The genetic association between personality and major depression or bipolar disorder. A polygenic score analysis using genome-wide association data

P. A. F. Madden  
*Washington University School of Medicine in St. Louis*

A. C. Heath  
*Washington University School of Medicine in St. Louis*

M. L. Pergadia  
*Washington University School of Medicine in St. Louis*

A. Agrawal  
*Washington University School of Medicine in St. Louis*

P. Lin  
*Washington University School of Medicine in St. Louis*

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Please let us know how this document benefits you.

Recommended Citation
https://digitalcommons.wustl.edu/open_access_pubs/1358

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Authors
P. A. F. Madden, A. C. Heath, M. L. Pergadia, A. Agrawal, P. Lin, R. A. Grucza, L. J. Bierut, and et al
The genetic association between personality and major depression or bipolar disorder. A polygenic score analysis using genome-wide association data

CM Middeldorp1,2,3,7, MHM de Moor1,2,3,7, LM McGrath5, SD Gordon6, DH Blackwood7, PT Costa8, A Terracciano8, RF Krueger9, EJC de Geus2, DR Nyholt10, T Tanaka8, T Esko11,12, PAF Madden13, J Derriére9, N Amin14, G Willemsen14, J-H Hottenga15, MA Distl11, M Ueda16, S Sanna16, P Spinhoven17, CA Hartman18, S Ripke19, PF Sullivan20, A Reale21, J Allik21, AC Heath13, ML Pergadia22, A Agrawal23, P Lin13, RA Grucca23, E Widen22, DL Cousminer22, JG Eriksson23,24,25,26,27, A Palotie9,28,29,30, PH Lee16, M Luciano16, A Tenesa31, G Davies31, LM Lopez31, NK Hansell16, SE Medland6, L Ferrucci8, D Schlessinger8, GW Montgomery6, MJ Wright6, YS Aulchenko14, AGJW Janssens14, BA Oostra33, A Metspalu10,11,12, GR Abecasis34, PJ Deary31, K Raikkonen35, LJ Bierut13, NG Martin6, NR Wray6, CM van Duijn14, JW Smoller8, BWJH Penninx15,17,18,36 and DI Boomsma12

The relationship between major depressive disorder (MDD) and bipolar disorder (BD) remains controversial. Previous research has reported differences and similarities in risk factors for MDD and BD, such as predisposing personality traits. For example, high neuroticism is related to both disorders, whereas openness to experience is specific for BD. This study examined the genetic association between personality and MDD and BD by applying polygenic scores for neuroticism, extraversion, openness to experience, agreeableness and conscientiousness to both disorders. Polygenic scores reflect the weighted sum of multiple single-nucleotide polymorphism alleles associated with the trait for an individual and were based on a meta-analysis of genome-wide association studies for personality traits including 13 835 subjects. Polygenic scores were tested for MDD in the combined Genetic Association Information Network (GAIN-MDD) and MDD2000 samples (N = 8921) and for BD in the combined Systematic Treatment Enhancement Program for Bipolar Disorder and Wellcome Trust Case-Control Consortium samples (N = 6329) using logistic regression analyses. At the phenotypic level, personality dimensions were associated with MDD and BD. Polygenic neuroticism scores were significantly positively associated with MDD, whereas polygenic extraversion scores were significantly positively associated with BD. The explained variance of MDD and BD, ~0.1%, was highly comparable to the variance explained by the polygenic personality scores in the corresponding personality traits themselves (between 0.1 and 0.4%). This indicates that the proportions of variance explained in mood disorders are at the upper limit of what could have been expected. This study suggests shared genetic risk factors for neuroticism and MDD on the one hand and for extraversion and BD on the other.

Translational Psychiatry (2011) 1, e50; doi:10.1038/tp.2011.45; published online 18 October 2011

1Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; 2Department of Psychology, Academic Medical Center, Amsterdam, The Netherlands; 3Department of Child and Adolescent Psychiatry, GGZ inGeest/VU Medical Center, Amsterdam, The Netherlands; 4Department of Psychiatry and Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 5Department of Genetics, University of Helsinki, Helsinki, Finland; 6Genetic Epidemiology Unit, Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 7Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK; 8National Institute on Aging, NIH, Baltimore, MD, USA; 9Department of Psychiatry, University of Minnesota, Baltimore, MD, USA; 10Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 11Estonian Biocentre, Tartu, Estonia; 12Estonian Genome Center of University of Tartu, Tartu, Estonia; 13Department of Psychiatry, Washington University School of Medicine, St Louis, Washington, USA; 14Department of Psychology, Erasmus University Medical Center, Rotterdam, The Netherlands; 15EMGO Institute for Health And Care Research, VU University Medical Center, Amsterdam, The Netherlands; 16Istituto di Neurogenetica e Neurofarmacologia, Cagliari, Italy; 17Departments of Clinical Psychology and Psychiatry, Leiden University, Leiden, The Netherlands; 18Department of Psychiatry, University Medical Center Groningen, Groningen, The Netherlands; 19Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 20Department of Genetics, University of North Carolina at Chapel Hill, NC, Chapel Hill, USA; 21Department of Psychology, University of Tartu, Tartu, Estonia; 22Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 23National Institute for Health and Welfare, Helsinki, Finland; 24Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; 25Vasa Central Hospital, Vasa, Finland; 26Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland; 27Folkhälsan Research Centre, Helsinki, Finland; 28Wellcome Trust Genome Campus, Wellcome Trust Sanger Institute, Cambridge, UK; 29Department of Medical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 30Department of Psychiatry, University of Cambridge, Cambridge, UK; 31Department of Psychology, Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh, UK; 32MRC Human Genetics Unit, The Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK; 33Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; 34Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 35Department of Psychology, University of Helsinki, Helsinki, Finland and 36Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands

