Bramon, E; Pirinen, M; Strange, A; Lin, K; Freeman, C; Bellenguez, C; Su, Z; Band, G; Pearson, R; Vukcevic, D; Langford, C; Deloukas, P; Hunt, S; Gray, E; Dronov, S; Potter, SC; Tashakkori-Ghanbaria, A; Edkins, S; Bumpstead, SJ; Arranz, MJ; Bakker, S; Bender, S; Bruggeman, R; Cahn, W; Chandler, D; Collier, DA; Crespo-Facorro, B; Dazzan, P; de Haan, L; di Forti, M; Dragovic, M; Giegling, I; Hall, J; Iyegbe, C; Jablensky, A; Kahn, RS; Kalaydjieva, L; Kravariti, E; Lawrie, S; Lins-Zen, DH; Mata, I; McDonald, C; McIntosh, A; Myin-Germeys, I; Ophoff, RA; Pariente, CM; Paunio, T; Picchioni, M; Ripke, S; Rujescu, D; Sauer, H; Shaikh, M; Sussmann, J; Suvisaari, J; Tosato, S; Touloupoulou, T; van Os, J; Walshe, M; Weisbrod, M; Whalley, H; Wiersma, D; Blackwell, JM; Brown, MA; Casas, JP; Corvin, A; Duncanson, A; Jankowski, JAZ; Markus, HS; Mathew, CG; Palmer, CNA; Plomin, R; Rautanen, A; Sawcer, SJ; Trembath, RC; Wood, NW; Barroso, I; Peltonen, L; Lewis, CM; Murray, RM; Donnelly, P; Powell, J; Spencer, CCA; Psychiat Genomics, C; Psychosis Endophenotypes Int, C; Wellcome Trust Case-Control, C (2013) A Genome-wide Association Analysis of a Broad Psychosis Phenotype Identifies Three Loci for Further Investigation. Biological psychiatry, 75 (5). pp. 386-397. ISSN 0006-3223 DOI: https://doi.org/10.1016/j.biopsych.2013.03.033

Downloaded from: http://researchonline.lshtm.ac.uk/1620481/

DOI: 10.1016/j.biopsych.2013.03.033

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/
A Genome-Wide Association Analysis of a Broad Psychosis Phenotype Identifies Three Loci for Further Investigation

Supplemental Information

Supplemental Methods

Principal Component Analysis

After the quality control, for the 4,835 individuals remaining with 695,193 single nucleotide polymorphisms (SNPs), we applied the following further SNP pruning filters: a 10% minor allele frequency, $10^{-3}$ Hardy-Weinberg equilibrium deviation threshold, and all SNPs within a 1500 SNP window had to have $r^2$ below 0.2 (window shift of 150 used). Thus a subset of 71,677 SNPs was selected for principal component analysis using EIGENSOFT version 3.0 (1). Three covariate vectors were obtained and a total of 356 individuals with non-European ancestry were removed (HapMap2 data and the average difference in the probe intensities across SNPs were used).

Bayesian Meta-Analysis

The Bayesian related effects model assumes a multivariate normal distribution with zero mean and covariance matrix $\sigma^2 R$ as the prior distribution for the genetic effects on the log-odds scale. In this study, we have used a value $\sigma = 0.2$ for the standard deviation of the effect size (see (2) for a justification) and in the correlation matrix $R$ we have set the correlation between our data and the two replication data sets to 0.9375 and between the two replication data sets to 0.99. This reflects the fact that the phenotype definition is more similar between the replication data sets than between our data and the replication data sets. The Bayes factor between this related effects model and the null model (all effects are a priori equal to zero) can be approximated by the ratio of two multivariate normal densities,

$$\frac{f(b \mid \text{mean} = 0, \text{var} = \sigma^2 R + S)}{f(b \mid \text{mean} = 0, \text{var} = S)}.$$
evaluated at the effect size estimates from the three studies (collected in vector $b$) and where $S$ is the diagonal matrix whose elements are the squared standard errors of those effect size estimates.

