ORIGINAL ARTICLE

708 Common and 2010 rare DISC1 locus variants identified in 1542 subjects: analysis for association with psychiatric disorder and cognitive traits

PA Thomson1,2,9, JS Parla3,9, AF McRae4,9, M Kramer5, K Ramakrishnan1, J Yao6, DC Soares1, S McCarthy3, SW Morris1, L Cardone3, S Cass1, E Ghibani3, W Hennah1,5, KL Evans1,2, D Rebolini3, JK Millar7, SE Harris1,2, JM Starr2, DJ MacIntyre6, Generation Scotland2, AM McIntosh6, JD Watson3, IJ Deary2, PM Visscher4,8, DH Blackwood6, WR McCombie3 and DJ Porteous1,2

A balanced t(1;11) translocation that transects the Disrupted in schizophrenia 1 (DISC1) gene shows genome-wide significant linkage for schizophrenia and recurrent major depressive disorder (rMDD) in a single large Scottish family, but genome-wide and exome sequencing-based association studies have not supported a role for DISC1 in psychiatric illness. To explore DISC1 in more detail, we sequenced 528 kb of the DISC1 locus in 653 cases and 889 controls. We report 2718 validated single-nucleotide polymorphisms (SNPs) of which 2010 have a minor allele frequency of <1%. Only 38% of these variants are reported in the 1000 Genomes Project European subset. This suggests that many DISC1 SNPs remain undiscovered and are essentially private. Rare coding variants identified exclusively in patients were found in likely functional protein domains. Significant region-wide association was observed between rs16856199 and rMDD (P = 0.026, unadjusted P = 6.3 × 10^{-9}, OR = 3.48). This was not replicated in additional recurrent major depression samples (replication P = 0.11). Combined analysis of both the original and replication set supported the original association (P = 0.0058, OR = 1.46). Evidence for segregation of this variant with disease in families was limited to those of rMDD individuals referred from primary care. Burden analysis for coding and non-coding variants gave nominal associations with diagnosis and measures of mood and cognition. Together, these observations are likely to generalise to other candidate genes for major mental illness and may thus provide guidelines for the design of future studies.

Molecular Psychiatry advance online publication, 4 June 2013; doi:10.1038/mp.2013.68

Keywords: DISC1; recurrent major depressive disorder; sequencing

INTRODUCTION

Schizophrenia (SZ), bipolar disorder (BD) and recurrent major depressive disorder (rMDD) are common forms of serious mental illness, each with a strong and overlapping genetic component.1–3 Genome-wide linkage, association, cytogenetic, copy number variant and, more recently, sequencing studies establish that the genetic architecture of psychiatric illness is complex and that there is extensive genetic heterogeneity, which is incompletely defined or understood (reviewed in Sullivan et al.9). We previously reported a t(1;11) translocation in a single large Scottish family that showed genome-wide significant linkage for SZ, rMDD and jointly with BD.5 The t(1;11) translocation is balanced and structurally simple, but the outcome is genetically complex, disrupting the protein coding gene Disrupted in schizophrenia 1 (DISC1), the antisense non-coding gene Disrupted in schizophrenia 2 (DISC2) and the non-coding gene DISC1FP1, creating a DISC1/DISC1FP1 fusion transcript.6–11 Several small independent studies have reported evidence for association of single DISC1 single-nucleotide polymorphisms (SNPs) (coding and non-coding) or haplotypes with SZ, BD, rMDD and other neuropsychiatric traits, including autism spectrum disorder, cognition, normative cognitive ageing, anxiety and structural and functional brain imaging phenotypes.9,12,13 Rare amino-acid substitution variants in DISC1 have been reported in cases of SZ,10,11 BD,14 rMDD,15 autism spectrum disorder15 and agenesis of the corpus callosum,16 as has an increased burden of rare missense variants in exon 11 of DISC1 for schizoaffective disorder,17 and for DISC1 pathway genes in SZ.10 In contrast, a meta-analysis of all known common variants within the DISC locus, from a total of 11,626 cases and 15,237 controls that involved the testing of 1241 SNPs, found no evidence that common variants at the DISC locus are significantly associated with SZ.18 Moreover, the DISC1 locus has not reached genome-wide significance in large-scale meta-analyses of linkage studies of SZ,19 nor have its common variants in large-scale genome-wide association studies of SZ, BD or rMDD.20–22 A recent exon-based study that sequenced 2.7 kb of DISC1 in 727 cases of SZ and 733 controls found 32 rare alleles (minor allele frequency (MAF) <0.01)
in SZ cases and 40 in European controls with no evidence for a significantly increased burden of likely pathogenic variants.\textsuperscript{23} DISC1, however, continues to feature strongly in attempts to assess genome-wide association results in terms of networks\textsuperscript{24} or in combination with known biological function.\textsuperscript{25–27}

