Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder

Genetics of Personality Consortium

**IMPORTANCE** Neuroticism is a pervasive risk factor for psychiatric conditions. It genetically overlaps with major depressive disorder (MDD) and is therefore an important phenotype for psychiatric genetics. The Genetics of Personality Consortium has created a resource for genome-wide association analyses of personality traits in more than 63,000 participants (including MDD cases).

**OBJECTIVES** To identify genetic variants associated with neuroticism by performing a meta-analysis of genome-wide association results based on 1000 Genomes imputation; to evaluate whether common genetic variants as assessed by single-nucleotide polymorphisms (SNPs) explain variation in neuroticism by estimating SNP-based heritability; and to examine whether SNPs that predict neuroticism also predict MDD.

**DESIGN, SETTING, AND PARTICIPANTS** Genome-wide association meta-analysis of 30 cohorts with genome-wide genotype, personality, and MDD data from the Genetics of Personality Consortium. The study included 63,661 participants from 29 discovery cohorts and 9,786 participants from a replication cohort. Participants came from Europe, the United States, or Australia. Analyses were conducted between 2012 and 2014.

**MAIN OUTCOMES AND MEASURES** Neuroticism scores harmonized across all 29 discovery cohorts by item response theory analysis, and clinical MDD case-control status in 2 of the cohorts.

**RESULTS** A genome-wide significant SNP was found on 3p14 in MAGI1 (rs35855737; \(P = 9.26 \times 10^{-9}\) in the discovery meta-analysis). This association was not replicated (\(P = .32\)), but the SNP was still genome-wide significant in the meta-analysis of all 30 cohorts (\(P = 2.38 \times 10^{-8}\)). Common genetic variants explain 15% of the variance in neuroticism. Polygenic scores based on the meta-analysis of neuroticism in 27 cohorts significantly predicted neuroticism (\(1.09 \times 10^{-12} < P < .05\)) and MDD (\(4.02 \times 10^{-9} < P < .05\)) in the 2 other cohorts.

**CONCLUSIONS AND RELEVANCE** This study identifies a novel locus for neuroticism. The variant is located in a known gene that has been associated with bipolar disorder and schizophrenia in previous studies. In addition, the study shows that neuroticism is influenced by many genetic variants of small effect that are either common or tagged by common variants. These genetic variants also influence MDD. Future studies should confirm the role of the MAGI1 locus for neuroticism and further investigate the association of MAGI1 and the polygenic association to a range of other psychiatric disorders that are phenotypically correlated with neuroticism.

JAMA Psychiatry. doi:10.1001/jamapsychiatry.2015.0554
Published online May 20, 2015.
Dimensions of personality have been linked with the liability to have psychiatric illness. Perhaps the strongest link between personality and psychiatric illness is the association of neuroticism with major depressive disorder (MDD). Neuroticism is also associated with other psychiatric disorders that entail emotional dysregulation, including personality, substance use, and anxiety disorders. Furthermore, neuroticism is associated with neurological diseases such as migraine and Alzheimer disease. Hence, neuroticism is a psychological risk factor of profound public health significance.

Neuroticism refers to the tendency to experience diverse and relatively more intense negative emotions. Neuroticism and similar traits such as harm avoidance and negative emotionality share an affective underpinning and are found in all main theories of personality. Twin studies of neuroticism, harm avoidance, or negative emotionality generally find that between 40% and 60% of the trait variance is explained by genomic variation, and it has been found that there are no large age-by-genotype or sex-by-genotype effects, modest assortative mating, and large genetic and phenotypic stability across the life span. These findings and the fact that neuroticism is strongly related to MDD make neuroticism an important phenotype for psychiatric genetic studies.

