Genome-wide association study of major depressive disorder

Citation for published version:

Digital Object Identifier (DOI):
10.1038/mp.2010.109

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher’s PDF, also known as Version of record

Published In:
Molecular Psychiatry

Publisher Rights Statement:
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License.
To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned

NR Wray1, ML Pergadia2, DHR Blackwood3, BWJH Penninx4, SD Gordon1, DR Nyholt1, S Ripke5,6, DJ MacIntyre3, KA McGhee3, AW Maclean3, JH Smit4, JH Hottenga4, G Willemsen4, CM Middeldorp4, EJC de Geus4, CM Lewis7, P McGuffin7, IB Hickie8, EJCG van den Oord9, JZ Liu1, S Macgregor1, BP McEvoy1, EM Byrne1, SE Medland1, DJ Statham1,11, AK Henders1, AC Heath2, GW Montgomery1, NG Martin1, DI Boomsma4, PAF Madden2 and PF Sullivan10

1 Genetic Epidemiology, Molecular Epidemiology, Psychiatric Genetics and Queensland Statistical Genetics Laboratories, Queensland Institute of Medical Research, Brisbane, QLD, Australia; 2 Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA; 3 Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK; 4 Department of Biological Psychology and Medical Center, VU University, Amsterdam, The Netherlands; 5 Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 6 Broad Institute of Harvard and MIT, Cambridge, MA, USA; 7 Department of Medical and Molecular Genetics, King’s College London, MRC SGDP Centre, Institute of Psychiatry, London, UK; 8 Clinical Research Unit, Brain and Mind Research Institute, University of Sydney, NSW, Australia; 9 Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth University, Richmond, VA, USA and 10 Department of Genetics, University of North Carolina, Chapel Hill, NC, USA

Major depressive disorder (MDD) is a common complex disorder with a partly genetic etiology. We conducted a genome-wide association study of the MDD2000+ sample (2431 cases, 3673 screened controls and >1 M imputed single-nucleotide polymorphisms (SNPs)). No SNPs achieved genome-wide significance either in the MDD2000+ study, or in meta-analysis with two other studies totaling 5763 cases and 6901 controls. These results imply that common variants of intermediate or large effect do not have main effects in the genetic architecture of MDD. Suggestive but notable results were (a) gene-based tests suggesting roles for adenylate cyclase 3 (ADCY3, 2p23.3) and galanin (GAL, 11q13.3); published functional evidence relates both of these to MDD and serotonergic signaling; (b) support for the bipolar disorder risk variant SNP rs1006737 in CACNA1C (P = 0.020, odds ratio = 1.10); and (c) lack of support for rs2251219, a SNP identified in a meta-analysis of affective disorder studies (P = 0.51).

We estimate that sample sizes 1.8- to 2.4-fold greater are needed for association studies of MDD compared with those for schizophrenia to detect variants that explain the same proportion of total variance in liability. Larger study cohorts characterized for genetic and environmental risk factors accumulated prospectively are likely to be needed to dissect more fully the etiology of MDD.

Molecular Psychiatry (2012) 17, 36–48; doi:10.1038/mp.2010.109; published online 2 November 2010

Keywords: major depressive disorder; depression; genome-wide association study; CACNA1C; ADCY3; GAL

Introduction

Major depressive disorder (MDD) is a common and debilitating disorder with pervasive impact on the quality of life for both sufferers and their families. Lifetime prevalence is estimated to be ~15%,1,2 and is consistently estimated to be twice as common in women as men.1,4 MDD is associated with high morbidity, reflected in estimates of burden of disease and years lost in productivity,1,5 and excess mortality from suicide6 and other causes. Our understanding of the etiology of MDD remains fragmented, despite wide-ranging research, but is the key to effective prevention and treatment.

MDD is familial, with heritability estimated to be 0.37 (95% confidence interval (CI) 0.31–0.42) and both early age of onset and recurrence of depression are associated with higher familial aggregation.7,8 This implied genetic etiology has motivated studies designed to identify specific genetic variants associated with MDD. Results from genome-wide linkage,
reviewed in Boomsma et al.,9 and candidate gene association studies10 have shown little consistency and hopes for new progress have spurred on a generation of genome-wide association studies (GWAS). Five GWAS for MDD have been published to date,11–15 each using control samples screened negative for MDD. None of these studies has identified variants that achieve genome-wide significance. Taken together, the results of these studies imply that specific genetic variants individually make very small contributions to the etiology of MDD. In this study, we present the largest GWAS for MDD to date, the MDD2000+ study comprising 2431 cases and 3673 screened controls. We compare our results with reports of the other published MDD GWAS and present a formal meta-analysis of our results with the two other largest studies (5763 cases and 6901 controls).

Materials and methods

Overview
The MDD2000+ project comprises a total of 2431 cases with MDD and 3673 screened controls from different sources and genotyped on different platforms (Tables 1 and 2). Samples were provided by the Queensland Institute of Medical Research (QIMR, Australia), The Netherlands Study of Anxiety and Depression (NESDA), The Netherlands Twin Registry (NTR), the University of Edinburgh (UK), and the Molecular Genetics of Schizophrenia study (controls only, US). Genotyping was conducted on different Illumina and Affymetrix platforms and because the overlap in genotyped single-nucleotide polymorphisms (SNPs) is limited, association analysis is based on a set of >1 M imputed SNPs. The number of SNPs for each analysis set (Table 1), represents genotyped SNPs surviving all quality control (QC) criteria that were used for imputation.