The biological functions of DISC1 fit well with current aetiological concepts in SZ-related major mental illness and cognition.\textsuperscript{28} DISC1 is a scaffold protein that interacts with, and modulates the activity of, multiple proteins with key roles in neurodevelopment, neurogenesis, neuronal migration, integration and signalling\textsuperscript{29–31} including the antidepressant and antipsychotic targets GSK3\textsuperscript{b} and PDE4.\textsuperscript{32} Several common and rare amino-acid polymorphisms of DISC1 have predicted deleterious effects on protein function and demonstrable biological effects in experimental settings.\textsuperscript{31} The 704C allele is associated with reduced activity of ERK1 and Akt kinases, altered binding affinities of DISC1 for NDE1 and NDE1 and variation in DISC1 oligomeric status.\textsuperscript{34–37} The 607F allele results in a reduced binding and centrosomal localisation of PCM1, (b) reduced noradrenaline neurotransmitter release in SH-SY5Y cells,\textsuperscript{38} (c) altered mitochondrial trafficking\textsuperscript{39} and (d) a partial shift from neuronal to glial expression in the brain.\textsuperscript{40} Furthermore, Singh et al.\textsuperscript{41} reported that 607F impacts negatively on neural progenitor proliferation in E16 mouse brain, correlates with aberrant wnt signalling in human lymphoblasts, and is associated with a neurodevelopmental phenotype in morpholino mutant zebrafish. R37W lies within an arginine-rich nuclear localisation motif and a partially overlapping functional domain for PDE4\textsuperscript{2} and GSK3\textsuperscript{b}.\textsuperscript{41,43} The 37W allele shows reduced nuclear DISC1 expression and altered DISC1 regulation of ATF4, a critical modulator of cAMP signalling and mediator of the stress response.\textsuperscript{44}

In summary, whereas the primary genetic evidence from the original Scottish t(1;11) family was significant at the genome-wide significance level for both SZ and rMDD and the experimental evidence and biological plausibility of DISC1 remains strong, the evidence from subsequent linkage and candidate gene association studies is however inconsistent and not supported by genome-wide association studies or meta-analysis.\textsuperscript{12} To explore these conflicting findings, we aimed here to establish the nature and frequency of DISC1 genomic sequence variants, identify rare variants in putative functional domains, and test for effects of these on cognitive traits and the risk of psychiatric illness. We comprehensively sequenced 528 kb covering the entire DISC1 locus, including TRAX (also known as TSNAX) for which there is evidence for intergenic splicing with DISC1\textsuperscript{45} and the intergenic region, which contains regulatory elements immediately 5' of DISC1.\textsuperscript{13,46–48}

\section*{MATERIALS AND METHODS}

A full summary of the methods can be found in Supplementary Information. Briefly, all study participants gave signed consent for their data and samples to be used in studies that have been approved by the appropriate Research Ethics Committee or the GS access Committee. Genomic DNA from each individual was whole genome amplified in triplicate, the products pooled and amplified with primer pairs tiled across 528 kb of TRAX/ DISC1 (hg18 chr1:229723339-230251606; hg19 chr1:231656716-232184983). For each sample, the pooled products were sheared, converted into paired-end Illumina libraries and sequenced on an Illumina GAII or HiSeq 2000 sequencer to >80% coverage and >30-fold depth. Sequences were aligned to the UCSC hg18 reference sequence, variants called using MAQ software\textsuperscript{49} and the variants in repeats removed. Ten per cent of all remaining variants were validated using Sanger sequencing chemistry on an ABI3730 sequencer, and the derived information used to optimise the quality control filters. After quality control screening, all exonic and low frequency (MAF <1\%) variants were also validated by Sanger chemistry sequencing as above. The variants were functionally annotated using SNPnexus\textsuperscript{50} (http://www.snp-nexus.org).