**Polygenic Score Analysis**

As explained in the main text, we performed a polygenic score analysis using the SNPs associated with schizophrenia in the Psychiatric GWAS Consortium (PGC) study. The 19,434 SNPs that were chosen as proxies for some PGC SNPs (see the main text) were aligned by the allele frequencies.

A logistic regression model regressing the case-control status on the three principal components covariates and the polygenic score had pseudo $R^2$ of 0.166 (Nagelkerke's pseudo $R^2$) and 0.0984 (McFadden's pseudo $R^2$), whereas for the model without polygenic score the corresponding values were 0.124 (Nagelkerke's) and 0.0723 (McFadden's). Here we interpret these values by saying that the variance explained by the polygenic scores is 4.2% (Nagelkerke’s scale) or 2.6% (McFadden’s scale). The $P$ value for the polygenic score was $6 \times 10^{-25}$.

Similar analyses were run adding further covariates accounting for the seven centers involved in the study and the $P$ value for polygenic score remained small ($5 \times 10^{-14}$). With the centers as covariates, the pseudo variances explained by the polygenic score were 1.7% (Nagelkerke’s pseudo $R^2$) and 1.4% (McFadden's pseudo $R^2$).
Figure S1. Principal components analysis of discovery data. Plotted is the projection of the study individuals on to the first two principal components (PC) of genetic structure. Individuals are colored according to recruitment locations as given in legend.
**Figure S2.** This Manhattan plot shows the evidence for association at all autosomal single nucleotide polymorphisms that passed quality control. Red line indicates a $P$ value of $5 \times 10^{-8}$ and blue line $1 \times 10^{-5}$. This analysis includes all samples passing quality control in our discovery study: 1,239 cases, 857 unaffected relatives and 2,739 healthy controls.
**Figure S3.** Forest plot showing the evidence of association in the discovery and replication cohorts. The plots show the estimated odds ratio (OR) and the 95% confidence interval at the three single nucleotide polymorphisms presented in Table 2 and Figure 3. Combined estimates use fixed effects meta-analysis. PGC, Psychiatric GWAS Consortium.
Figure S4. Variance explained (McFadden’s pseudo $R^2$) by the polygenic scores by *P*-value threshold. The variance explained is plotted by *P* value threshold applied to the discovery Psychiatric GWAS Consortium schizophrenia sample (3) from a logistic regression model whose only covariates are the 3 principal components. The corresponding significance of the polygenic score at each threshold ranged between $4.0 \times 10^{-11}$ and $5.7 \times 10^{-25}$. SNPs, single nucleotide polymorphisms.
Table S1. Samples prior to quality control and centers where they were collected.

<table>
<thead>
<tr>
<th>Center Location</th>
<th>Number of Samples</th>
<th>Total (%)</th>
<th>Details of Centers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients</td>
<td>Relatives</td>
<td>Controls</td>
</tr>
<tr>
<td>London</td>
<td>406</td>
<td>352</td>
<td>522</td>
</tr>
<tr>
<td>Holland</td>
<td>649</td>
<td>649</td>
<td>1300</td>
</tr>
<tr>
<td>Perth</td>
<td>376</td>
<td>205</td>
<td>194</td>
</tr>
<tr>
<td>Santander - Pamplona</td>
<td>309</td>
<td>1</td>
<td>323</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>48</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>Heidelberg</td>
<td>32</td>
<td>17</td>
<td>47</td>
</tr>
<tr>
<td>Munich</td>
<td>0</td>
<td>0</td>
<td>1457</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1820</td>
<td>1224</td>
<td>3891</td>
</tr>
</tbody>
</table>
### Table S2. Quality control and sample exclusions in the discovery cohort.

