Genome-wide significant locus for Research Diagnostic Criteria Schizoaffective Disorder Bipolar Type.

Elaine K Green¹, Arianna Di Florio²,³, Liz Forty², Katherine Gordon-Smith⁴, Detelina Grozeva⁵, Christine Fraser², Alex Richards³, Jennifer L Moran⁶, Shaun Purcell⁶,⁷, Pamela Sklar⁶,⁷ George Kirov², Michael J Owen², Michael C O’Donovan², Nick Craddock², Lisa Jones⁴, Ian R Jones².

¹School of Biomedical and Healthcare Sciences, Plymouth University Peninsula Schools of Medicine and Dentistry, Portland Square Building, Drake Circus, Plymouth PL4 8AA
²MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, CF24 4HQ, UK
³Department of Psychiatry, University of North Carolina at Chapel Hill, USA
⁴Department of Psychological Medicine, Worcester University, WR2 6AJ, UK
⁵Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK
⁶Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
⁷Division of Psychiatric Genomics in the Department of Psychiatry, Friedman Brain Institute, and Institute for Genomics and Multiscale Biology, Icahn School of Medicine, Mount Sinai, New York, NY, USA

Running Title: GWS association with RDC-SABP.

*Correspondence: Dr E K Green, School of Biomedical and Healthcare Sciences, Plymouth University Peninsula Schools of Medicine and Dentistry, Portland Square Building, Drake Circus, Plymouth PL4 8AA. UK (Email: elaine.green@plymouth.ac.uk, tel: 44 (0)1752584465, fax: 44 (0)1752584605).
Abstract
Studies have suggested that Research Diagnostic Criteria for Schizoaffective Disorder Bipolar type (RDC-SABP) might identify a more genetically homogenous subgroup of bipolar disorder. Aiming to identify loci associated with RDC-SABP we have performed a replication study using independent RDC-SABP cases (n=144) and controls (n=6,559), focusing on the 10 loci that reached a P-value <10^{-5} for RDC-SABP in the Wellcome Trust Case Control Consortium (WTCCC) bipolar disorder sample. Combining the WTCCC and replication datasets by meta-analysis (combined RDC-SABP, n=423, Controls, n=9,494) we observed genome-wide significant association at one SNP, rs2352974, located within the intron of the gene TRAIP on chromosome 3p21.31 (P-value, 4.37x10^{-8}). This locus did not reach genome-wide significance in bipolar disorder or schizophrenia large Psychiatric Genomic Consortium datasets, suggesting that it may represent a relatively specific genetic risk for the bipolar subtype of schizoaffective disorder.

Key Words:
Schizoaffective disorder bipolar type, TRAIP, Research Diagnostic Criteria, SABP
Introduction

In the recent years many successful large scale genome studies have identified genetic susceptibility loci, including common, rare and structural variants which confer risk individually to schizophrenia (SZ) and bipolar disorder (BD), and a combination of such disorders (PGC-BD, 2011; PGC-SZ, 2014; Kirov, 2015; Green et al., 2016). Often such studies, in addition to probands with BD and SZ, include a group which meet criteria for schizoaffective disorder. Traditional family studies have shown evidence of familial overlap between schizoaffective disorder and both SZ and BD, which has been confirmed by Scandinavian population registered based studies (Cardno & Owen, 2014). Depending on the proportion of cases diagnosed as schizoaffective disorder, the inclusion of these individuals may impact on the genetic findings when considering the relationship between SZ and BD (Cardno & Owen, 2014).

The concept and diagnostic criteria for Schizoaffective Disorder (SAD) have been controversial with DSM-IV-TR noting that the SAD category “fills a necessary and important hole in the diagnostic system, but unfortunately it does not do its job very well”. The approaches taken by the various classification systems have differed, with Research Diagnostic Criteria (RDC) for schizoaffective disorder bipolar type broader than those proposed by the World Health Organisation and the American Psychiatric Association (Spitzer et al., 1978)). RDC does not require psychotic symptoms in the absence of a mood episode and rather gives a list of specific psychotic symptoms “suggesting schizophrenia” occurring in a manic episode that would point to a diagnosis of SABP. These differences result in more patients being diagnosed in the schizoaffective categories under RDC compared to the ICD and DSM systems.