Correspondence: Dr C Middeldorp, Department of Biological Psychology, VU University Amsterdam, Van der Boechorststraat 1, Amsterdam 1081BT, The Netherlands. E-mail: cm.middeldorp@psy.vu.nl

These authors contributed equally to this work.

Keywords: bipolar disorder; genetic correlation; genome-wide association; polygenic score analysis; personality-major depression

Received 23 May 2011; revised 19 August 2011; accepted 31 August 2011
Introduction

The relationship between major depressive disorder (MDD) and bipolar disorder (BD) remains controversial. Both are characterized by depressive episodes, whereas BD is, in addition, characterized by manic episodes. Research into risk factors for MDD and BD reports similarities and differences.

This also applies to studies investigating the association between personality and MDD or BD. Most of these studies have focused on the personality traits neuroticism/negative emotionality (N), extraversion/positive emotionality (E) and, to a lesser extent, agreeableness (A), conscientiousness (C) and openness to experience (O). N is commonly defined as a tendency toward emotional instability. E is characterized by a disposition toward positive emotions, gregariousness and the tendency to be active, seek out stimulation and enjoy the company of others. O involves active imagination, aesthetic attentiveness, variety preference and intellectual curiosity. A can be defined as the tendency to be cooperative and compassionate rather than suspicious and antagonistic toward others. Finally, the dimension ‘C’ reflects traits of self-discipline, carefulness, thoroughness, organization, deliberation and achievement.

In a recent meta-analysis on the relationship between personality and MDD, MDD was significantly associated with higher N and with lower C with a Cohen’s d of 1.33 for N and −0.90 for C. The association with C became weaker (Cohen’s d −0.59) after controlling for N, but remained significant. Although a negative link between E and MDD has often been reported, the effect was modest and not significant in the meta-analysis (Cohen’s d −0.62). The associations with O and A were not significant either.

Studies on the association between personality and BD are sparser, but have consistently shown higher levels on N and O and lower levels of C compared with normal controls. This suggests that subjects with MDD and BD are similar regarding N and C and differ regarding O. This is supported by studies directly comparing personality profiles for MDD and BD. All studies showed the same trend with higher O in BD subjects than in MDD subjects. This was significant in only one of these studies, but the other samples included far fewer subjects and probably did not have the power to detect the effect (<100 BD subjects versus ~1000 subjects). Most of these studies have been performed in MDD or BD subjects in an euthymic phase; thus, the results do not reflect a state effect of mood on personality.

Mood disorders and personality traits are partly influenced by genetic risk factors. Heritability estimates are ~40% for MDD, 50% for personality traits and between 60 and 90% for BD. This raises the question whether associations between personality and mood disorders are explained by shared genetic risk factors. So far, this has only been investigated for MDD. Twin studies have provided considerable support for overlapping genetic risk factors influencing N and MDD (reviewed in Middeldorp et al.). Fewer twin studies have investigated the association with other personality traits and MDD suggesting a smaller, but significant genetic correlation between C and O and MDD, but not between E and MDD.

Genome-wide association (GWA) data also provide an opportunity to investigate whether traits are influenced by overlapping genetic risk factors. On the basis of GWA results for one trait, for instance, neuroticism, performed in one sample (the discovery sample), a polygenic score is calculated for each individual in another sample (the target sample). These polygenic scores are obtained by taking a set of top single-nucleotide polymorphisms (SNPs), for example, all SNPs with P-values below 0.1 and multiplying the individual’s genotypic score (0, 1 or 2) by the effect of the SNP. If the polygenic scores are significantly related to a second trait, for instance, MDD, in the target sample, this indicates that the two traits, namely neuroticism and MDD, are influenced by overlapping genetic risk factors. In this manner, a genetic relationship was observed for schizophrenia and BD and for MDD and anxiety. In the former study, polygenic scores based on a GWA study in schizophrenia explained between 1 and 2% in BD. In the latter study, polygenic scores based on a MDD GWA study explained ~2% of the variance in anxiety disorders.

This study investigates the genetic association between the five personality traits N, E, O, A and C and MDD, as well as BD. On the basis of the results of a GWA meta-analysis of N, E, O, A and C in >13 000 subjects, individual polygenic scores were calculated and tested for their effect on case-control status in 2 combined target MDD samples and in 2 combined target BD samples totaling 8921 and 6329 subjects, respectively. We first asked whether the genetic association between N and MDD as found in the twin studies is confirmed using polygenic score analysis and to what extent the other personality traits are genetically associated with MDD. Second, we investigated the genetic relationship between personality and BD. Finally, we asked what the differences are between the genetically mediated personality profiles underlying BD and MDD.

