.211	0.09	1.22	0.357
All	All	0.10	3347	1.05	0.913	0.0005	1.33	1.629
Deletions	1	0.003	437	1.33	0.107	0.0002	3.04	0.135
Deletions	2–6	0.45	250	1.05	0.068	0.12	1.52	0.069
Deletions	All	0.023	1244	1.13	0.328	0.0001	2.18	0.324
Duplications	1	0.65	695	0.97	0.196	0.44	1.02	0.524
Duplications	2–6	0.84	426	0.91	0.123	0.31	1.10	0.250
Duplications	All	0.50	2103	1.00	0.585	0.10	1.12	1.305
BD								
All	1	0.08	817	1.14	0.269	0.032	1.37	0.560
All	2–6	0.60	614	0.98	0.211	0.71	0.90	0.357
All	All	0.16	2705	1.05	0.913	0.27	1.07	1.629
Deletions	1	0.013	343	1.35	0.107	0.010	2.00	0.135
Deletions	2–6	0.60	198	0.99	0.068	0.89	0.50	0.069
Deletions	All	0.15	981	1.08	0.328	0.15	1.23	0.324
Duplications	1	0.30	582	1.06	0.196	0.33	1.08	0.524
Duplications	2–6	0.87	347	0.88	0.123	0.51	0.99	0.250
Duplications	All	0.31	1724	1.03	0.585	0.42	1.02	1.305
CNV type	Size (kb)	P-values	Total	Case/control ratio	Controls	P-values	Case/control ratio	Controls
(b) SCZ								
All	100–200	0.72	1656	0.97	0.467	0.59	0.96	0.455
All	200–500	0.13	1201	1.07	0.324	0.49	1.00	0.617
All	500+	0.0048	490	1.28	0.122	0.0001	2.00	0.557
Deletions	100–200	0.38	731	1.03	0.201	0.28	1.28	0.127
Deletions	200–500	0.056	369	1.19	0.095	0.13	1.28	0.122
Deletions	500+	0.0035	144	1.64	0.032	0.0001	5.19	0.075
Duplications	100–200	0.87	925	0.93	0.266	0.95	0.84	0.328
Duplications	200–500	0.38	832	1.03	0.229	0.72	0.94	0.495
Duplications	500+	0.12	346	1.15	0.091	0.013	1.51	0.483
BD								
All	100–200	0.21	1383	1.05	0.467	0.93	0.84	0.455
All	200–500	0.14	971	1.09	0.324	0.57	0.97	0.617
All	500+	0.71	351	0.94	0.122	0.08	1.35	0.557
Deletions	100–200	0.43	591	1.02	0.201	0.76	0.83	0.127
Deletions	200–500	0.079	295	1.21	0.095	0.36	1.11	0.122
Deletions	500+	0.38	95	1.10	0.032	0.039	2.12	0.075
Duplications	100–200	0.19	792	1.07	0.266	0.91	0.85	0.328
Duplications	200–500	0.36	676	1.04	0.229	0.66	0.94	0.495
Duplications	500+	0.81	256	0.89	0.091	0.18	1.23	0.483

Abbreviations: CNV, copy number variants; SCZ, schizophrenia; BD, bipolar disorder.

shared risk loci to a particular diagnosis is also of critical importance. As genetic associations are identified and validated, their biological ramifications will become increasingly important to understand as we strive toward developing better diagnostic and therapeutic strategies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are deeply grateful for the participation of all subjects contributing to this research, and to the collection team that worked to recruit them: Emma Flordal-Thelander, Ann-Britt Holmgren, Marie Hallin, Marie Lundin, Ann-Kristin Sundberg, Christina Pettersson, Radja Satgunathan-Dawoud, Sonja Hassellund, Malin Rådström, Birgitta Ohlander, Leila Nyrén, Isabelle Kizling, Louise Frisén, Inger Röhmer, Catharina Lavebratt, Malin Kärrn, Martina Wennberg and Agneta Carswärd-Kjellin. We would also like to thank Professor Martin Schalling for facilitating collection of many subjects with BD. Funding support was provided by the Stanley Center for Psychiatric

Research, Broad Institute from a grant from Stanley Medical Research Institute, NIMH MH077139 (PFS), the Karolinska Institutet, Karolinska University Hospital, the Swedish Research Council, ALF grant from Swedish County Council and Söderström Königskga Foundation.

REFERENCES

- Collins PY, Patel V, Joestl SS, March D, Insel TR, Daar AS *et al*. Grand challenges in global mental health. *Nature* 2011; **475**: 27–30.
- Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going, but still not gone. *Br J Psychiatry* 2010; **196**: 92–95.
- Association AP. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn Association AP: Washington, DC, 1994.
- Perlis RH, Miyahara S, Marangell LB, Wisniewski SR, Ostacher M, DelBello MP *et al*. Long-term implications of early onset in bipolar disorder: data from the first 1000 participants in the systematic treatment enhancement program for bipolar disorder (STEP-BD). *Biol Psychiatry* 2004; **55**: 875–881.
- Angermeyer MC, Kuhn L. Gender differences in age at onset of schizophrenia. An overview. *Eur Arch Psychiatry Neurol Sci* 1988; **237**: 351–364.
- Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med* 2005; **2**: e141.