Non-coding variants were annotated using the UCSC table browser for the following tracks: ‘RepeateMasker’\textsuperscript{51} ‘CpG island’, ‘TFBS conserved’, ‘7x Reg Potential’ (which substantially overlaps with DNase hypersensitivity sites) and/or ‘28-Way Most conserved—PlacMammal’ (http://genome.ucsc.edu/). Sequence variants classified as coding were mapped to the DISC1 L isoform and potential pathogenicity ascribed using Pmut,\textsuperscript{52} Panther\textsuperscript{53} and PolyPhen-2.\textsuperscript{54} The coding sequence variants were mapped onto a list of known curated DISC1-interactor binding sites\textsuperscript{31} and with other functional elements (for example, phosphorylation sites\textsuperscript{55}). Case–control association was tested on the combined case samples as well as individually for SZ, BP and rMDD using Fisher’s exact test. Permutation was used to derive region-wide P-values and significance thresholds. Quantitative trait association analyses using LBC1936 samples were performed by linear regression of the trait residuals (adjusted for age and sex) on the number of minor alleles at each SNP, with empirical P-values estimated by permutation to avoid issues with the test statistic distribution caused by the combination of rare variants and slight deviations from normality in the phenotypes. All association analyses were performed using PLINK.\textsuperscript{56} Mark-recapture analysis followed the Lincoln-Petersen and Modified Petersen methods\textsuperscript{57,58} with 95\% confidence intervals calculated following Chapman.\textsuperscript{59} Burden analysis was performed in PLINK/SEQ to implement BURDEN and VTEST with empirical P-values estimated using permutation. Genotyping of the replication and familial samples was performed by the Edinburgh Wellcome Trust Clinical Research Facility Genetics Core using TaqMan SNP genotyping assay C_33950433_10 with concurrent genotyping of known heterozygotes.

\section*{RESULTS}

\subsection*{Sequence analysis}

We sequenced 1542 Caucasians from Scotland comprising 240 cases of SZ, 221 cases of BD, 192 cases of rMDD and, as controls, 889 members of the Lothian Birth Cohort 1936 (LBC1936), which have been extensively phenotyped.\textsuperscript{60} Each sample was sequenced to >80\% coverage at a minimum of 30-fold read depth by long-range PCR and sequencing on either Illumina GAII or HiSeq 2000 sequencers. To ensure a robust data set, all variants within repetitive regions were removed. Final quality score thresholds for the data were derived from capillary sequence validation of 10\% of the remaining variants. All variants with an MAF <1\% were validated by ABI3730 sequencing. After quality control, there was no evidence for sequencing bias between cases and controls (Supplementary Figure S1). Allele frequencies from our sample showed strong concordance to those from the European subset of the 1000 Genomes Project\textsuperscript{61} (Supplementary Figure S2). We report 2718 SNPs in the 1542 samples analysed, 708 at \( \geq 1\% \) and 2010 at \( <1\% \) MAF (Supplementary Table S1). Only 1027 of the 2718 SNPs (38\%) were previously reported in the European subset of the 1000 Genomes Project.\textsuperscript{61}

As defined and annotated by the UCSC genome browser (http:// genome.ucsc.edu), 489 SNPs mapped to regions of regulatory potential, 177 to non-coding exons (including DISC2) and 36 to coding regions of exons. Of these 36 variants, 12 were synonymous changes, 23 were non-synonymous changes, with one producing a stop codon consistent with the DISC1 Es isoform (Figure 1; Supplementary Figure S3; Supplementary Tables S1 and S2). Supplementary Table S3 summarises the overlap between variants identified in this study and other DISC1 sequencing studies and relevant association studies.\textsuperscript{10,11,13,14,16,17,23,46}

Association and segregation of common variants with psychiatric illness and related quantitative traits

Genome-wide association studies of SZ, BP and rMDD are most consistent with a polygenic liability for common variants, but they
The risk allele for rs16856199 did not segregate with rMDD in 10 families of carriers identified from the Generation Scotland replication sample (Supplementary Figure S5). This suggests that there is increased evidence for association of rs16856199 in the more severely affected individuals.