Genetic studies focusing on individuals rated by Research Diagnostic Criteria as Schizoaffective Disorder Bipolar type (RDC-SABP) have previously been performed. Examining the Wellcome Trust Case Control Consortium dataset (WTCCC, 2007), we noted that RDC-SABP stood out from other subsets of the BD sample as having a significantly excess of strong association signals (P<10^-5) and hence ‘may be of particular use to for identifying common susceptibility loci GWAS’ (Hamshere et al., 2009). In addition, variation at genes encoding GABA_A–receptor subunits were associated with risk of RDC-SABP and this association was relatively specific to this diagnostic subset, with no association to SZ or BD (Craddock et al., 2010). This finding was replicated in an independent study (Breuer et al., 2011). Finally, polygenic score analysis of RDC-SABP using SZ derived polygenic scores showed that the polygenic influences on SZ had a greater overlap with SABP than those for the remaining BD individuals (Hamshere et al., 2011).

Aiming to identify loci that are associated with RDC-SABP at statistically stringent levels of significance (P-value < 5x10^-8) we have genotyped a replication sample using the Illumina Infinium
HD genotyping array, the ImmunoChip, focusing on the 10 SNPs that reached a P-value threshold of < 10^{-5} in the WTCCC study (Hamshere et al., 2009), and combined the 2 datasets by meta-analysis (total RDC-SABP cases, n= 423 and controls, n= 9,494).

Materials and Methods

Samples

All participants were unrelated, white European, living in the British Isles. The protocols and procedures were approved by the relevant ethics review panels where patients were recruited.

Original WTCCC sample set.

The WTCCC bipolar disorder sample and dataset has been previously reported, and as such the sample and collection information is not included (WTCCC, 2007). Analyses of subsets of these BD samples, including those individuals rated as RDC-SABP have also been reported (n = 279) (Hamshere et al., 2009; Craddock et al., 2010; Green et al., 2010). RDC is a broader definition of SABP, and provide more delineation between individuals on the basis of the pattern of mood psychotic symptomatology than rating by DSM-IV (APA, 1994; Hamshere et al., 2011). Inter-rater reliability was formally assessed using 20 random cases. Mean kappa statistics were 0.86 for RDC.

Replication sample set

The independent replication RDC-SABP (n=144) sample set was part of the bipolar disorder sample, the BDRN sample (n=1,849). A description of this BD collection has been detailed in Green et al., (2013, 2016). WTCCC2 set was used as the control sample (n = 6,599), and the characteristics and recruitment of which have been described in WTCCC 2007. These controls are not screened to exclude the presence of psychiatric illness.

Genotyping

The genotyping was performed using the custom Illumina Infinium HD genotyping array, the ImmunoChip. The ImmunoChip BD genotyping study has been previously reported for the 1218 BD cases and 2913 controls (Green et al., 2013) but not for the subset of RDC SABP cases. Additional samples were genotyped at University College London (UCL) to increase the sample size, including 631 BD cases (44 RDC-SABP cases) and 3,646 WTCCC2 controls. In total, the replication RDC-SABP sample consists of 144 RDC-SABP cases and 6,556 WTCCC2 controls, which are independent of the original WTCCC GWAS.
This study focuses on 10 SNPs that were included on the ImmunoChip as part of the ‘investigator-specific SNP selection for replication’ and were independently associated SNPs ($r^2 < 0.2$) with an association signal of $P < 10^{-5}$ for RDC-SABP cases against controls in our previous study Hamshere et al., 2009. It is worth noting that in this study the Cochran-Armitage trend test of genotype distributions with disease was employed, whereas the data presented here was analysed using logistic regression of disease state with a genomic inflation factor ($\lambda$) of 1.06. As such the P-values and ORs stated may differ slightly from the original publication and the P-values for 2 SNPs are slightly $> 10^{-5}$. We note that following Principal Component analysis (data shown in supplementary methods) SNP rs2352974 remains genome-wide significant when combined with the replication dataset by meta-analysis.

**Statistical analysis**

A brief summary of the methodology is described here and more detailed description is available in the Supplementary Materials section. The BDRN sample set was genotyped on the ImmunoChip at either the Sanger Institute or UCL sequencing facility. The bipolar disorder dataset genotyped at the Sanger Institute has been published (Green et al., 2013) and this genotype calling and quality control pipeline was implemented for the sample genotyped at UCL. Briefly, the genotypes were called by GenoSNP software (Giannoulatou et al., 2008). The data management and quality control assessment was performed using PLINK (v1.07) (Purcell et al., 2007) and a series of shell scripts initially for all BD and control samples.