- 7 Ileva EI, Morris DW, Moates AF, Suppes T, Thaker GK, Tamminga CA. Genetics and intermediate phenotypes of the schizophrenia–bipolar disorder boundary. *Neurosci Biobehav Rev* 2010; **34**: 897–921.
- 8 Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet C Semin Med Genet* 2003; **123C**: 48–58.
- 9 Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF *et al*. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* 2009; **373**: 234–239.
- 10 Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF *et al*. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
- 11 Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I *et al*. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009; **460**: 753–757.
- 12 Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D *et al*. Common variants conferring risk of schizophrenia. *Nature* 2009; **460**: 744–747.
- 13 Steinberg S, de Jong S, Andreassen OA, Werge T, Borglum AD, Mors O *et al*. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum Mol Genet* 2011; **20**: 4076–4081.
- 14 Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA *et al*. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**: 969–976.
- 15 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.
- 16 Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L *et al*. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 2008; **40**: 1056–1058.
- 17 Cichon S, Muhleisen TW, Degenhardt FA, Mattheisen M, Miro X, Strohmaier J *et al*. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet* 2011; **88**: 372–381.
- 18 Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N *et al*. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 2011; **43**: 977–983.
- 19 Guilmatre A, Dubourg C, Mosca AL, Legallac S, Goldenberg A, Drouin-Garraud V *et al*. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry* 2009; **66**: 947–956.
- 20 Mefford HC, Muhle H, Ostertag P, von Spiczak S, Buysse K, Baker C *et al*. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS Genet* 2010; **6**: e1000962.
- 21 Moreno-De-Luca D, Mulle JG, Kaminsky EB, Sanders SJ, Myers SM, Adam MP *et al*. Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am J Hum Genet* 2010; **87**: 618–630.
- 22 Sebat J, Levy DL, McCarthy SE. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet* 2009; **25**: 528–535.
- 23 McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S *et al*. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* 2009; **41**: 1223–1227.
- 24 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.
- 25 Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J *et al*. Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry* 2011; **168**: 302–316.
- 26 Kristiansson E AP, Wistedt B. Validity of the diagnosis of schizophrenia in a psychiatric inpatient register. *Nordisk Psykiatrik Tidsskrift* 1987; **41**: 229–234.
- 27 Dalman C, Broms J, Cullberg J, Allebeck P. Young cases of schizophrenia identified in a national inpatient register—are the diagnoses valid? *Soc Psychiatry Psychiatr Epidemiol* 2002; **37**: 527–531.
- 28 Ekholm B, Ekholm A, Adolfsson R, Vares M, Osby U, Sedvall GC *et al*. Evaluation of diagnostic procedures in Swedish patients with schizophrenia and related psychoses. *Nord J Psychiatry* 2005; **59**: 457–464.
- 29 Sellgren C, Landen M, Lichtenstein P, Hultman CM, Langstrom N. Validity of bipolar disorder hospital discharge diagnoses: file review and multiple register linkage in Sweden. *Acta Psychiatr Scand* 2011; **124**: 447–453.
- 30 Ryden E, Thase ME, Straht D, Aberg-Wistedt A, Bejerot S, Landen M. A history of childhood attention-deficit hyperactivity disorder (ADHD) impacts clinical outcome in adult bipolar patients regardless of current ADHD. *Acta Psychiatr Scand* 2009; **120**: 239–246.
- 31 First MB SRL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)*. American Psychiatric Press, Inc: Arlington, VA, USA, 1996.
- 32 Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S *et al*. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008; **40**: 1253–1260.
- 33 Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 2007; **81**: 1084–1097.
- 34 Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F *et al*. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**: 52–58.
- 35 McMichael AJ, Jones EY. Genetics. First-class control of HIV-1. *Science* 2010; **330**: 1488–1490.
- 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 analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 37 Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genome-wide association studies of nearly all common variants. *Genet Epidemiol* 2008; **32**: 381–385.
- 38 Lee PH, O'Dushlaine C, Purcell SM. Inrich: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* advance online publication, 17 April 2012 (e-pub ahead of print).
- 39 Moskvina V, O'Dushlaine C, Purcell S, Craddock N, Holmans P, O'Donovan MC. Evaluation of an approximation method for assessment of overall significance of multiple-dependent tests in a genome-wide association study. *Genet Epidemiol* 2011; **35**: 861–866.
- 40 Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD *et al*. Global variation in copy number in the human genome. *Nature* 2006; **444**: 444–454.
- 41 Handsaker RE, Korn JM, Nemesh J, McCarroll SA. Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nat Genet* 2011; **43**: 269–276.
- 42 International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**: 237–241.
- 43 Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W *et al*. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry* 2009; **14**: 755–763.
- 44 Lee KW, Woon PS, Teo YY, Sim K. Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: What have we learnt? *Neurosci Biobehav Rev* 2012; **36**: 556–571.
- 45 O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D *et al*. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry* 2011; **16**: 286–292.
- 46 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.
- 47 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.
- 48 McQuillin A, Bass N, Anjorin A, Lawrence J, Kandaswamy R, Lydall G *et al*. Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. *Eur J Hum Genet* 2011; **19**: 588–592.
- 49 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.
- 50 Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA *et al*. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J Med Genet* 2010; **47**: 332–341.
- 51 Schaaf CP, Goin-Kochel RP, Nowell KP, Hunter JV, Aleck KA, Cox S *et al*. Expanding the clinical spectrum of the 16p11.2 chromosomal rearrangements: three patients with syringomyelia. *Eur J Hum Genet* 2011; **19**: 152–156.
- 52 Sampson MG, Coughlin II CR, Kaplan L, Conlin LK, Meyers KE, Zackai EH *et al*. Evidence for a recurrent microdeletion at chromosome 16p11.2 associated with congenital anomalies of the kidney and urinary tract (CAKUT) and Hirschsprung disease. *Am J Med Genet A* 2010; **152A**: 2618–2622.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)