SNP rs16856199 is on the Affymetrix 6.0 array, but the best tagging SNP on the Illumina 660W-Quad, Human Hap, Human1M-Duo arrays is rs16856189. SNP rs16856189 has an $r^2$ of 0.27 with rs16856199, which may explain in part why this association has not been reported previously in genome-wide association studies. SNP rs6678723, which lies 2.1 kb distal to this SNP within intron 11, showed the most significant association of DISC1 sequence complexity on the Illumina Trio arrays.

The LBC1936 has quantitative measures of symptoms of anxiety, depression and the personality trait of neuroticism, plus psychometrically tested measures of cognitive ability (fluid (age sensitive) and crystallised (non-age sensitive)) and cognitive ageing, which have been shown to be highly heritable and polygenic. Association of these traits with DISC1 was tested by linear regression analyses, co-varied for age at testing and sex (Supplementary Figures S6–S8). There were no region-wide significant findings for any of these quantitative traits.

Estimating the net pool of DISC1 variants

To estimate the effective pool of common and rare sequence variants in the European population, we applied a ‘mark-recapture’ approach (see Supplementary methods) to our data and that of the 1000 Genomes Project after appropriate checks on read depth and Sanger sequence validation (Supplementary Table S5; Supplementary Figure 9).
The total number of DISC1 SNPs $>1\%$ MAF was estimated at 905 (95%CI = 905 ± 5), of which 901 (99.5%) are known (Supplementary Table S6). The number of rare variants ($<1\%$ MAF) is less confidently predicted, but is likely to be substantially higher (95%CI = 3777 ± 252) (Supplementary Table S6). Thus, despite the ~2500 European genomes in which the DISC1 locus has been completely sequenced and the 2305 rare DISC1 variants now known, ~40% or more remain to be discovered, and will be essentially 'private'.

Rare amino-acid substitutions
Of the 17 rare coding SNPs previously reported for DISC1,10,16,17,68,69 we identified 12 (70.6%) plus an additional 8 non-synonymous variants, of which 5 are also absent from the European samples of both the 1000 Genomes Project (v3.20101123)71 and the Exome Variant Server (NHBLI GO Exome Sequencing Project (ESP), Seattle, WA; http://evs.gs.washington.edu/EVS/; September, 2012), and previous DISC1 sequencing studies.10,11,13,14,16,17,23,46,70 (Figure 1; Supplementary Figure S3; Supplementary Tables S2 and S3).

Five variants, R37W, T453M, T603M, L607F and S704C, are predicted to be deleterious by all three applied prediction algorithms (PolyPhen-2, Pmut and Panther; Supplementary Table S2). From this set, the functional effects of the common variants L607F (rs6675281) and S704C (rs821616) have been well documented in the literature, as mentioned earlier. R37W lies within a defined nuclear localisation signal71 and PDE4B binding site72 and is seen in a single case of rMDD (discussed at the end of this section). T453M is present at low frequency in cases and control individuals, both in this study and others (Supplementary Table S3). T603M was only identified in a single control, but Song et al.11 reported a T603I variant in a schizophrenic individual that was absent in their set of control individuals (Supplementary Table S3).

Five variants were only observed in affected individuals from our study (R37W, A83V, W160L, R233K and R418H), but not in 889 control individuals. Four of these non-synonymous case-only singletons are located in the largest coding exon, DISC1 exon 2, and the remaining variant is in DISC1 exon 4 (Figure 1). A83V was seen in a single individual with BD; this variant is predicted to be deleterious by PolyPhen-2 and Pmut and has been shown to affect Wnt signalling.41 It was however observed at low frequency in controls and in individuals with partial agenesis of the corpus callosum in previous studies (Supplementary Table S3).11,14,16 Apart from R37W and A83V, none of the three remaining non-synonymous variants, W160L, R233K and R418H, are consistently predicted by the three prediction algorithms to have functional effects (Supplementary Table S2). 160L and 418H were detected in single SCZ individuals and have also previously been reported in individuals with SCZ; but 160L has also been detected in control individuals.10 The variant 233K has not been previously reported, and was identified in an individual with rMDD. No non-synonymous variants in TRAX were found in cases only. A single...
Evidence for familial segregation was sought for five rare exonic variants where additional family members were available, but none segregated perfectly or unequivocally with diagnosis (Supplementary Figure 9). Of note however was the identification of the non-synonymous amino-acid variant R37W (rs137948488), first reported\(^4\) in a subject with SZ, and seen here in a single case of rMDD. R37 is strictly conserved among orthologues and recent publications, including our own, have demonstrated biological effects of 37W on DISC1 interactions\(^1,2,7\) and shown a dominant-negative effect on the sub-cellular distribution of DISC1.\(^4\) Five additional family members of the 37W members were available for genotyping diagnosed with rMDD, generalised anxiety disorder, bipolar II or no psychiatric diagnosis at the time of assessment. The R37W mutation was present in relatives with rMDD and generalised anxiety disorder, but not in a relative with bipolar II, or any unaffected individual (Figure 3).