Genome-wide association (GWA) studies require large sample sizes to have sufficient statistical power, which is often achieved by aggregating results in multiple cohorts in a meta-analysis. However, this requires a measurement scale that is comparable across cohorts. We recently showed for neuroticism that different personality instruments could be linked through item response theory analysis to assess the same underlying constructs. Personality item data were harmonized in more than 160,000 participants from the Genetics of Personality Consortium. A meta-analysis of data from more than 29,000 twin pairs from 6 of the participating cohorts showed that the heritability of the harmonized neuroticism scores was 48%. This estimate was based on twin correlations that ranged between 0.39 and 0.53 for monozygotic twin pairs and between 0.11 and 0.26 for dizygotic twin pairs across cohorts and sexes. The opposite-sex twin correlations were not significantly lower than the same-sex dizygotic twin correlations, illustrating that the same genetic factors influence neuroticism in men and women.

Gene-finding studies for MDD and neuroticism-like personality traits have had limited success to date. There have been 2 meta-analytic GWA studies for personality traits, including neuroticism and harm avoidance. The sample sizes were small by current standards (N = 11,590 in the study by Service et al and N = 17,375 in the study by de Moor et al) and single-nucleotide polymorphisms (SNPs) were imputed using HAPMAP as a reference. The largest GWA studies for MDD are those from the Psychiatric Genomics Consortium, with 92,400 MDD cases and 9519 controls in the discovery phase of the study and 67,833 MDD cases and 50,695 controls in the replication phase, and imputation based on HAPMAP. These studies did not detect genome-wide significant SNPs.

To assess whether gene-finding efforts are likely to have success, techniques have been developed that test whether common variants tagged by genome-wide SNP arrays contribute to variation in the phenotype. Two such studies for neuroticism found effects of common SNPs, explaining about 6% of the phenotypic variance, and another study for MDD found that common SNPs explain 28% to 32% of the phenotypic variance. In young children, genome-wide SNPs explained 13% to 43% of the variance in internalizing problems.

Herein, we report results of the largest GWA study for neuroticism so far, to our knowledge, conducted in 63,661 participants from 29 cohorts. A replication cohort of 9786 participants was also included. Imputation was performed against the 1000 Genomes reference panel. The main aim of the study was to identify genetic variants for neuroticism by performing a meta-analysis of GWA results. Additional aims were to estimate SNP-based heritability in 2 of the largest cohorts to establish that the sets of SNPs contain information to detect genetic variants and to test whether these variants predict MDD status in a large cohort of clinically assessed MDD cases and screened controls.

**Methods**

**Cohorts**

The meta-analysis included 29 discovery cohorts, with 21 cohorts from Europe, 6 from the United States, and 2 from Australia. The analyses were conducted between 2012 and 2014. All participants were of European descent. The total sample size was 63,661 for the GWA meta-analysis. The Generation Scotland: Scottish Family Health Study cohort (n = 9786) was included for replication of GWA top results. For more information on each cohort, see eAppendix 1 and eTable 1 in the Supplement. Approval by local institutional review boards was obtained in all studies and informed consent was obtained from all participants.

**Phenotyping**

After harmonizing all item data on neuroticism from multiple instruments, comparable neuroticism scores were obtained for all cohorts. These scores were estimated for all participants after conducting item response theory analysis on the available item data for neuroticism from the NEO Personality Inventory, Eysenck Personality Questionnaire, and International Personality Item Pool inventory, all item data for harm avoidance from Cloninger’s Tridimensional Personality Questionnaire, and all item data for negative emotionality from the Multidimensional Personality Questionnaire (eAppendix 1 in the Supplement). For the Generation Scotland cohort, phenotypes were summed scores on the neuroticism scale of the Eysenck Personality Questionnaire Revised Short Form.

**Genotyping and Imputation**

An overview of SNP genotyping, quality control, and imputation is given in eTable 2 in the Supplement. Quality control of genotype data was performed in each study independently, using comparable but study-specific criteria. Basic quality control steps included checks for European ancestry, sex incon-
Genotype data were further checked based on Hardy-Weinberg equilibrium, minor allele frequencies (MAFs), SNP, and sample call rates. Genotype data were imputed using the 1000 Genomes phase 1 version 3 (build 37, hg19) reference panel with standard software packages such as IMPUTE, MACH, or Minimac (eTable 2 in the Supplement).