Subject recruitment

QIMR. Study participants were adult twins and their families recruited through the Australian Twin Registry (http://www.twins.org.au). Only unrelated individuals were included in MDD2000+. All participants provided written informed consent under study protocols approved by the QIMR Human Research Ethics Committee. MDD cases were identified through psychiatric questionnaires, either the shortened Composite International Diagnostic Interview16 or the SSAGA-OZ interview instrument (a version of the Semi-Structured Assessment for the Genetics of Alcoholism17 modified for use in Australia), a comprehensive psychiatric interview designed to assess MDD and other psychiatric disorders17 according to DSM-III18 and DSM-IV19 criteria. Structured interviews were administered by trained telephone interviewers, closely supervised by a clinical psychologist. Briefly, from 1988 to 1990, study participants were mailed an extensive health and lifestyle questionnaire, which included the shortened revised Eysenck personality questionnaire.20 Sum scores of 12 item responses in each personality domain resulted in quantitative scores for neuroticism. Between 1992–2000, an unselected subset of these participants were interviewed by telephone using the SSAGA-OZ.21 Over the period 1996–9 sibling pairs that were either concordant or discordant for extreme neuroticism scores (one sibling in the top or bottom decile, the other sibling in the top or bottom quintile) were recruited to complete the Composite International Diagnostic Interview, which provides DSM-IV lifetime diagnoses of MDD.22 Finally, some study participants completed the SSAGA-OZ telephone questionnaire in 2003–2007 as part of alcohol and nicotine dependence studies (the nicotine addiction genetics and Inter-related Project Grant studies described in Table 2 of Hansell et al.23). The Inter-related Project Grant/nicotine addiction genetics studies captured 28% of families who had already participated in earlier studies ascertaining those with either (a) large sibship size or containing a proband with either (b) nicotine dependence or (c) alcohol dependence. For this study, all cases met DSM-IV lifetime criteria for MDD. Screening items for mania were not consistent across interviews and screening items for psychosis were not included; the ability to assess accurately these less common criteria is difficult in large-scale community settings. Therefore, it is possible that a small number of individuals with a primary diagnosis of bipolar disorder or schizophrenia are included in the case group. If multiple cases were present within families then one case was selected in the following order of preference: age of onset <31 years, multiple episodes of depression, co-morbid anxiety disorders and high neuroticism score. Unrelated controls were selected as genotyped individuals from families in which no individuals qualified for diagnoses of MDD or anxiety disorders. If multiple controls were available from a family, the individual with the lowest neuroticism score was preferentially selected, otherwise an individual was selected at random.

NESDA and NTR. Additional MDD cases were selected from two parallel studies, NESDA (http://www.nesda.nl) and NTR (http://www.tweelingenregister.org); NTR also provided screened controls. These samples do not overlap with those included in a prior MDD GWAS24 but are drawn from the same parent studies (although a small number of related individuals are included, see meta-analysis section below). Details of the data collection methods are described elsewhere.9 Similar inclusion and exclusion criteria were used to select MDD cases from both the NESDA and NTR studies. Inclusion criteria were a lifetime diagnosis of DSM-IV MDD as determined by the Composite International Diagnostic Interview,25 age 18–65 years, and self-reported western European ancestry. Those not fluent in Dutch or with a primary diagnosis of schizophrenia
### Table 1: Analysis set statistics: numbers of samples, numbers of SNPs, ages at interview and MDD onset

<table>
<thead>
<tr>
<th>Analysis set</th>
<th>Platform and approximate number of SNPs genotyped</th>
<th>Sample source</th>
<th>N</th>
<th>No. of males</th>
<th>No. of females</th>
<th>No of recurrent and early onset&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Age at interview mean ± s.d.</th>
<th>Age of onset mean ± s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I317</td>
<td>Illumina 317 k</td>
<td>QIMR cases&lt;sup&gt;b&lt;/sup&gt;</td>
<td>84</td>
<td>0</td>
<td>84</td>
<td>31</td>
<td>47.5 ± 12.0</td>
<td>34.5 ± 12.5</td>
</tr>
<tr>
<td></td>
<td>Post-quality control autosomal SNPs</td>
<td>QIMR controls&lt;sup&gt;b&lt;/sup&gt;</td>
<td>237</td>
<td>0</td>
<td>237</td>
<td>—</td>
<td>45.4 ± 13.2</td>
<td>—</td>
</tr>
<tr>
<td>I370</td>
<td>Illumina 370 k</td>
<td>QIMR cases&lt;sup&gt;c&lt;/sup&gt;</td>
<td>737</td>
<td>325</td>
<td>412</td>
<td>264</td>
<td>42.9 ± 9.1</td>
<td>27.8 ± 10.4 (N = 735)</td>
</tr>
<tr>
<td></td>
<td>Post-quality control X chromosome SNPs</td>
<td>QIMR controls&lt;sup&gt;c&lt;/sup&gt;</td>
<td>795</td>
<td>443</td>
<td>352</td>
<td>—</td>
<td>41.0 ± 12.6 (N = 794)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Illumina 610 k</td>
<td>NTR controls&lt;sup&gt;f&lt;/sup&gt;</td>
<td>577</td>
<td>269</td>
<td>308</td>
<td>—</td>
<td>49.7 ± 13.4</td>
<td>—</td>
</tr>
<tr>
<td>I610</td>
<td>Illumina 610 k</td>
<td>QIMR cases&lt;sup&gt;d&lt;/sup&gt;</td>
<td>169</td>
<td>38</td>
<td>131</td>
<td>45</td>
<td>42.8 ± 10.5</td>
<td>31.2 ± 11.5</td>
</tr>
<tr>
<td></td>
<td>Post-quality control X chromosome SNPs</td>
<td>QIMR controls&lt;sup&gt;d&lt;/sup&gt;</td>
<td>428</td>
<td>132</td>
<td>296</td>
<td>—</td>
<td>41.9 ± 12.4</td>
<td>—</td>
</tr>
<tr>
<td>A6.0</td>
<td>Affymetrix 600 k</td>
<td>QIMR cases&lt;sup&gt;e&lt;/sup&gt;</td>
<td>941</td>
<td>298</td>
<td>643</td>
<td>480</td>
<td>42.2 ± 9.9</td>
<td>25.4 ± 9.8 (N=940)</td>
</tr>
<tr>
<td></td>
<td>NESDA/NTR cases&lt;sup&gt;e&lt;/sup&gt;</td>
<td>127</td>
<td>45</td>
<td>82</td>
<td>36</td>
<td>41.3 ± 11.7 (N=126)</td>
<td>28.3 ± 10.5 (N=92)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UoE cases&lt;sup&gt;e&lt;/sup&gt;</td>
<td>373</td>
<td>151</td>
<td>222</td>
<td>289</td>
<td>31.8 ± 14.5 (N=352)</td>
<td>23.3 ± 10.9 (N=352)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MGS controls</td>
<td>1636</td>
<td>918</td>
<td>718</td>
<td>—</td>
<td>52.5 ± 17.2</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>Imputed</td>
<td>Cases</td>
<td>2431</td>
<td>857</td>
<td>1574</td>
<td>1145</td>
<td>41.1 ± 11.4 (N=2409)</td>
<td>26.7 ± 10.8 (N=2372)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Controls</td>
<td>3673</td>
<td>1762</td>
<td>1911</td>
<td>—</td>
<td>48.0 ± 15.8 (N=3613)</td>
<td>—</td>
</tr>
</tbody>
</table>

Abbreviations: MAF, minor allele frequency; MDD, major depressive disorder; MGS, Molecular Genetics of Schizophrenia; NESDA, The Netherlands Study of Anxiety and Depression; NTR, The Netherlands Twin Registry; QIMR, Queensland Institute of Medical Research; SNP, single-nucleotide polymorphism; UoE, University of Edinburgh.