**Burden analysis for putative functional variants**

To explore the burden of SNPs of potential functional significance, all variants with MAF <1% were first validated by ABI3730 sequencing. There was no significant overall difference in the number of singleton variants (Supplementary Figure S10) or in the overall number of minor alleles by diagnosis (see Supplementary Methods). SNPs were classified on the basis of bioinformatic annotation into seven functional classes: those in exons including untranslated exons, coding sequence, non-synonymous coding SNPs, conserved regions, regions with regulatory potential, conserved transcription factor binding sites and increased conserved regions with MAF <0.18% (P = 0.022). Nominally significant association was found in the LBC1936 data between the burden of minor alleles across all frequencies for SNPs in conserved transcription factor binding sites and increased symptoms of depression, a measure of depressed mood at the time of testing (Table 1). In addition, a nominally significant increase in burden of minor alleles for SNPs in CpG islands or coding SNPs was observed with Moray House Test scores, measures of cognitive ability (Supplementary Table S9). Summaries of the nominally significant results are given in Table 1.

**DISCUSSION**

Diagnoses of SZ, BP and rMDD were all present in the original Scottish family carrying the translocation that disrupts DISC1. All three DSMIV diagnoses have a strong and overlapping genetic component, but robust statistical analysis of gene-level contributions to risk are complicated by extensive genetic heterogeneity within and between diagnoses.\(^4\) We have provided the most comprehensive landscape of genetic variation at the DISC1 locus to date in patients with this spectrum of psychiatric illness and in healthy population controls with quantitative measures of mood and cognition. Comparison between our sequencing study and that of the 1000 Genomes Project confirms that current genome-wide association studies effectively captures the majority of common (but not rare) variants in the European population. Our sample size is large by current sequencing study standards, but we lack power to detect genome-wide significant P-values for either common or rare variants (see Supplementary Information and Supplementary Figure S11 for further details and also Kiezun et al.).\(^7\) Indeed, the predicted abundance of independent rare variants at this (and any other given) locus makes it highly improbable that any one will contribute to illness in the population at a frequency that will be statistically significant, given the numbers of patients we can afford to analyse by direct sequencing.\(^7\)

We observed no evidence for association at the whole-genome level of statistical significance between individual rare or common variants and either psychiatric illness or cognition. This is consistent with recent findings,\(^7,17\) which suggest that much larger samples would be required to detect such associations. Burden analysis of multiple rare and/or deleterious putative functional variants also failed to show association with these traits. We do report both functional and putative regulatory variants that are both individually, and by functional classification, nominally associated with rMDD and/or cognitive ability at the locus-wide level of significance.

Our study identified a novel association between intronic SNP rs16856199 and rMDD in hospital-referral subjects. Segregation with diagnosis in the relatives of these probands corroborated the association, but further studies are required to understand the lack of replication in population-based cohorts with depression. This may be due to inherent differences between patients recruited from hospital-referral compared with those from population-based cohorts. Cohorts from primary care are more likely to have a family history of depression,\(^7,5\) and may have more physical and psychiatric comorbidity in general. Conversely, the population-based sample may have fewer, less severe episodic, and less hospital-based cohorts.\(^7\) However, given the modestly significant P-value for rMDD in the discovery cohort, the number of psychiatric traits examined and the lack of replication, it is possible that the observed association is due to chance.

The nominal associations of the burden of common (threshold MAF = 30.7%) and rare potentially regulatory variants (threshold MAF = 0.060%) to measures of cognitive ability merit further study. A yet-to-be-defined subset of these is likely to have critical roles is spatial and temporal regulation of transcription and splicing.\(^46,78\) This highlights the need for annotation tools with improved predictive value for non-coding variants.\(^79\)

More importantly, our study demonstrates that substantial coding and non-coding genetic variation at the DISC1 locus...
remains undiscovered. Despite sequencing over 1500 subjects, we have probably captured only ~40% of the extant DISC1 variants in just the European population. Crowley et al. sequenced 2.7 kb of DISC1 exons and 5’ and 3’ regulatory sequence in 1460 samples of European or African origin. We observed 13/38 (34%) of the variants genotyped in the replication phase, supporting the argument for an abundance of rare variants.