Statistical Analysis

GWA Analyses in Each Cohort
The GWA analyses were conducted in each cohort using linear regression (additive model, with sex and age as covariates) with the aim to identify single common genetic variants that influence neuroticism in both men and women of different ages. Depending on the characteristics of the cohort, additional covariates such as principal components were added. Different software packages were used to run the association analysis (eTable 2 in the Supplement). Uncertainty of the imputed genotypes was taken into account. In those cohorts that included related individuals, the dependency among participants was accounted for. Locations of SNPs are reported on build 37 (hg19).

Meta-analysis of GWA Results Across Cohorts
A meta-analysis of the GWA results of the discovery cohorts was conducted using the weighted inverse variance method in METAL (http://csg.sph.umich.edu/abecasis/Metal/). This is a fixed-effects model in which effect sizes (β) are weighted by the inverse of their variance and then summed over cohorts. This model is appropriate if phenotypes are on a similar scale, which was the case for the harmonized neuroticism scores. Poorly imputed SNPs ($r^2 < 0.30$ or proper_info $< 0.40$) and SNPs with low MAF (MAF $< \sqrt{5/n}$, which corresponds to fewer than 5 estimated individuals in the least frequent genotype group, under the assumption of Hardy-Weinberg equilibrium) were excluded, resulting in a total number of 1.1 million to 6.6 million SNPs across cohorts. The number of unique SNPs available for meta-analysis was 7,480,565. For 530,951 SNPs, association results were available in 1 cohort only and were discarded, leading to a final 6,949,614 SNPs for which results are reported. Genomic control inflation factors ($\lambda$) and Manhattan and quantile-quantile plots per cohort are provided in eTable 3, eFigure 1, and eFigure 2 in the Supplement. The SNPs with $P = 5 \times 10^{-8}$ or smaller were considered genome-wide significant. In the Generation Scotland cohort, all SNPs with $P < 1 \times 10^{-5}$ were tested for replication. For these SNPs, a meta-analysis of all 30 cohorts was conducted. Because sum scores were available for neuroticism in the Generation Scotland cohort, this meta-analysis was based on combining $P$ values, taking into account the direction of effect and weighting by sample size, rather than combining effect sizes.
Variance Explained by Common SNPs
In 2 large cohorts included in the meta-analysis, the Netherlands Twin Register (NTR; n = 3599 unrelated individuals) cohort and the QIMR Berghofer Medical Research Institute (QIMR; n = 3369) adult cohort, genomic-relatedness-matrix restricted maximum likelihood analysis in the Genome-wide Complex Trait Analysis (GCTA) software was applied to estimate the proportion of variance in neuroticism that can be explained by common SNPs. The GCTA analysis was based on best-guess genotypes obtained in PLINK using a threshold of a maximum genotype probability > 0.70, and additionally filtering on r² > 0.80. Next, in estimating the genetic relationship matrix in the GCTA software, SNPs with MAF < 0.05 were excluded. The additive genetic relationship matrix for all individuals in the data sets estimated based on SNPs was used to estimate the proportion of phenotypic variance due to additive genetic variance. Sex, age, and population-specific principal components were included as covariates.