<table>
<thead>
<tr>
<th>Clinical Group</th>
<th>Patients</th>
<th>Relatives</th>
<th>Controls</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passed quality control &amp; genotyped</td>
<td>1239</td>
<td>857</td>
<td>2739</td>
<td>4835</td>
<td>69.7</td>
</tr>
<tr>
<td>Genotyping failure</td>
<td>204</td>
<td>83</td>
<td>735</td>
<td>1022</td>
<td>14.7</td>
</tr>
<tr>
<td>Excluded – &gt;2% SNPs missing</td>
<td>67</td>
<td>59</td>
<td>88</td>
<td>214</td>
<td>3.1</td>
</tr>
<tr>
<td>Excluded – duplicate or monozygotic twin</td>
<td>29</td>
<td>19</td>
<td>22</td>
<td>70</td>
<td>1.0</td>
</tr>
<tr>
<td>Excluded – heterozygosity</td>
<td>33</td>
<td>13</td>
<td>24</td>
<td>70</td>
<td>1.0</td>
</tr>
<tr>
<td>Excluded – ancestry</td>
<td>139</td>
<td>107</td>
<td>110</td>
<td>356</td>
<td>5.1</td>
</tr>
<tr>
<td>Excluded – sex mismatch</td>
<td>22</td>
<td>9</td>
<td>26</td>
<td>57</td>
<td>0.8</td>
</tr>
<tr>
<td>Excluded – clinical reasons</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>0.3</td>
</tr>
<tr>
<td>Failed DNA quality control</td>
<td>66</td>
<td>77</td>
<td>147</td>
<td>290</td>
<td>4.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1820</td>
<td>1224</td>
<td>3891</td>
<td>6935</td>
<td></td>
</tr>
</tbody>
</table>

### Table S3. Quality control and exclusions of SNPs.

<table>
<thead>
<tr>
<th>Quality Control of SNPs</th>
<th>SNP Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosomes X, Y and mitochondrial DNA exclusions</td>
<td>38,895</td>
<td>4.2</td>
</tr>
<tr>
<td>SNPs with Mendelian inheritance errors</td>
<td>26,858</td>
<td>2.9</td>
</tr>
<tr>
<td>SNP excluded if &gt; 5% of individuals failed genotyping</td>
<td>11,610</td>
<td>1.2</td>
</tr>
<tr>
<td>SNP excluded if MAF &lt;2%</td>
<td>145,097</td>
<td>15.6</td>
</tr>
<tr>
<td>SNP excluded if HWE ( p ) value (&lt; 1 \times 10^{-6})</td>
<td>2,404</td>
<td>0.3</td>
</tr>
<tr>
<td>SNPs excluded due to poor calling (manually inspected in Evoker)</td>
<td>9,499</td>
<td>1.0</td>
</tr>
<tr>
<td>SNPs that passed all quality control filters</td>
<td>695,193</td>
<td>74.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>929,556</td>
<td></td>
</tr>
</tbody>
</table>

HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency; SNP, single nucleotide polymorphism.
Table S4. Replication of previously published loci for schizophrenia and/or bipolar disorder (sign tests). The table shows the number of previously reported loci showing an effect size in the same direction at the best tag single nucleotide polymorphism (SNP) in Table 1. The sign test is a one-sided binomial test for an increase in the fraction of SNPs where the risk allele is the same. Results are shown for all SNPs (top) and thinned to remove SNPs within 100kb and to exclude the MHC regions (bottom). Figure 1 provides a plot of these data. * indicates SNPs associated with both schizophrenia and bipolar disorder. P values in bold are significant at the 5% level.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Loci with published evidence of association</th>
<th>Loci with same risk allele in our sample (%)</th>
<th>Binomial test P value for enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>All SNPs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia only</td>
<td>24</td>
<td>19 (79%)</td>
<td><strong>0.0033</strong></td>
</tr>
<tr>
<td>Bipolar disorder only</td>
<td>10</td>
<td>5 (50%)</td>
<td>0.623</td>
</tr>
<tr>
<td>Schizophrenia &amp; bipolar disorder *</td>
<td>10</td>
<td>6 (60%)</td>
<td>0.377</td>
</tr>
<tr>
<td>Schizophrenia (including *)</td>
<td>34</td>
<td>25 (73%)</td>
<td><strong>0.0045</strong></td>
</tr>
<tr>
<td>Bipolar disorder (including *)</td>
<td>20</td>
<td>11 (55%)</td>
<td>0.41</td>
</tr>
<tr>
<td>Thinned SNPs excluding MHC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia only</td>
<td>17</td>
<td>13 (76%)</td>
<td><strong>0.025</strong></td>
</tr>
<tr>
<td>Schizophrenia (including *)</td>
<td>27</td>
<td>19 (70%)</td>
<td><strong>0.026</strong></td>
</tr>
<tr>
<td>Bipolar disorder (including *)</td>
<td>18</td>
<td>11 (61%)</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Supplemental References