We planned to combine the data genotyped at the two centres. In order to highlight any potential ‘batch effects’ problems that might prevent the combining of the data, we included 9 identical samples from the first centre to be genotyped by UCL. The concordance rate for the 9 samples across overlapping SNPs ($n=96,184$) was very high reaching 99.997%. Thus we felt confident in combining the datasets.

From the total BD dataset, 144 RDC-SABP and controls were extracted ($n = 6,556$) and quality control analysis performed. Principal Component Analysis (PCA) was performed with Eigenstrat on the combined sample and any individual outliers that did not cluster near to the HapMap European individuals were removed in order to maximise the ethnic homogeneity of our sample (Patterson et al., 2006; Price et al., 2006). The genomic inflation factor was calculated using 43K SNPs in relative linkage ($\alpha=1$).

**Meta-analysis**
The RDC-SABP replication dataset was combined with the original RDC-SABP (WTCCC 2007) dataset by fixed-effects meta-analysis using PLINK (v1.07) (Purcell et al., 2007) to estimate a common odds ratio weighted by individual study standard errors (SE).

**Results**

An independent replication sample of 144 cases (RDC-SABP) and 6,559 controls SNPs were genotyped on the ImmunoChip Illumina array. We have focused on 10 SNPs that showed an independent association ($r^2 < 0.2$) signal at $P < 10^{-5}$ for RDC-SABP against controls in our previous study of the WTCCC dataset previously (Hamshere et al., 2009). We combined our replication data with the WTCCC SNP data by fixed effect meta-analysis. SNP, rs2352974, on chromosome 3p21.31 met genome-wide association ($P$-value = $4.37 \times 10^{-8}$, OR = 0.67). This SNP resides within the intronic region of the gene, *TRAIP* (TRAF interacting protein).

A meta-analysis of all SNPs on the ImmunoChip was also performed (data not presented). No additional individual SNP was associated at levels that exceed the accepted genome-wide levels of significance ($P < 5 \times 10^{-8}$).
Table 1. Meta-analysis of WTCCC RDC-SABP (cases n=279, controls n=2,938) and ImmunoChip RDC-SABP data (cases n=144, controls n=9,497).

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>BP</th>
<th>A1</th>
<th>A2</th>
<th>WTCCC RDC-SABP data</th>
<th>Immunochip, RDC_SABP data</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Cases</strong></td>
<td><strong>Controls</strong></td>
<td><strong>OR [95%CI]</strong></td>
</tr>
<tr>
<td>rs4027132</td>
<td>2</td>
<td>12037492</td>
<td>G</td>
<td>A</td>
<td>0.357</td>
<td>0.459</td>
<td>7.69E-06</td>
</tr>
<tr>
<td>rs2352974</td>
<td>3</td>
<td>49890613</td>
<td>T</td>
<td>C</td>
<td>0.387</td>
<td>0.499</td>
<td>1.01E-06</td>
</tr>
<tr>
<td>rs4279178*</td>
<td>4</td>
<td>47068330</td>
<td>A</td>
<td>G</td>
<td>0.401</td>
<td>0.589</td>
<td>2.96E-06</td>
</tr>
<tr>
<td>rs7680321</td>
<td>4</td>
<td>47145107</td>
<td>C</td>
<td>T</td>
<td>0.136</td>
<td>0.079</td>
<td>1.02E-05</td>
</tr>
<tr>
<td>rs13154602</td>
<td>5</td>
<td>76395917</td>
<td>A</td>
<td>C</td>
<td>0.363</td>
<td>0.269</td>
<td>3.59E-06</td>
</tr>
<tr>
<td>rs1171115</td>
<td>6</td>
<td>84229623</td>
<td>C</td>
<td>T</td>
<td>0.357</td>
<td>0.264</td>
<td>9.81E-06</td>
</tr>
<tr>
<td>rs7900962</td>
<td>13</td>
<td>43399245</td>
<td>G</td>
<td>A</td>
<td>0.253</td>
<td>0.173</td>
<td>6.92E-06</td>
</tr>
<tr>
<td>rs16942644</td>
<td>15</td>
<td>89612736</td>
<td>A</td>
<td>G</td>
<td>0.171</td>
<td>0.102</td>
<td>2.00E-06</td>
</tr>
<tr>
<td>rs4780811</td>
<td>16</td>
<td>6132787</td>
<td>G</td>
<td>A</td>
<td>0.041</td>
<td>0.014</td>
<td>1.10E-05</td>
</tr>
<tr>
<td>rs4818065</td>
<td>21</td>
<td>41037723</td>
<td>A</td>
<td>G</td>
<td>0.272</td>
<td>0.182</td>
<td>6.97E-07</td>
</tr>
</tbody>
</table>

Table 1. Meta-analysis of WTCCC RDC-SABP (cases n=279, controls n=2,938) and ImmunoChip RDC-SABP data (cases n=144, controls n=9,497).