Polygenic Risk Score Analysis
Polygenic risk score (PGS) analyses were conducted to test the predictive power of the meta-analysis results for neuroticism itself and for MDD. The PGSs were computed in the NTR cohort and the Netherlands Study of Depression and Anxiety (NESDA) cohort and were based on the meta-analysis results excluding the NTR and NESDA cohorts, further referred to as the PGS discovery set. The PGSs were calculated for all individuals of the NTR and NESDA target set by taking a set of most significant SNPs from the analysis in the PGS discovery set, multiplying the individual’s genotypic score (0, 1, or 2 for genotyped SNPs, or any value between 0-2 for imputed SNPs) by the effect size of a particular SNP (unstandardized regression coefficient based on the meta-analysis), and summing this over SNPs. The PGSs were calculated for 6 P value thresholds (P < 1 × 10⁻⁵, P < 1 × 10⁻⁴, P < 1 × 10⁻³, P < .01, P < .05 and P < .50). Next, linear regression was conducted to predict neuroticism from the PGSs in 8648 NTR participants. Logistic regression was conducted to predict MDD status in 1859 unrelated MDD cases and 2391 unrelated controls from the NTR and NESDA cohorts. The MDD case-control status was determined to be unrelated MDD cases and 2391 unrelated controls from the NTR and NESDA cohorts. The MDD case-control status was determined.

Results

Meta-analysis of GWA Results for Neuroticism
Meta-analysis of GWA results across the 29 cohorts revealed 1 genome-wide significant SNP (rs35855737; P = 9.26 × 10⁻⁹). The SNP is located on 3p14 in an intron of MAGI1 (Figure 1). The pooled regression effect was −0.04 with the minor allele C coded as the effect allele (Figure 2). Imputation quality was very high (r² or proper_info > 0.94) in all cohorts, except in the Erasmus Rucphen Family (ERF) cohort (r² = 0.63). The MAF of the SNP ranged from 0.13 to 0.22 across cohorts with imputation quality greater than 0.94 and showed a mean (SD) of 0.18 (0.02), which corresponds to the MAF for this SNP in the 1000 Genomes reference set. The MAF in the ERF cohort was 0.07. Eleven other SNPs in MAGI1 showed suggestive genome-wide significance (P < 1 × 10⁻⁵); all SNPs were intronic; 1 SNP was in very high linkage disequilibrium with rs35855737 (rs1404544; r² > 0.80; P = 8.59 × 10⁻⁹); and 3 SNPs were in high linkage disequilibrium with rs35855737 (rs1524970, rs1880522, and rs6799284; r² > 0.60; 3.64 × 10⁻⁶ < P < 8.54 × 10⁻⁷). The

Figure 2. Forest Plot for Genome-wide Significant Single-Nucleotide Polymorphism rs35855737 in MAGI1 on Chromosome 3 for Neuroticism
Manhattan and quantile-quantile plots are shown in Figure 3 and Figure 4. A list with all 127 suggestively genome-wide significant SNPs is provided in eTable 4 in the Supplement (full results of the meta-analysis can be downloaded from http://www.tweelingenregister.org/GPC).

Results of the follow-up analysis for all SNPs with $P < 1 \times 10^{-5}$ in the Generation Scotland cohort are displayed in eTable 4 in the Supplement. The SNP rs35855737 is not significantly associated with neuroticism in the Generation Scotland cohort, but the direction of the effect is the same ($\beta = -0.02$ for effect allele C; $P = .32$). A meta-analysis of the results from all 30 cohorts shows that rs35855737 remains genome-wide significant ($\beta = -0.04; P = 2.38 \times 10^{-8}$).

Variance in Neuroticism Explained by Common SNPs
In the NTR cohort, 14.7% of the variance in neuroticism was explained by all SNPs ($P = .02; 95\% CI, 0.002-0.29$). In the QIMR cohort, 15.7% of the variance was explained by SNPs ($P = .18; 95\% CI, 0-0.47$).

