Priority Communication

Biological Psychiatry

Genome-wide Regional Heritability Mapping Identifies a Locus Within the *TOX2* Gene Associated With Major Depressive Disorder

Yanni Zeng, Pau Navarro, Masoud Shirali, David M. Howard, Mark J. Adams, Lynsey S. Hall, Toni-Kim Clarke, Pippa A. Thomson, Blair H. Smith, Alison Murray, Sandosh Padmanabhan, Caroline Hayward, Thibaud Boutin, Donald J. MacIntyre, Cathryn M. Lewis, Naomi R. Wray, Divya Mehta, Brenda W.J.H. Penninx, Yuri Milaneschi, Bernhard T. Baune, Tracy Air, Jouke-Jan Hottenga, Hamdi Mbarek, Enrique Castelao, Giorgio Pistis, Thomas G. Schulze, Fabian Streit, Andreas J. Forstner, Enda M. Byrne, Nicholas G. Martin, Gerome Breen, Bertram Müller-Myhsok, Susanne Lucae, Stefan Kloiber, Enrico Domenici, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Ian J. Deary, David J. Porteous, Chris S. Haley, and Andrew M. McIntosh

ABSTRACT

BACKGROUND: Major depressive disorder (MDD) is the second largest cause of global disease burden. It has an estimated heritability of 37%, but published genome-wide association studies have so far identified few risk loci. Haplotype-block-based regional heritability mapping (HRHM) estimates the localized genetic variance explained by common variants within haplotype blocks, integrating the effects of multiple variants, and may be more powerful for identifying MDD-associated genomic regions.

METHODS: We applied HRHM to Generation Scotland: The Scottish Family Health Study, a large family- and population-based Scottish cohort ($N = 19,896$). Single-single nucleotide polymorphism (SNP) and haplotype-based association tests were used to localize the association signal within the regions identified by HRHM. Functional prediction was used to investigate the effect of MDD-associated SNPs within the regions.

RESULTS: A haplotype block across a 24-kb region within the *TOX2* gene reached genome-wide significance in HRHM. Single-SNP- and haplotype-based association tests demonstrated that five of nine genotyped SNPs and two haplotypes within this block were significantly associated with MDD. The expression of *TOX2* and a brain-specific long noncoding RNA RP1-269M15.3 in frontal cortex and nucleus accumbens basal ganglia, respectively, were significantly regulated by MDD-associated SNPs within this region. Both the regional heritability and single-SNP associations within this block were replicated in the UK–Ireland group of the most recent release of the Psychiatric Genomics Consortium (PGC), the PGC2–MDD (Major Depression Dataset). The SNP association was also replicated in a depressive symptom sample that shares some individuals with the PGC2–MDD.

CONCLUSIONS: This study highlights the value of HRHM for MDD and provides an important target within *TOX2* for further functional studies.

Keywords: Genome-wide analyses, Haplotype block, HRHM, MDD, Regional heritability, *TOX2*

http://dx.doi.org/10.1016/j.biopsych.2016.12.012

Major depressive disorder (MDD) is ranked as the second leading contributor to the global disease burden in terms of years lived with disability (1). The narrow sense heritability of MDD has been estimated to be 37% by twin studies (2), suggesting a substantial contribution from genetic factors. In efforts to identify specific genetic risk factors for MDD, family-based linkage studies have identified several significant peaks in certain families, but the findings have been inconsistent (3). Genome-wide association studies (GWASs) of unrelated participants have successfully identified hundreds of loci associated with other psychiatric disorders (4), but for MDD only four genome-wide significant and replicable loci have been identified by two large GWASs: one on a refined MDD phenotype for Chinese women and one on self-report-based depression using less intensive phenotyping in a much larger European sample (5–7).

Several factors may be responsible for the comparatively sparse GWAS results in MDD. First, MDD is likely to have a highly polygenic genetic architecture where the disease risk is conferred by many causal variants of small effect (8,9).

ISSN: 0006-3223

Combined with the high prevalence of MDD (10) and the possible incomplete linkage disequilibrium (LD) between genotyped single nucleotide polymorphisms (SNPs) and causal SNPs, single-SNP-based genome-wide association tests may have insufficient power to detect individual causal variants (11). Second, clinical heterogeneity has been shown in MDD between populations (6,12), and this may lead to difficulties in identifying causal variants across cohorts (13). Whereas GWAS sample sizes for MDD are increasing and efforts to refine the MDD phenotype are in progress (5,7), alternative methodologies for detecting the signal arising from causal variants within and across families may also be productive.