<sup>a</sup>Early onset < 31 years.

<sup>b,c,d,e</sup>Samples with the same superscript had cases and controls genotyped together sometimes in multiple batches.

<sup>f</sup>X chromosome imputed SNPs with MAF > 0.01.

The NTR controls were genotyped on the Illumina Human610-Quad but were included in the I370 analysis set to balance proportions of cases and controls in each set.
Table 2 Descriptive statistics by analysis set and sample source

<table>
<thead>
<tr>
<th>Analysis set</th>
<th>Sample source</th>
<th>Female (%)</th>
<th>Early onset &lt;31 years (%)</th>
<th>Recurrent (%)</th>
<th>Recurrent and early onset (%)</th>
<th>Recurrent and/or early onset and/or family history (%)</th>
<th>Education level (% low, middle, high)</th>
<th>Partner status (% with partner)</th>
<th>Regular smoker (%)</th>
<th>Neuroticism score Mean ± s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QIMR cases (N=84)</td>
<td>100</td>
<td>46</td>
<td>48</td>
<td>37</td>
<td>25</td>
<td>58</td>
<td></td>
<td></td>
<td>40.52 ± 9.81 (N=81)</td>
</tr>
<tr>
<td></td>
<td>QIMR controls (N=237)</td>
<td>100</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>38.50.12 (N=228)</td>
<td>81</td>
<td>16</td>
<td>73 ± 20 (N=237)</td>
</tr>
<tr>
<td></td>
<td>QIMR cases (N=737)</td>
<td>56</td>
<td>65</td>
<td>43</td>
<td>36</td>
<td>40</td>
<td>76</td>
<td></td>
<td></td>
<td>27.39 ± 3.49 (N=724)</td>
</tr>
<tr>
<td></td>
<td>QIMR controls (N=795)</td>
<td>44</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>28.38 ± 3.4 (N=744)</td>
<td>81</td>
<td>59</td>
<td>70 ± 69 (N=737)</td>
</tr>
<tr>
<td></td>
<td>NTR controls (N=577)</td>
<td>53</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>7.56 ± 3.7 (N=573)</td>
<td>92</td>
<td>17</td>
<td>— ± 0.43 (N=565)</td>
</tr>
<tr>
<td></td>
<td>QIMR cases (N=169)</td>
<td>78</td>
<td>55</td>
<td>37</td>
<td>27</td>
<td>22</td>
<td>69</td>
<td></td>
<td></td>
<td>35.40 ± 2.5 (N=161)</td>
</tr>
<tr>
<td></td>
<td>QIMR Controls (N=428)</td>
<td>69</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>25.51 ± 2.4 (N=410)</td>
<td>86</td>
<td>38</td>
<td>— ± 0.29 ± 0.0 (N=405)</td>
</tr>
<tr>
<td></td>
<td>QIMR cases (N=941)</td>
<td>68</td>
<td>77</td>
<td>66</td>
<td>51</td>
<td>35</td>
<td>95</td>
<td></td>
<td></td>
<td>24.39 ± 3.7 (N=924)</td>
</tr>
<tr>
<td></td>
<td>NESDA/NTR Cases (N=127)</td>
<td>65</td>
<td>71</td>
<td>38</td>
<td>28</td>
<td>85 (N=106)</td>
<td>91</td>
<td></td>
<td></td>
<td>27.37 ± 3.5 (N=124)</td>
</tr>
<tr>
<td></td>
<td>UoE cases (N=373)</td>
<td>60</td>
<td>78</td>
<td>100</td>
<td>78</td>
<td>46 (N=362)</td>
<td>100</td>
<td></td>
<td></td>
<td>27.37 ± 3.5 (N=124)</td>
</tr>
<tr>
<td></td>
<td>MGS controls (N=1636)</td>
<td>44</td>
<td>78</td>
<td>100</td>
<td>78</td>
<td>46 (N=362)</td>
<td>100</td>
<td></td>
<td></td>
<td>27.37 ± 3.5 (N=124)</td>
</tr>
<tr>
<td></td>
<td>ALL Cases (N=2431)</td>
<td>65</td>
<td>71</td>
<td>60</td>
<td>47</td>
<td>39</td>
<td>86</td>
<td></td>
<td></td>
<td>40.52 ± 9.81 (N=2431)</td>
</tr>
<tr>
<td></td>
<td>Controls (N=3673)</td>
<td>52</td>
<td>71</td>
<td>60</td>
<td>47</td>
<td>39</td>
<td>86</td>
<td></td>
<td></td>
<td>40.52 ± 9.81 (N=3673)</td>
</tr>
</tbody>
</table>

Abbreviations: MDD, major depressive disorder; MGS, Molecular Genetics of Schizophrenia; NA, not available; NESDA, The Netherlands Study of Anxiety and Depression; NTR, The Netherlands Twin Registry; QIMR, Queensland Institute of Medical Research; UoE, University of Edinburgh.

*Family history for QIMR and NTR cases and controls is based on direct psychiatric interview assessment of family members; and was a screening criterion for controls. For NESDA and UoE cases family history is self-report.

For QIMR cases and controls standardized residuals from regression of arcsin-transformed scores on age and sex in a population sample of > 18,000. For NESDA/NTR cases, standardized scores across based on these cases and GAIN-MDD cases and controls. For NTR controls z-transformed scores based on a population sample of > 19,000.
or schizoaffective disorder, obsessive–compulsive disorder, bipolar disorder or severe substance use or dependence were excluded. All subjects provided written informed consent after study approval by the appropriate Institutional Review Boards.

University of Edinburgh. MDD cases were recruited through in- and out-patient services of psychiatric hospitals in Scotland and were tertiary referrals from primary care. All patients were interviewed by an experienced psychiatrist using the Schedule for Affective Disorders and Schizophrenia-Lifetime, supplemented by hospital case note review and information from informants. Final determination of MDD as the primary DSM-IV diagnosis was made by consensus of two psychiatrists. All cases had a lifetime history of recurrent MDD and IQ >70. The study was approved by the Central Office of Research Ethics Committees in Scotland and all subjects gave informed written consent for the collection of DNA samples for use in genetic studies.