Polygenic Risk Score Analysis for Neuroticism and MDD
The results of the polygenic risk score analyses are presented in Figure 5. In the NTR cohort, polygenic risk scores are significantly ($P < .05$) associated with neuroticism when polygenic scores are based on SNP sets with thresholds of
Discussion

The meta-analysis of GWA results for neuroticism showed a genome-wide significant SNP on 3p14 in an intron of MAGI1. This gene is expressed in neuronal tissue, in particular the hippocampus, and is found at the synaptic plasma membrane.50 Also, MAGI1 acts as a scaffolding protein in the neurite growth factor receptor-mediated signaling pathway.51 Interestingly, MAGI1 has previously been implicated in bipolar disorder, schizophrenia, and episodicity in MDD,52-54 disorders that in part share their genetic etiology.55 A genome-wide linkage scan for early-onset bipolar disorder.56 The SNP-based genetic similarity across individuals accounted for approximately 15% of the variance in neuroticism. This estimate is larger than those in earlier reports, of about 6%.44,46 Heritability estimates from twin studies are usually larger and range between 40% and 55%.36

Polygenic risk scores based on the GWA meta-analysis significantly predicted MDD status in a large independent target set consisting of MDD cases and screened controls. The polygenic scores for neuroticism reassuringly also predicted neuroticism in controls for MDD of this same independent set. Neuroticism and MDD were explained about equally well by neuroticism polygenic scores (up to 1.6% explained variance). These findings are consistent with previous reports that studied the prediction of MDD and bipolar disorder based on polygenic scores derived from Big Five neuroticism GWA results.56-57

This study demonstrates that increasing the number of participants and SNPs in a meta-analysis was successful in identifying a novel locus for neuroticism. As expected, the effect size of the identified SNP is very small (pooled regression coefficient of -0.04 for the harmonized score with a variance of approximately 1). Together with our findings of an SNP-based heritability of approximately 15% and an increase in explained variance in the polygenic risk score analysis when polygenic scores are based on larger sets of SNPs, this suggests that neuroticism is highly polygenic. Our results further indicate that the heritability of neuroticism likely consists not only of common SNPs with small effects. Rare variants, repeat polymorphisms, and indels may also influence neuroticism, possibly in gene-by-gene interactions (epistasis). As a consequence, to further our understanding of the genetic and molecular basis of neuroticism (and associated psychiatric disorders), different routes need to be taken in future studies. One route would be to increase the number of participants and SNPs to further identify common variants. This route was shown to be very successful for schizophrenia.58 Also, the study of variants other than common SNPs should be pursued. Alternative routes could include pathway analyses and genetic studies that are informed by results from the animal literature on basic emotions such as fear, sadness, and anger.59-62

This study more than tripled the sample size compared with the previously published meta-analysis on personality,38 providing a substantial increase in power to detect variants. The power to detect variants that explain 0.23% of the variance (corresponding to the effect size for the most significant SNP in the previous meta-analysis38) increased from 84% to 100%. In addition, with a sample size of 63,661 individuals there is 80% power to detect variants that explain at least 0.063% of the variance in neuroticism, compared with 1.6% power given the 17,375 participants who were included in the previous meta-analysis.38 The large increase in sample size was possible because an item response theory approach enabled harmonization of personality data obtained from different personality questionnaires, which may serve as an example for gene-finding studies for other psychological, cognitive, and psychiatric traits where harmonization is required to increase sample size (eg, symptoms of depression or attention-deficit/hyperactivity disorder measured by different questionnaires).

Conclusions

The results for neuroticism were predictive for MDD. Future analyses may focus on whether the MAGI1 locus and polygenic variance for neuroticism are also associated with psychiatric disorders that are phenotypically associated with neuroticism, such as borderline personality disorder, bipolar disorder, schizophrenia, attention-deficit/hyperactivity disorder, and substance use disorders. This could be achieved by combining data from the Genetics of Personality Consortium with those available within the Psychiatric Genomics Consortium63 and the Social Science Genetic Association Consortium.64 Novel methods will be needed to test whether neuroticism represents a causal risk factor for MDD and other disorders, whether reverse causality is also present, or whether the genetic association between neuroticism and psychiatric disorders reflects an underlying common liability.55,65-67 It is expected that such studies will increase our understanding of the role that emotional instability plays in the occurrence and course of psychiatric disorders and other important health outcomes.
Experimental Medicine, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh (van den Berg); Generation Scotland, University of Edinburgh, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, Scotland (Hayward); Department of Psychiatry and Psychotherapy, HELIOS Hospital Stralsund, Stralsund, Germany (Graves); Institute of Public Health, University of Southern Denmark, Odense, Denmark (McGue).