Regional heritability mapping (RHM) is a method used to identify small genomic regions accounting for a significant proportion of the phenotypic variance in a trait of interest (14). In contrast to single-SNP-based tests, RHM integrates effects from multiple SNPs by using a regional genetic relationship matrix estimated from SNPs within a region. The matrix is constructed for each region defined by a sliding window across the genome and is then used to estimate the variance explained by the variants within the region in a linear mixed model (14). The major advantage of RHM is that the regional genetic relationship matrices not only tag the effect of genotyped variants but also measure the effect of ungenotyped and rare variants, including those associated with the SNPs but with individual effects too small to be detected by GWASs (14,15). Previous studies have shown that RHM has greater power to detect rare variants and multiple alleles in regions where GWASs provided null findings (15–17). In 2014, Shirali *et al.* developed a haplotype-block-based RHM (HRHM) method as an improved version of RHM. HRHM uses haplotype blocks as the unit of mapping; therefore, the identified blocks have less complex local LD structures (18).

In this study, we applied HRHM to a homogeneous sample of approximately 20,000 Scottish participants containing both closely and distantly related subjects with genome-wide genotyping data and a standardized structured clinical MDD diagnosis (19). We sought to identify genomic regions conferring risk for MDD, which were then further explored using single-SNP- and haplotype-based association tests. We then examined the functional effects of the MDD-associated SNPs within the identified block. Finally, replication analyses were performed in independent samples for both the regional heritability and SNP association results.

METHODS AND MATERIALS

The Tayside Research Ethics Committee (reference 05/S1401/89) provided ethical approval for the study. Participants all gave written consent after having an opportunity to discuss the project and before any data or samples were collected.

Datasets

Discovery Sample: Generation Scotland: The Scottish Family Health Study. Generation Scotland: The Scottish Family Health Study (GS:SFHS) contains 21,387 subjects (n_{male} = 8772, n_{female} = 12,615; age_{mean} = 47.2 years, SD = 15.1) who were recruited from the registers of collaborating general practices in Glasgow, Tayside, Ayrshire, Arran, and Northeast regions of Scotland, United Kingdom. At least one first-degree relative aged 18 years or over was required to be identified for each participant (19,20). A structured clinical interview was used for the diagnosis of lifetime DSM-IV mood disorders (21,22). Details of MDD diagnosis, genotyping, quality control, and imputation methods are described in the Supplement. In total, 561,125 genotyped and 8,642,105 post-imputation autosomal SNPs that passed quality control criteria were available for 19,896 participants (2659 MDD cases and 17,237 control subjects) for subsequent analyses.

Replication Sample 1: UK Biobank. Data used in this study were provided as part of the UK Biobank project (reference no. 4844). Details for genotyping, quality control, imputation, and phenotyping are described in the Supplement. In brief, genotyping data were available for 152,729 UK Biobank participants recruited in the United Kingdom (23). The probable MDD phenotype was created based on the putative MDD definition established in Smith *et al.* using responses to a touchscreen questionnaire (24), from self-report information, and from inpatient records via linkage to hospital episode data (see Supplement). After quality control and removing subjects who were in both the GS:SFHS and UK Biobank datasets, and one of each pair of close relatives (relatedness >0.05) of GS:SFHS participants or the remaining UK Biobank participants, 1,198,327 SNPs for 24,015 subjects with the putative MDD phenotype available (8143 cases and 15,872 control subjects) remained in downstream analyses.

Replication Sample 2: Psychiatric Genomics Consortium Major Depression Dataset. The Psychiatric Genomics Consortium (PGC) provided individual genotypes (best guess) of imputed SNPs for participants from 22 cohorts in the PGC Major Depression Dataset (PGC2–MDD) (Supplemental Table S1). All cases met DSM-IV criteria for life MDD; the majority of them were ascertained clinically. Most control samples were screened, and participants with lifetime MDD were removed (Supplemental Table S1). Details for genotyping, quality control, imputation, and phenotyping are described in the Supplement. After quality control and removing subjects who overlapped with the GS:SFHS and UK Biobank datasets, 32,554 subjects of European ancestry (13,261 cases and 19,293 control subjects) were used in downstream analysis. Consistent with earlier work (25,26), we grouped the 22 cohorts into 7 groups based on the country of ancestor information for regional heritability analysis (Supplemental Table S1).

Replication Sample 3: Depressive Symptom Datasets. The depressive symptom (DS) sample contains overlapping individuals with replication samples 1 and 2. Okbay *et al.* carried out a GWAS meta-analysis (N = 180,866) on three samples using depressive symptoms as the trait of interest (27). The ascertained MDD diagnosis information was available for two samples: PGC1–MDD (n_{cases} = 9240, $n_{controls}$ = 9519) and the Resource for Genetic Epidemiology Research on Aging (n_{cases} = 7231, $n_{controls}$ = 49,316) (27). For the third sample, UK Biobank (N = 105,739), a continuous phenotype measuring the severity of depressive symptom had been created and used in the meta-analysis (27). Although this sample overlapped with the PGC2–MDD and UK Biobank samples, it provided results based on a nondiagnostic quantitative measure of depressive symptoms and involved