Molecular Genetics of Schizophrenia controls. Briefly, random digit dialing was used to achieve a representative sample of individuals from the United States. Participants completed an online questionnaire including the short form Composite International Diagnostic Interview, supplemented by questions about schizophrenia, psychosis and bipolar disorder. Controls were required to never have met criteria for MDD. All subjects provided informed consent. These controls have been used in GWAS for multiple psychiatric disorders including MDD.

Genotyping and QC

Full details are given in the Supplementary File 1. All analysis sets were put through a common QC pipeline based on that used by the Psychiatric GWAS Consortium. SNPs were removed based on the following criteria: minor allele frequency (MAF) <0.01, Hardy–Weinberg equilibrium test \( P < 1 \times 10^{-6} \) in controls, missingness >0.01 or 0.02, missingness difference between cases and controls of 0.01 or 0.02, difference in frequency with HapMap of 0.07 or 0.15 (more stringent thresholds were applied when cases and controls were not genotyped concurrently, and were chosen to balance SNPs lost versus genomic control inflation factors). Samples were removed with missingness >0.02, if related to other samples or if identified as European ancestry outliers. Imputation was conducted in four analysis sets (I317, I370, I610, A6.0) in batches of ~300 mixed cases and controls to a common set of SNPs present in HapMap3 CEU/TSI, consisting of 410 haplotypes, using Beagle 3.04.

Statistical analyses

Association analysis was conducted on 2431 MDD cases and 3673 screened controls using allelic dosages for imputed SNPs passing QC. The test of association was logistic regression, including the first three ancestry principal components (PC) and analysis set as covariates. The INFO score (ratio of observed to expected dosage variance), a measure of the quality of imputation of SNPs was used to interpret results. A second association analysis was conducted restricting cases to those with recurrent early onset MDD (REO, age of MDD onset <31 years with multiple episodes of MDD). Separate analyses were conducted by sex, as different prevalences of MDD between males and females could imply existence of sex-specific genetic risk variants.

Statistical power

Detailed power calculations are provided in the Supplementary File 1. The MDD2000+ sample affords >90% power to detect an associated variant with MAF 0.36 and genotype relative risk (GRR) 1.33 (the median values across all published complex disease/trait GWAS, with \( P < 5 \times 10^{-8} \)).

Comparison of association results with other studies including meta-analysis

We compared our results with those published for other GWAS of MDD. In particular, we conducted a formal genome-wide meta-analysis for autosomal SNPs from MDD2000+, GAIN–MDD12 and UK15 studies (that is, the three largest MDD GWAS). The GAIN–MDD sample, after removing individuals related to Dutch cases or controls of MDD2000+, comprised 1696 MDD cases and 1634 screened controls and the UK sample comprised 1636 MDD cases and 1594 screened controls. Meta-analysis30 on the logistic regression results from each study, restricting the imputed SNPs to those with INFO scores between 0.8 and 1.1.

Gene-based tests

To determine whether any genes harbored an excess of SNPs with small \( P \)-values, we undertook genome-wide gene-based tests which account for both gene length and linkage disequilibrium between SNPs using VEGAS. SNPs were allocated to one or more autosomal genes using gene boundaries ± 50 kb. We investigated 183 candidate genes for MDD.

Results

Of 1,251,157 imputed SNPs, 1,079,979 (86%) had MAF>0.01 and INFO>0.8. The quantile–quantile plots of the observed versus expected \(-\log(P)\) from the four (total, male, female and REO) association analyses are presented in Figure 1 and Supplementary File 1. The genomic control \( \lambda \) (the median \( \chi^2 \) association statistic divided by the median expected under the null) shows no evidence for inflation of the test statistics. For example, for the full analysis, \( \lambda = 1.05 \) (and standardized to a sample size of 1000, \( \lambda_{1000} = 1.02 \)). There were no genome-wide significant results with \( P < 5 \times 10^{-8} \). The Manhattan plot of association \( P \)-values for all subjects is presented in the Supplementary File 1 and Table 3 shows associations with \( < 1 \times 10^{-5} \) in any analysis.
Regional association and forest plots are provided in Supplementary File 2. SNPs associated with $P < 0.001$ in any analysis are provided in Supplementary File 3.

**Comparison of association results with other studies including meta-analysis**

We compared the MDD2000+ results with other published studies. Muglia et al,\(^{11}\) listed 27 SNPs with $P < 1 \times 10^{-3}$; 25 SNPs overlapped with our analysis but none had $P < 0.05$ and association with the same allele. Of the top 200 SNPs in one or both of their study samples, ~90% were analysed in our study but ~5% had $P < 0.05$, consistent with chance expectations. As the MDD2000+ controls were also used in GenRed\(^{13}\) and STARD,\(^{14}\) we did not make formal comparisons of our results, but we note that ~5% of their top SNPs ($P < 1 \times 10^{-3}$) had $P < 0.05$ in MDD2000+. The SNP rs2251219 identified as a mood disorder risk factor in a GWAS meta-analysis (three bipolar studies plus the GAIN–MDD study)\(^{12}\) but was not associated in MDD2000+ ($P = 0.51$, MAF = 0.40, odds ratio (OR) = 0.97, 95% CI 0.90–1.05). PCLO was the top finding in the GAIN–MDD study,\(^{12}\) replicated in Australian (QIMR)\(^{12}\) and Dutch\(^{33}\) samples, but we found no evidence for association with PCLO variants (for example, $P = 0.51$ for rs2522833), despite partial overlap (562 cases and 264 controls) in the QIMR samples used here and in the original replication sample (see Supplementary File 1 for details). Meta-analysis results of the three largest MDD GWAS—427362 SNPs in MDD2000+, GAIN–MDD and the UK studies—yielded $\lambda_{t003} = 1.01$ (Figure 1iii). In all, 19 SNPs gave $P < 5 \times 10^{-5}$ compared with ~21 expected under the null hypothesis (ignoring linkage disequilibrium) and are listed in the Supplementary File 1.

**Gene-based tests**

MDD2000+ SNPs mapped to 18454 genes. We list all genes with $P < 1 \times 10^{-4}$ and all lie within linkage regions identified in the SLEP\(^{24}\) database of research findings for psychiatric disorders. Gene ontology annotations and names are provided for these genes (Table 4). Three are plausible candidate genes for MDD (GAL, ADCY3, PDK4) and others have a role in ion binding (ZDHHC19, ZNF83). None of the genes listed in Table 4 shows any evidence for association in the gene-based test applied to GAIN–MDD or the UK MDD GWAS results, although association at $P < 0.001$ is retained for GAL in the gene-based test applied to the meta-analysis results.