Abbreviations: CHR, chromosome; BP, position in base pairs for UCSC Build hg19; A1, allele 1; A2, allele 2; WTCCC, Wellcome Trust Case Control Consortium; MAF, minor allele frequency; OR, odds ratio; RDC, Research Diagnostic Criteria; SABP, Schizoaffective disorder, Bipolar type; Gene, gene symbol is followed by the distance between the SNP and the reference sequence gene location.

The SNPs listed are those with an independent association ($r^2 < 0.2$) signal at $P < 10^{-5}$ for Research Diagnostic Criteria Schizoaffective disorder, Bipolar type (RDC-SABP) against controls previously reported by Hamshere et al., (2009) analysed originally using the Cochran-Armitage Trend test, here an updated analysis of the WTCCC RDC-SABP dataset has been performed using logistic regression of disease state with a genomic inflation factor ($\lambda$) of 1.06, as such P-values and OR may alter slightly from the original publication. Note: rs4786811 is included in the meta-analysis although the MAF is < 0.05 in both cases and controls. rs6414684 was merged with rs4279178*. 
Discussion

Combining an independent sample with our previous dataset (Hamshere et al., 2009), we report a novel locus reaching genome-wide significant association with RDC-SABP at the intronic SNP rs2352974 (P-Value = 4.37x10^{-8}, OR= 0.67) on chromosome 3p21.31 at TRAIP (TRAF interacting protein). In comparison, this loci was not genome-wide significantly associated in either the large Psychiatric Genomics Consortium (PGC) BD (P=0.39, OR=1.01) or SZ meta-analysis data (P=0.037, OR=0.98) (PGC-BD, 2011, PGC-SZ, 2014), suggesting that it may be a relatively specific genetic risk for bipolar subtype of schizoaffective disorder. The gene TRAIP encodes an E3 RING ubiquitin ligase. A recent study has reported that mutations within TRAIP are associated with microcephalic primordial dwarfism, and identified TRAIP as a component of the DNA damage response replication blocking DNA lesions (Harley et al., 2016).

There is much debate around the clinical usefulness and the nosological status of diagnostic category schizoaffective disorder. Discussions include whether schizoaffective disorder is a form of schizophrenia, affective disorder, a combination of the two or should be regarded as a separate disease entity (Craddock et al., 2009). To add to this there are concerns over the poor reliability of diagnosis (Maj et al. 2000; Santelmann et al., 2015) and apparent low diagnostic stability over time (Schwartz et al., 2000; Laursen et al., 2005). Our findings here, of a susceptibility locus specific (i.e. not identified in BD or SZ datasets) for RDC-SABP, combined with our previous genetic findings for SABP do further support the notion that SABP is a partly independent diagnostic category, with some specific biological characteristics not shared with other phenotypes (Craddock et al., 2009, 2010; Hamshere et al., 2009). Larger well phenotypically defined samples, although challenging to collect, we envisage will enable the identification of additional risk loci that are specific to SABP, and loci that also confer risk to both or either BD and/or SZ.

In summary, within our UK RDC-SABP sample we have identified a genome-wide significantly associated locus at an intronic SNP in TRAIP. Our findings further indicate the importance of research examining clinical diagnostic phenotypes, which in turn will be ultimately important for clinical practice.

Acknowledgements

We are indebted to all individuals who have participated in, or helped with, our research. We particularly thank those involved with Bipolar UK-the Bipolar Organization and the Bipolar Disorder Research Network (www.BDRN.org). Sample collection and analysis was supported by the Wellcome Trust (grant 078901) and a grant from the Stanley Medical Research Institute via the Stanley Centre for Psychiatric Research at the Broad Institute of MIT and Harvard. Genotyping was supported by the Wellcome Trust under WTCCC.
Conflict of Interest Disclosure

Michael C O'Donovan has received a consultancy fee from Roche for participation in a discussion about using genetics to identify drug targets.

References


Craddock N, O'Donovan MC, Owen MJ. 2009. Psychosis genetics: modeling the relationship between schizophrenia, bipolar disorder, and mixed (or "schizoaffective") psychoses. Schizophr Bull 35: 482-490.


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575.


Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 44