The level of sequence variation identified in our study is unlikely to be exceptional, and indeed is consistent with evidence emerging from other genome sequencing studies. Consequently, it will be challenging to demonstrate robust (replicated) association by statistical evidence alone in case–control studies, exceptionally so with the numbers of patients that are currently affordable for sequencing. The original t(1;11) family illustrates the added issue of variable penetrance and cross-boundary diagnosis for a given mutation: ~70% of carriers had SZ, BP or rMDD, but ~30% had no formal psychiatric diagnosis, yet t(1;11) carriers, including both affected and unaffected, had ERP P300 measures in the range typical of individuals with SZ. The original identification of 37W in a case of SZ and here in a case of rMDD (and two offspring, one with rMDD, the other unaffected) may suggest variable penetrance of this biologically functional variant.

Note that the 37W variant was not observed in 10 000 control individuals, the 1000 Genomes project, the NHLBI GO Exome Sequencing Project (ESP), nor any functional. Overall, these results demonstrate a high level of sequence variation in DISC1, a subset of which may contribute to psychiatric disorder in some individuals who will be typically rare in the population precluding classical statistical analysis and requiring biological validation. This predicts a population-specific contribution of rare casual variants to risk. Our results indicate the potential value of sequencing non-coding regions of the genome, which may harbour disease-associated regulatory variants. Our findings of both functional and putative regulatory variants nominally associated with depression and cognitive ability merit replication in independent samples and biological exploration.

CONFLICT OF INTEREST

WRM has participated in Illumina sponsored meetings over the past four years and received travel reimbursement and an honorarium for presenting at these events.

ACKNOWLEDGEMENTS

We acknowledge the contribution of Prof Walter M Muir (deceased) for the collection of the clinical samples used in this study. We thank Gail Davies, Dave Liewald, Lee Murphy, Alan Maclean, Pamela Linksted, Archie Campbell, Shona Kerr, Andrew McQuillan and Stuart Macgregor for help in preparing samples, data and statistical advice and Generation Scotland for access to GSS/FSHS samples and data. We thank the LBC1936 cohort members. We thank the LBC1936 study secretary Paula Davies and Alan Gow, Janie Corley and Caroline Brett, for cognitive data collection and data entry. We thank the nurses and staff at the WTCRF where the medical phenotypic entry. We thank the staff at the Lothian Health Board, and the staff at the SRE Centre, University of Glasgow. The LBC1936 research was supported by a programme grant from Research Into Ageing and continues with programme grants from Help the Aged/Research Into Ageing (Disconnected Mind). The research was done within the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, supported by the BBSRC, EPSRC, ESRC and MRC, as part of the cross-council Lifelong Health and Wellbeing Initiative (G0700704/84698). Generation Scotland receives core support from the Chief Scientist Office of the Scottish Government Health Directorates CZD/16/6/2 and the Scottish Funding Council. We are grateful to the GS Executive Committee Professors Andrew D Morris, Blair H Smith, Anna F Dominiczak, David J Porteous and Drs Lynne J Hocking and McQuillan and Stuart Macgregor for help in preparing this manuscript. This work was supported by a gift from T and V Stanley to CSHL, the UK Medical Research Council (G0902166), the Wellcome Trust (WT 008179), the Chief Scientists Office of the Scottish Government Department of Health (CZD/16/6/1), a Scottish Funding Council Strategic Research Development Grant award (HR03006) and a Lifelong Health and Wellbeing (G0700704/84698). AMcR and PMV acknowledge support from the Australian National Health and Medical Research Council.