Of the 183 candidate genes investigated in other GWAS for MDD, we could test for association with 180 (Supplementary File 4). None were associated with $P < 0.00028$ (0.05/180). The six most associated were IL10, OPRM1, HTT, HTR1B, GRIN1 and CACNA1C. Of these, OPRM1 was reported to have $P < 0.05$ in the GAIN–MDD gene-based test and CACNA1C was top ranked in the GenRed\(^{14}\) study; the other studies all used a gene test based on best single associated SNP corrected for gene length. In this study, CACNA1C ranked 808 out of 18454 (4th percentile). Our top associated SNP in CACNA1C is rs98545 ($P = 0.0019$, OR = 0.83, MAF = 0.15) which is 387 kb from, and in near linkage equilibrium ($r^2 = 0.06$) with, rs1006737 ($P = 0.020$, OR = 1.10, MAF = 0.35), the SNP first identified in genome-wide association studies of bipolar disorder;\(^{32}\) an interaction between these SNPs was not significant ($P = 0.21$). The 3rd ranked of the candidate genes (Huntingtin, HTT) was included in the candidate list because mood disorders are characteristic of presymptomatic carriers of the Huntington’s disease trinucleotide repeat polymorphisms.\(^{21}\) HTT ranked in the top 0.7% of all genes; the SNP showing highest individual association within the gene is rs363099 (also known as rs4690074, $P = 4.7 \times 10^{-4}$, OR = 0.86, MAF = 0.31) located in exon 29, 85 kb from the exon 1 CAG repeat.

**Discussion**

**Study results**

The MDD2000+ study is the largest GWAS for MDD reported to date, comprising 2431 cases with MDD and 3673 screened controls. Our analysis was of >1M SNPs but none achieved genome-wide significance. Indeed, the quantile–quantile plots (Figure 1) show that the distribution of observed associations closely follows that expected under the null hypothesis. These results are consistent with other GWAS of MDD.\(^{11–15}\) Furthermore, comparison of our results with these other studies including formal meta-analysis of results from MDD2000+ and the next two largest studies\(^{12,15}\) (a total of 5763 cases and 6901
Table 3 Regions containing at least one SNP with a P-value of \( P < 10^{-5} \) in any one of the four association analyses—total sample (all), males only, females only and recurrent early onset (REO) MDD ordered by chromosomal position

<table>
<thead>
<tr>
<th>Association</th>
<th>SNP</th>
<th>Chromosome band</th>
<th>Position</th>
<th>MAF</th>
<th>Minor allele</th>
<th>Other allele</th>
<th>All Odds</th>
<th>Male Odds</th>
<th>Female Odds</th>
<th>REO Odds</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>rs182358</td>
<td>1p21</td>
<td>97235738</td>
<td>0.50</td>
<td>C</td>
<td>T</td>
<td>0.83</td>
<td>9.9E-06</td>
<td>0.78</td>
<td>3.3E-04</td>
<td>0.86</td>
</tr>
<tr>
<td>All</td>
<td>rs7551221</td>
<td>1q23</td>
<td>160267022</td>
<td>0.21</td>
<td>T</td>
<td>C</td>
<td>1.25</td>
<td>5.8E-06</td>
<td>1.13</td>
<td>1.2E-01</td>
<td>1.31</td>
</tr>
<tr>
<td>All</td>
<td>rs2384061</td>
<td>2p23.3</td>
<td>24989124</td>
<td>0.42</td>
<td>A</td>
<td>G</td>
<td>1.20</td>
<td>6.2E-06</td>
<td>1.24</td>
<td>7.0E-04</td>
<td>1.18</td>
</tr>
<tr>
<td>All</td>
<td>rs3732293</td>
<td>2p13.1</td>
<td>74503442</td>
<td>0.08</td>
<td>C</td>
<td>A</td>
<td>1.41</td>
<td>1.5E-06</td>
<td>1.45</td>
<td>1.0E-03</td>
<td>1.38</td>
</tr>
<tr>
<td>All</td>
<td>rs8052848</td>
<td>6p21.3</td>
<td>31790008</td>
<td>0.05</td>
<td>A</td>
<td>G</td>
<td>1.49</td>
<td>8.2E-06</td>
<td>1.32</td>
<td>4.3E-02</td>
<td>1.62</td>
</tr>
<tr>
<td>All</td>
<td>rs10815165</td>
<td>9p24.1</td>
<td>7475343</td>
<td>0.46</td>
<td>G</td>
<td>A</td>
<td>0.84</td>
<td>6.3E-06</td>
<td>0.90</td>
<td>7.9E-02</td>
<td>0.78</td>
</tr>
<tr>
<td>All</td>
<td>rs2031616</td>
<td>10q23–q33</td>
<td>94726947</td>
<td>0.39</td>
<td>A</td>
<td>G</td>
<td>0.84</td>
<td>9.1E-06</td>
<td>0.79</td>
<td>2.7E-04</td>
<td>0.86</td>
</tr>
<tr>
<td>All</td>
<td>rs2636463</td>
<td>12q22–23</td>
<td>88193916</td>
<td>0.31</td>
<td>A</td>
<td>G</td>
<td>1.22</td>
<td>2.9E-06</td>
<td>1.22</td>
<td>2.9E-03</td>
<td>1.21</td>
</tr>
<tr>
<td>All</td>
<td>rs17226852</td>
<td>20p12.1</td>
<td>13921760</td>
<td>0.08</td>
<td>C</td>
<td>T</td>
<td>1.41</td>
<td>1.5E-06</td>
<td>1.52</td>
<td>2.6E-04</td>
<td>1.36</td>
</tr>
</tbody>
</table>

Abbreviations: LD, linkage disequilibrium; MAF, minor allele frequency; MDD, major depressive disorder; SNP, single-nucleotide polymorphism.