Submitted for Publication: September 2, 2014; final revision received March 5, 2015; accepted March 7, 2015.

Author Contributions: Drs de Moor and van den Berg had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs de Moor and van den Berg share first authorship. Study concept and design: de Moor, van den Berg, van Duijn, Terracciano, McGue, Boomsma. Acquisition, analysis, or interpretation of data: de Moor, van den Berg, Verweij, Krueger, Luciano, Arias Vasquez, Matteson, Derringer, Esko, Amin, Hansell, Hart, Seppälä, Huffman, Korte, Lahti, Lee, Miller, Nutile, Tanaka, Teumer, Viktørin, Wedenojo, Abecasis, Adkins, Agrawal, Allik, Appel, Bigdeli, Busonero, Campbell, Costa, Davy, Smith-Davies, Davies, de Wit, Ding, Engelhardt, Eriksson, Fedkin, Ferrucci, Giegling, Grucza, Hartmann, Heath, Heienonen, Henders, Homuth, Hottenga, Iacono, Jokela, Karlsson, Kemp, Kirkpatrick, Latvala, Liewald, Madden, Magri, Magnusson, Maren, Medland, Mihalov, Milaneschi, Montgomery, Nauck, Ouwens, Palotie, Pettersson, Polasek, Qian, Pulski-Ràbkit, Reikard, Rose, Slutske, Starr, Wright, Porteous, Rujescu, Rudan, Kaprio, Deary, Räikkönen, Wilson, Grabe, van Duijn, Schlessinger, Pedersen, McGue, Penninx, Boomsma.

Statistical analysis: de Moor, van den Berg, Verweij, Luciano, Arias Vasquez, Matteson, Derringer, Esko, Amin, Hansell, Hart, Seppälä, Huffman, Korte, Lahti, Lee, Miller, Nutile, Tanaka, Teumer, Viktørin, Wedenojo, Adkins, Allik, Bigdeli, Ding, Engelhardt, Fedkin, Ferrucci, Jokela, Karlsson, Kemp, Magnusson, Maren, Medland, Mihalov, Milaneschi, Polasek, Qian, Ruggiero, Sorice, Woda, Zgaga, Ciullo, Hartmann, Hejtema, Evans, Terracciano, McGue, Martin.

Submitted for Publication: September 2, 2014; final revision received March 5, 2015; accepted March 7, 2015.

Author Contributions: Drs de Moor and van den Berg had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs de Moor and van den Berg share first authorship. Study concept and design: de Moor, van den Berg, van Duijn, Abecasis, Davey Smith, Franke, Giegling, Heath, Heienonen, Iacono, Jokela, Karlsson, Kemp, Magnusson, Maren, Medland, Mihalov, Milaneschi, Polasek, Qian, Ruggiero, Sorice, Woda, Zgaga, Ciullo, Hartmann, Hejtema, Evans, Terracciano, McGue, Martin.

Conflict of Interest Disclosures: Dr Grabe reported having received funding by the German Research Foundation, the German Federal Ministry of Education and Research, and the DAMP Foundation as well as speaker’s honoraria from Servier and Eli Lilly and Co. No other disclosures were reported.

Disclaimer: This publication is the work of the authors, and they will serve as guarantors for the contents of this article.

Additional Information: Analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is supported by grant NWO 480-05-003 from the Netherlands Organization for Scientific Research. Generation Scotland is a collaboration between the University Medical Schools and the National Health Service (NHS), Aberdeen, Dundee, Edinburgh, and Glasgow, United Kingdom. Acknowledgments by the cohorts are listed in Appendix 2 in the Supplement.

Additionl Contributions: We thank all the participants.

REFERENCES