Table 1. Summary of nominally significant burden results

<table>
<thead>
<tr>
<th>Case–control</th>
<th>Trait</th>
<th>SNP subset</th>
<th>Test</th>
<th>Average excess of variants in cases</th>
<th>Optimised MAF threshold (N. SNP)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rMDD</td>
<td></td>
<td>RegPot</td>
<td>BURDEN</td>
<td>1.39</td>
<td>---</td>
<td>0.0018 (31)</td>
</tr>
<tr>
<td>rMDD</td>
<td></td>
<td>PhastCon</td>
<td>VTTEST</td>
<td>1.20</td>
<td>---</td>
<td>0.044</td>
</tr>
<tr>
<td>Quantitative traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moray House Test at age 70</td>
<td>Symptoms of depression</td>
<td>TFBS</td>
<td>BURDEN</td>
<td>0.062</td>
<td>---</td>
<td>0.032</td>
</tr>
<tr>
<td>Moray House Test at age 11</td>
<td>Moray House Test at age 11 adjusted for the</td>
<td>PhastCon</td>
<td>VTTEST</td>
<td>0.074</td>
<td>0.00013 (21)</td>
<td>0.030</td>
</tr>
<tr>
<td>Moray House Test at age 11</td>
<td>Moray House Test at age 11</td>
<td>CpG</td>
<td>VTTEST</td>
<td>0.064</td>
<td>0.000060 (2)</td>
<td>0.047</td>
</tr>
<tr>
<td>Moray House Test at age 11</td>
<td>Moray House Test at age 11</td>
<td>Coding</td>
<td>VTTEST</td>
<td>0.101</td>
<td>0.000070 (12)</td>
<td>0.028</td>
</tr>
<tr>
<td>Moray House Test at age 11</td>
<td>Moray House Test at age 11</td>
<td>CpG</td>
<td>VTTEST</td>
<td>0.097</td>
<td>0.31 (3)</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Abbreviations: MAF, minor allele frequency; rMDD, recurrent major depressive disorder; SNPs, single-nucleotide polymorphisms.

See main text for trait descriptions and Supplementary Table S8 for results on all diagnoses and traits. Quantitative trait analysis was one-tailed under the hypothesis that an increased burden of minor alleles would reduce scores for cognitive traits and increase scores for anxiety, depression and neuroticism. References for all quantitative traits are available in the open-access Lothian Birth Cohort protocol paper (Deary et al. Sequenced: SNPs in protein-coding regions of exons including both synonymous and non-synonymous SNPs. PhastCon, SNPs within regions of conservation in placental mammals (PhastCon, UCSC). RegPot: SNPs with putative regulatory potential (UCSC, 7 x regulatory potential). CpG: SNPs in CpG islands commonly associated with promoter regions. Tested using the Hospital Anxiety Depression Scales. The Moray House Test is the general cognitive test—mostly a verbal reasoning, IQ-type test—that was used in the Scottish Mental Survey 1947.
AUTHOR CONTRIBUTIONS
DJP, PT and WRM designated the study, interpreted the data and wrote the paper. WJM, DJM, AMM and DHB were responsible for the psychiatric evaluation, selection and provision of case samples. DJP, PT, UJ, DLM, AMM and DHB were responsible for the provision of GSSFS samples and associated phenotypic measures. SEH, JMS and UD were responsible for the provision of LBC1936 samples and associated phenotypic measures. JSP developed the sample processing methodology and edited the paper. MK developed the sequence analysis pipeline and conducted the quality control of the sequence data and edited the paper. PT, AFM and PMW conducted the statistical analysis. KR, DCS, SM, SC and PT conducted the bioinformatics analysis. All authors contributed to revise the paper.

REFERENCES
13 Mathisson I, Munafò MR, Flint J. Meta-analysis indicates that common variants at the DISC1 locus are not associated with schizophrenia. Mol Psychiatry 2012; 17: 634–641.
15 Williams JM, Beck TF, Pearson DM, Proud MB, Cheung SW, Scott DAA. 1q42
22 Tiwary KR. The severity of mental disorders is linked to interaction among candidate genes. Integr Biol (Camb) 2012; 4: 1096–1101.
37 Millar JK, Pickard BS, Mackie S, James R, Christie S, Buchanan SR et al. DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling.[see comment]. Science 2005; 310: 1187–1191.
40 Millar JK, Christie S, Semple CA, Porteous DJ. Chromosomal location and genomic structure of the human translin-associated factor X gene (TRAX; TSNAX) revealed
Some properties of the hypergeometric distribution with applications

Lincoln FC. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D


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.

Petersen CGJ. The yearly immigration of young place into the Limfjord from the German Sea. Danish Biol Station 1896; 6: 1–77.