For SNPs with INFO > 0.8. If multiple SNPs in high LD qualified for inclusion in the table, the one with the smallest P is listed. Regional association plots are provide for all regions with a SNP \( P < 10^{-5} \) in Supplementary File 2.
<table>
<thead>
<tr>
<th>Rank</th>
<th>Band</th>
<th>Gene ID</th>
<th>No. of SNPs</th>
<th>Length (kb)</th>
<th>Start (bp) of gene</th>
<th>Stop (bp) of gene</th>
<th>Gene test P SNP with minimum P-value</th>
<th>SNP stats</th>
<th>Gene name</th>
<th>Molecular function from GO annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2p23.3</td>
<td>ADCY3</td>
<td>a. 84</td>
<td>100</td>
<td>24 895 554</td>
<td>24 995 559</td>
<td>rs2384061</td>
<td>6.3E–06</td>
<td>Adenylate cyclase 3</td>
<td>Nucleotide binding; magnesium ion binding; calmodulin binding; ATP binding; calcium- and calmodulin-responsive adenylate cyclase activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 169</td>
<td></td>
<td>c. 29</td>
<td>d. 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5q35</td>
<td>NPM1</td>
<td>a. 52</td>
<td>23</td>
<td>170 747 402</td>
<td>170 770 493</td>
<td>rs1134697</td>
<td>2.7E–04</td>
<td>Nucleophosmin</td>
<td>Nucleic acid binding; DNA binding; RNA binding; protein binding; rRNA binding; ribosomal large subunit binding; ribosomal small subunit binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 52</td>
<td></td>
<td>c. 16</td>
<td>d. 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6q27</td>
<td>UNC93A²</td>
<td>a. 95</td>
<td>25</td>
<td>167 624 890</td>
<td>167 649 491</td>
<td>rs2076008</td>
<td>1.3E–04</td>
<td>Unc-93 homolog A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 147</td>
<td></td>
<td>c. 31</td>
<td>d. 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6q27</td>
<td>TTI12²</td>
<td>a. 84</td>
<td>18</td>
<td>167 658 563</td>
<td>167 676 167</td>
<td>rs2076008</td>
<td>1.3E–04</td>
<td>Tubulin tyrosine ligase-like family; member 2</td>
<td>Catalytic activity; tubulin-tyrosine ligase activity; ATP binding; ligase activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 131</td>
<td></td>
<td>c. 27</td>
<td>d. 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7q21.3</td>
<td>PDK4⁴</td>
<td>a. 77</td>
<td>13</td>
<td>95 050 744</td>
<td>95 063 861</td>
<td>rs1535170</td>
<td>3.4E–05</td>
<td>Pyruvate dehydrogenase kinase, isozyme 4</td>
<td>Two-component sensor activity; nucleotide binding; pyruvate dehydrogenase (acetyl transferring) kinase activity; ATP-binding; transferase activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 133</td>
<td></td>
<td>c. 29</td>
<td>d. 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>7q21–q22</td>
<td>ASB4²</td>
<td>a. 104</td>
<td>52</td>
<td>94 953 219</td>
<td>95 005 007</td>
<td>rs1535170</td>
<td>3.4E–05</td>
<td>Ankyrin repeat and SOCS box-containing 4</td>
<td>SOCS = suppressor of cytokine signalling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 182</td>
<td></td>
<td>c. 46</td>
<td>d. 46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9q31.1</td>
<td>TEX10</td>
<td>a. 51</td>
<td>51</td>
<td>102 104 190</td>
<td>102 154 995</td>
<td>rs1930243</td>
<td>2.7E–04</td>
<td>Testis expressed 10</td>
<td>Binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 90</td>
<td></td>
<td>c. 17</td>
<td>d. 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>11q13.3</td>
<td>GAL</td>
<td>a. 48</td>
<td>7</td>
<td>68 208 558</td>
<td>68 215 219</td>
<td>rs2156464</td>
<td>2.7E–05</td>
<td>Galanin</td>
<td>Neuropeptide hormone activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 52</td>
<td></td>
<td>c. 11</td>
<td>d. 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>15q22.3</td>
<td>USP3</td>
<td>a. 74</td>
<td>87</td>
<td>61 583 862</td>
<td>61 670 716</td>
<td>rs7183892</td>
<td>4.5E–04</td>
<td>Ubiquitin-specific peptidase 3</td>
<td>Ubiquitin thiolesterase activity; ubiquitin-specific protease activity; peptidase activity; cysteine-type peptidase activity; zinc ion binding; metal ion binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 121</td>
<td></td>
<td>c. 23</td>
<td>d. 23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20p12.1</td>
<td>SEL1L2</td>
<td>=</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. 98</td>
<td>141</td>
<td>13 778 049</td>
<td>13 919 262</td>
<td>rs17226852</td>
<td>1.5E–06</td>
<td>Sel-1 suppressor of lin-12-like 2</td>
<td>Binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. 209</td>
<td></td>
<td>c. 46</td>
<td>d. 42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: ATP, adenosine triphosphate; GO, gene ontology; MAF, minor allele frequency; MDD, major depressive disorders; SNP, single-nucleotide polymorphism; rRNA, ribosomal RNA; a, MDD2000+; b, GAIN-MDD; c, UK GWAS; d, meta-analysis. *The ± 50 kb regions of UNC93A and TTI12 overlap. **The ± 50 kb regions of PDK4 and ASB4 overlap. P < 1 × 10⁻² in bold. Top associated SNPs lie in region ± 50 kb from the listed start and stop positions. Regional association plots for ADCY3, GAL and CACNA1C are in Supplementary File 2. SNP stats: P-value, minor allele, MAF, odds ratio.
controls) showed little evidence for replication between studies.