Membership of the Psychiatric Genomics Consortium (PGC)

Ingrid Agartz\textsuperscript{1,2}, Farooq Amin\textsuperscript{3,4}, Ole A. Andreassen\textsuperscript{1,5}, Maria H. Azevedo\textsuperscript{6}, Nicholas Bass\textsuperscript{7}, Donald W. Black\textsuperscript{8}, Douglas H.R. Blackwood\textsuperscript{9}, Richard Bruggeman\textsuperscript{10}, Nancy G. Buccola\textsuperscript{11}, William Byerley\textsuperscript{12,13}, Wiepke Cahn\textsuperscript{14}, Rita M. Cantor\textsuperscript{15,16}, Khalid Choudhury\textsuperscript{7}, Sven Cichon\textsuperscript{17,18,19}, C. Robert Cloninger\textsuperscript{20}, Aiden Corvin\textsuperscript{21}, Nicholas Craddock\textsuperscript{22,23}, David Curtis\textsuperscript{24}, Mark Daly\textsuperscript{25}, Susmita Datta\textsuperscript{7}, Lieuwe De Haan\textsuperscript{26}, Srdjan Djurovic\textsuperscript{1,27}, Jubao Duan\textsuperscript{28,29}, Frank Dudbridge\textsuperscript{30}, Ayman H. Fanous\textsuperscript{31,32,33}, Robert Freedman\textsuperscript{34}, Nelson B. Freimer\textsuperscript{15}, Marion Friedl\textsuperscript{35}, Pablo V. Gejman\textsuperscript{38,29}, Lyudmila Georgieva\textsuperscript{22,23}, Ina Giegling\textsuperscript{25}, Michelle Gill\textsuperscript{31}, Hugh Gurling\textsuperscript{7}, Marlan I. Hamshere\textsuperscript{22,23}, Thomas Hansen\textsuperscript{36}, Annette M. Hartmann\textsuperscript{35}, Peter A. Holmans\textsuperscript{22,23}, Christina M. Hultman\textsuperscript{35}, Andrés Ingason\textsuperscript{20}, Anna Kahler\textsuperscript{37}, René S. Kahn\textsuperscript{14}, Matthew C. Keller\textsuperscript{34}, Kenneth S. Kendler\textsuperscript{31,38,39}, Elaine Kenny\textsuperscript{21}, Yunjung Kim\textsuperscript{40}, George K. Kirov\textsuperscript{22,23}, Bettina Konté\textsuperscript{39}, Lydia Krabbendam\textsuperscript{41}, Robert Krasucki\textsuperscript{7}, Jacob Lawrence\textsuperscript{7}, Phil Hyoun Lee\textsuperscript{25}, Todd Lencz\textsuperscript{42,43,44}, Douglas F. Levinson\textsuperscript{45}, Jeffrey A. Lieberman\textsuperscript{46}, Dan-Yu Lin\textsuperscript{47}, Don H Linszen\textsuperscript{26}, Patrik Magnusson\textsuperscript{37}, Wolfgang Maier\textsuperscript{48}, Anil K. Malhotra\textsuperscript{42,43,44}, Manuel Mattheisen\textsuperscript{17,19,49}, Morten Mattingsdal\textsuperscript{1,50}, Steve McCarroll\textsuperscript{51}, Andrew M. McIntosh\textsuperscript{9}, Andrew McQuillin\textsuperscript{7}, Helena Medeiros\textsuperscript{57}, Ingrid Melle\textsuperscript{1,5}, Vihra Milanova\textsuperscript{51}, Derek W. Morris\textsuperscript{21}, Valentina Moskvin\textsuperscript{a2,22,23}, Bryan J. Mowry\textsuperscript{52,53}, Inez Myin-Germeys\textsuperscript{41}, Benjamin M. Neale\textsuperscript{54}, Markus M. Nöthen\textsuperscript{17,19}, Michael C. O'Donovan\textsuperscript{22,23}, Colm T. O'Dushlaine\textsuperscript{21}, Ann Olincy\textsuperscript{34}, Line Olsen\textsuperscript{36}, Roel A. Ophoff\textsuperscript{35,56,15}, Michael J. Owen\textsuperscript{22,23}, Carlos N. Pato\textsuperscript{57}, Michele T. Pato\textsuperscript{57}, Benjamin S. Pickard\textsuperscript{58}, Jonathan Pimm\textsuperscript{7}, Danielle Posthuma\textsuperscript{59,60}, Shaun Purcell\textsuperscript{54}, Vinay Puri\textsuperscript{7}, Digby Quested\textsuperscript{41}, Henrik B. Rasmussen\textsuperscript{36}, Marcella Rietschel\textsuperscript{48,62}, Stephan Ripke\textsuperscript{25}, Lizzy Rossin\textsuperscript{25}, Douglas M. Ruderfer\textsuperscript{25}, Dan Rujescu\textsuperscript{35}, Alan R. Sanders\textsuperscript{28}, Thomas G. Schulze\textsuperscript{62,63}, Jianxin Shi\textsuperscript{64}, Jeremy M. Silverman\textsuperscript{65,66}, Pamela Sklar\textsuperscript{25,65}, David St Clair\textsuperscript{67}, T. Scott Stroup\textsuperscript{46}, Patrick F. Sullivan\textsuperscript{37,40,68,69}, Srinivasa Thirumalai\textsuperscript{70}, Jim Van Os\textsuperscript{41}, Peter M. Visscher\textsuperscript{71}, Thomas Werge\textsuperscript{36}, Durk Wiersma\textsuperscript{10}, Stan Zammit\textsuperscript{22,23}.