To determine whether any genes harbored an excess of low-associated variants, we undertook gene-based tests which account for both gene length and linkage disequilibrium between SNPs. Our top associated genes were not replicated in the gene-based test when applied to the GAIN–MDD or UK MDD results, although some genes retained low association \( P \)-values when applied to the meta-analysis results (Table 4). Despite this there are some suggestive results worthy of note. Among the top ranked genes is \( GAL \), which encodes the neuropeptide galanin, proposed by Weiss et al.\(^\text{36} \) as having an important role in MDD. Their hypothesis, based on animal models is that \( GAL \) is a regulator of brain serotonin and 5-HT1A receptor-mediated transmission,\(^\text{37} \) agonists of \( GAL \) receptors have been proposed as potential drug targets for MDD.\(^\text{38} \) \( GAL \) also has an important role in the hippocampal processing of cognition.\(^\text{39} \) Our most associated SNP in \( GAL \) (rs2156464, \( P = 2.7 \times 10^{-5} \), OR = 1.24, CI = 1.12–1.37, MAF = 0.19) lies in the same haplotype block as an association reported for panic disorder.\(^\text{40} \) Another top ranked gene (\( ADCY3 \)) encodes the enzyme adenylate cyclase 3 that catalyses synthesis of cyclic adenosine monophosphate, the association underpinned by SNP rs2384061 (\( P = 6.3 \times 10^{-6} \), OR = 1.20, CI = 1.11–1.29, MAF = 0.42, Table 3). \( ADCY3 \) is an plausible candidate gene because depressed patients display reduction in platelet \( ADCY \) activity.\(^\text{41} \) Perhaps most interesting is that \( ADCY \) and \( GAL \) activity are inter-related\(^\text{42} \) and \( ADCY3 \) binds \( CACNA1C \) (calcium channel, voltage dependent, L type, \( \alpha-1C \) subunit).\(^\text{43} \) A SNP (rs1006737, intron 3) within the gene \( CACNA1C \) is associated with bipolar disorder\(^\text{44} \) schizophrenia and MDD.\(^\text{45} \) In this study, \( CACNA1C \) ranked in the top 4% of genes and we replicated the association with rs1006737 (\( P = 0.020 \), OR = 1.10, CI = 1.01–1.19, MAF = 0.34), but the top associated SNP in \( CACNA1C \) was rs98545 (\( P = 0.0019 \), OR = 0.83, CI = 0.75–0.94, MAF = 0.15, intron 29). rs98545 lies 387 kb from rs1006737; they are in near linkage equilibrium (\( r^2 = 0.06 \)) and in our data their effects are additive in a logistic regression. Neither rs1006737 nor rs98545 were genotyped in the UK MDD or GAIN–MDD studies; imputation of GAIN–MDD genotypes showed weak association with rs1006737 (\( P = 0.070 \), OR = 1.10, CI = 0.99–1.21, MAF = 0.32) but not with rs98545 (\( P = 0.52 \)). We cannot exclude that the association of rs98545 has occurred by chance in our sample. Regional association plots for \( ADCY3 \), \( GAL \) and \( CACNA1C \) can be found in Supplementary File 2.

Despite the large sample size of MDD2000+, a number of limitations may reduce the realized power for detection of association. First, our study combines data sets genotyped on different primary platforms, and the power of our sample will vary between SNPs depending on the linkage disequilibrium structure between genotyped and imputed SNPs. Second, some of our cases and controls were genotyped separately because of economic constraints. We were fortunate that we had >300 QIMR cases genotyped on both the Affymetrix and Illumina platforms, which we used extensively for QC. Third, when combining cases and controls genotyped as separate sets, it was necessary to impose more stringent QC constraints than for cases and controls genotyped on the same platform. These constraints were successful in removing artefactual differences between the case–control sets, but could also have eliminated true associations. Fourth, in MDD2000+ with the aim of maximizing sample size we have combined the clinical cases from the University of Edinburgh cohort with the community samples from QIMR and NTR. The extent to which this is important depends on the unknown genetic etiology of MDD. Lastly, 26% of both cases and controls contributed by the QIMR sample (which makes the largest contribution in this study) were from families ascertained on the basis of a sibling with nicotine dependence, as reflected by the high proportion of smokers in the QIMR samples. Smoking is associated with depression and if some genes are involved in smoking behavior and MDD, then the power to detect genetic variants for MDD may be reduced.

Implications

Our analysis and meta-analysis represent the largest and most powerful investigation into the genetic architecture of MDD to date. The lack of clear-cut evidence for association allows us to exclude genetic main effects that are common (MAF > 0.1) with moderate to strong effect sizes (GRR > 1.4). As noted earlier, the MDD2000+ sample alone possessed >90% power to detect the median MAF (0.36) and GRR (1.33) of GWAS associations published to date with \( P < 5 \times 10^{-8} \). Moreover, as we would expect to detect >20% of variants with OR 1.2 and MAF 0.2–0.7 (Supplementary File 1), and having detected none that are genome-wide significant, we can probably draw an even more stringent boundary on the risk allele architecture under a main effect model. These results are informative for researchers planning association studies of MDD.

These results are unsurprising in the light of results for a range of complex genetic disorders published since this study was conceived in 2007.\(^\text{46} \) Results from GWAS of psychiatric disorders\(^\text{47} \) have, on the whole, yielded replicated associations only for low prevalence/high heritability disorders, that is, autism, bipolar disorder and schizophrenia, with tobacco dependence being the exception.\(^\text{48} \) As a high prevalence/low heritability disorder, the challenge for MDD was always going to be harder. From Yang et al.,\(^\text{49} \) we estimate that sample sizes 2.4-fold greater are needed for GWAS of MDD (prevalence 0.15) compared with schizophrenia (prevalence 0.007) to...
identify a variant that explains the same proportion of phenotypic variance to liability of these disorders (Supplementary File 1). This result reflects the smaller mean difference in phenotypic liability between cases and controls for MDD compared with schizophrenia. Hospital-based MDD cohorts may represent a more extreme phenotype, with lower prevalence and higher heritability. Using a prevalence for such clinical samples to be 0.06 (the average across sexes) still requires a sample size ~1.8 times greater for a case–control study of MDD compared with one for schizophrenia.

Further, for main effect models, as the heritability on the liability scale \( h^2 \) is a function of the number of variants \((n)\), their frequencies \((q_i)\) and their effect sizes \((a_i)\),

\[
h^2 = 2 \sum_{i=1}^{n} q_i (1 - q_i) a_i^2
\]  

(with total variance of liability of 1), then the lower heritability of MDD (~0.37, although this may be higher in clinical samples) compared to schizophrenia (~0.81), must be explained by fewer risk alleles, lower risk allele frequencies and/or smaller effect sizes. It seems most plausible that MDD might have smaller effect sizes compared with schizophrenia (except in the unlikely event that the number of loci for MDD is substantially fewer than for schizophrenia). We estimate that sample sizes 4–5 times those needed for schizophrenia studies may be needed for MDD to detect variants that explain an equal proportion of the known genetic variance (Supplementary File 1).

**Study designs**

It is clear that larger sample sizes are needed to achieve the power required to detect common variants of smaller effect, and that larger sample sizes will be needed for MDD compared with less prevalent but more heritable disorders. The prize is major: any replicated hit provides a priceless window into the etiology of this idiopathic disorder. However, alternative and complementary strategies are possible, particularly those that consider a broader range of genetic models and relax further assumptions of etiological homogeneity.

First, recognizing that effect size is the average effect of a variant across all genetic and environmental backgrounds, it may be possible to forgo larger sample sizes to concentrate on more homogeneous subsets in which larger effect sizes may be detectable. There has been much debate about the genetic heterogeneity of MDD. Some genetic heterogeneity is consistent with the polygenic model; each affected individual could potentially harbor a different combination of risk variants which could imply a spectrum of phenotypic symptoms, with those presenting similar symptom profiles carrying more similar profiles of risk alleles, consistent with symptom sharing of closely related individuals. For this reason, most GWAS for MDD have prioritized genotyping of the less prevalent and more heritable recurrent early onset MDD, but even so genome-wide significant associations have been elusive. Stratification by sex has also not yielded consistent results. Other possibilities certainly exist but phenotypes may be unavailable (for example, accurate delineation of typical versus atypical MDD) or prohibitively expensive (for example, magnetic resonance imaging). Use of quantitative scores of severity and reliability may be the best strategy to balance sample size with phenotype definition, as MDD symptom profiles are consistent with a quantitative rather than dichotomous liability.

Second, in addition to achieving a more homogenous genetic background, larger estimated effect sizes may come from selecting more homogenous environmental backgrounds. The low heritability of liability for MDD implies an important role for environmental risk factors. Although genotype × environment interaction cannot explain the so-called ‘missing heritability’, it can contribute to small effect sizes. Although genotype × environment studies are conceptually attractive, the lessons learned from the most studied genotype × environment hypothesis for MDD (5HTTLPR and stressful life event) are sobering. However, this broad conclusion has been challenged with Uher and McGuffin arguing that replication has been achieved when stressful life event were recorded objectively, temporally and before the onset to MDD. More mileage might be gained by considering MDD in the context of more clearly defined exposures, such as pregnancy, in which women with perinatal and post-partum MDD have been exposed to a similar event.

Whichever way we look at it, and whether risk variants are common or rare, it seems that the challenge for MDD will be much harder than for the less prevalent more heritable psychiatric disorders. Larger samples are required whether we attempt to identify associated variants with small effect across average backgrounds or attempt to enhance detectable effects sizes by selection of homogeneity of genetic or environmental background. In the long-term, a greater understanding of the etiology of MDD will require large prospective, longitudinal, uniformly and broadly phenotyped and genotyped cohorts that allow the joint dissection of the genetic and environmental factors underlying MDD.

**Conflict of interest**

The authors declare no conflict of interest.

**Acknowledgments**

We thank the twins and their families registered at the Australian Twin Registry for their participation in the many studies that have contributed to this research. We thank Dixie Statham (sample collection); Leanne Wallace, Anthony Caracella and staff of the Molecular Epidemiology Laboratory (DNA processing); David...
Smyth, Harry Beeby, and Daniel Park (IT support). Funding was provided by the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496675, 496739, 552485, 552498, 613608), the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254), and the US National Institutes of Health (NIH grants AA07535, AA10248, AA13320, AA13321, AA13326, AA14041, MH66206, DA12854, DA019951). A portion of the genotyping on which this study was based (Illumina 370K scans on 4300 individuals) was carried out at the Center for Inherited Disease Research, Baltimore through an access award to our late colleague Dr Richard Todd (Psychiatry, Washington University School of Medicine, St Louis). Statistical analyses were partly conducted at the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003). SEM, SM and GWM are supported by the National Health and Medical Research Council Fellowship Scheme. NRW and DRN are supported by the Australian Research Council Future Fellowship Scheme.

Funding support was provided by the Netherlands Scientific Organization (904-61-090, 904-61-193, 480-04-004, 400-05-717, Centre for Medical Systems Biology (NWO Genomics), the Neuroscience Campus Amsterdam and the EMGO + institute; the European Union (EU/WLRT-2001-01254), ZonMW (Geestkracht program, 10-000-1002), NIMH (RO1 MH059160) and matching funds from participating institutes in NESDA and NTR. The NTR controls were genotyped in the Genomics platform (certified service provider (CSPro(R)) for Illumina) at the LIFE and BRAIN Center Bonn (funded by NWO-SPI 56-464-1419).

We would like to record our appreciation for much practical support from our late colleague Walter J Muir, Professor of Developmental Psychiatry at the University of Edinburgh and especially for his many stimulating ideas for genetic studies in depression. We thank Margaret Van Beck for sample collection and Lee Murphy, Wellcome Trust Clinical Research Facility, Western General Hospital, Crewe Road South, Edinburgh, for sample management. The collection of the Edinburgh cohort was supported by grants from The Wellcome Trust, London and the Chief Scientist Office of the Scottish Government.

Additional funding was provided by National Institute of Mental Health Schizophrenia (NIMH) grant MH080403 to EJCGvdO and PFS. Control subjects from the NIMH Genetics Initiative, data and biomaterials are being collected by the ‘Molecular Genetics of Schizophrenia II’ collaboration. The investigators and coinvestigators are: ENH/Northwestern University, Evanston, IL, MH059571, Pablo V Gejman, MD (Collaboration Coordinator; PI), Alan R Sanders, MD; Emory University School of Medicine, Atlanta, GA, MH59587, Farooq Amin, MD (PI); Louisiana State University Health Sciences Center; New Orleans, Louisiana, MH067257, Nancy Buccola APRN, BC, MSN (PI); University of California-Irvine, Irvine, CA, MH60870, William Byerley, MD (PI); Washington University, St Louis, MO, U01, MH060879, C Robert Cloninger, MD (PI); University of Iowa, Iowa, IA, MH59566, Raymond Crowe, MD (PI), Donald Black, MD; University of Colorado, Denver, CO, MH59565, Robert Freedman, MD (PI); University of Pennsylvania, Philadelphia, PA, MH61675, Douglas Levinson, MD (PI); University of Queensland, Queensland, Australia, MH59588, Bryan Mowry, MD (PI); Mt. Sinai School of Medicine, New York, NY, MH59586, Jeremy Silverman, PhD (PI).

Author contributions


Analysis Set QC: DRN, SDG, NRW, BPMcE, SEM Joint analysis set QC, imputation, association analysis and regional plots of MDD2000+:: SR Post-association analysis analyses: NRW, SDG, EMB, JZL, SM

Manuscript preparation: NRW, PFS and all other authors

References

1 Australian Bureau of Statistics 4326.0—National Survey of Mental Health and Wellbeing: Summary of Results, 2007 http://www.abs.gov.au/AUSSTATS/abs@nsf/nf/43260


30 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, 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.


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