PGC Affiliations

1 Psychiatry Section, Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
2 Department of Research, Diakonhjemmet Hospital, Oslo, Norway.
3 Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia, USA.
4 Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Atlanta, Georgia, USA.
5 Department of Psychiatry, Oslo University Hospital, Oslo, Norway.
6 Faculty of Medicine, University of Coimbra, Coimbra, Portugal.
7 Molecular Psychiatry Laboratory, Research Department of Mental Health Sciences, University College London Medical School, Windeyer Institute of Medical Sciences, London, UK.
8 Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa, USA.
9 Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK.
10 University Medical Center Groningen, Department of Psychiatry, University of Groningen, Groningen, The Netherlands.
11 School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA.
12 Department of Psychiatry, University of California at San Francisco, San Francisco, California, USA.
13 NCIRE (Northern California Institute for Research and Education), San Francisco, California, USA.
14 Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands.
15 University of California at Los Angeles (UCLA) Center for Neurobehavioral Genetics, University of California at Los Angeles, Los Angeles, California, USA.
16 Department of Human Genetics, University of California at Los Angeles, Los Angeles, California, USA.
59 Vrije Universiteit (VU), Center for Neurogenomics and Cognitive Research (CNCR), Department of Functional Genomics, Amsterdam, The Netherlands.
60 VU Medical Centre, Department of Medical Genomics, Amsterdam, The Netherlands.
61 Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford, UK.
62 Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany.
63 Department of Psychiatry and Psychotherapy, Georg-August-University, Göttingen, Germany.
64 Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA.
65 Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, USA.
66 Department of Psychiatry, Veterans Affairs Medical Center, New York, New York, USA.
67 Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen, UK.
68 Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
69 Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
70 West Berkshire National Health Service (NHS) Trust, Reading, UK.
71 Queensland Statistical Genetics Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland, Australia.