Materials and methods

Subjects, measurement instruments and genotyping

Discovery samples for personality. The GWA meta-analyses were performed on personality data collected from nine samples: SardiNIA–Italy, Erasmus Rucphen Family study (ERF)–The Netherlands, Study of Addiction: Genetics and Environment (SAGE)–United States of America, Helsinki Birth Cohort Study (HBCS)–Finland, Nicotin Addiction Genetics Study/Interactive Research Project Grants (NAG/IRPG) study–Australia, Queensland Institute of Medical Research (QIMR) adolescent study–Australia, Lothian Birth Cohort 36 (LBC36)–United Kingdom, Baltimore Longitudinal Study of Aging (BLSA)–United States of America and Estonian Genome Project of University of Tartu (EGPUT)–Estonia. For a detailed description of these samples, we refer to de Moor et al. The total number of subjects available for the meta-analyses was 13 835. Sample sizes ranged from 600 to 3972 individuals. Mean age ranged from 19 to 70 years. In 5 studies, the mean age was between 40 and 50 years, in 1 study the mean age was 19 years, and in 3 studies, the mean age was between 60 and 70 years. It must be noted that the meta-analysis as described in de Moor et al. also included the GAIN-MDD sample. This sample was excluded in the personality traits meta-analyses for this
study as the GAIN-MDD set served as one of the target samples (see below for the description of the sample).

Personality scores were assessed with NEO Personality Inventory—Revised (NEO-PI-R), NEO-PI-3 or the NEO Five-Factor Inventory. In each study, scores for the 5 factors N, E, O, A and C were based on the 60 items of the NEO Five-Factor Inventory (12 items per phenotype). Summed scores were computed for all five personality dimensions.

DNA was extracted from blood samples. Genotyping was performed on Illumina platforms (Illumina, San Diego, CA, USA) in all studies, except in SardiNIA in which an Affymetrix platform (Affymetrix, Santa Clara, CA, USA) was used. Genotype data were checked in each study independently, using slightly different inclusion criteria. Among the basic checks that were performed are checks for European ancestry, Mendelian errors, gender inconsistencies and high genome-wide homozygosity. Genotype data were further cleaned based on Hardy–Weinberg equilibrium, minor allele frequencies, SNP call rate (% of subjects with missing SNPs per subject). Imputation to $\sim 2.5\text{M}$ common SNPs included in HapMap was performed using the HapMap phase II CEU data as the reference sample (NCBI build 36/UCSC hg18, Bethesda, MD, USA). Imputation was carried out using IMPUTE for SAGE and EGPUT samples. For the other samples, genotype data were imputed using MACH software.

Target samples for MDD and BD. Polygenic scores were tested in two MDD case–control samples: GAIN-MDD and MDD2000 + and in two BD case–control samples: Wellcome Trust Case–Control Consortium (WTCCC) and Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). The four samples have been described in detail elsewhere. MDD and BD diagnoses were assessed with commonly used standardized interviews.

GAIN-MDD. This sample consisted of subjects from two large-scale longitudinal studies, the NESDA (Netherlands Study of Depression and Anxiety) and the NTR (Netherlands Twin Register). The mean ages of cases ($N = 1738$) and controls ($N = 1802$) were 43 and 45 years, respectively. Inclusion criteria for MDD cases were a lifetime diagnosis of MDD2000 +. The second MDD target sample consisted of 2101 cases and 3280 screened controls, a subset of the MDD2000 + sample after excluding samples that overlapped with the discovery and GAIN-MDD samples. Samples were provided by the Queensland Institute of Medical Research (QIMR, Brisbane, QLD, Australia), NESDA, NTR, the University of Edinburgh (UoE, Scotland, UK) and the MGS (Molecular Genetics of Schizophrenia) study (controls only, United States). Control subjects from NTR who also participated in the GAIN-MDD study ($N = 223$) were excluded, as well as cases and controls from the NAG/IPRG study that had been included in the personality traits meta-analysis ($N = 500$). Mean ages of cases and controls were 41 and 48 years, respectively. MGS controls completed an online questionnaire including the short-form CIDI, supplemented by questions about schizophrenia, psychosis and BD. Controls were required to never have met the criteria for these disorders or MDD.

Genotyping was conducted on different Illumina and Affymetrix platforms. Imputation was conducted in four analysis sets (I317, I370, I610 and A6.0) to a common set of SNPs present in HapMap3 CEU/TSI, using Beagle 3.04.

Wellcome Trust Case–Control Consortium. The sample comprises 1868 BD cases. They were over the age of 16 years and of European descent. Individuals who had been in contact with mental health services were recruited if they suffered from a major mood disorder in which clinically significant episodes of elevated mood had occurred, such as bipolar I disorder (71% cases), schizoaffective disorder bipolar type (15% cases), bipolar II disorder (9% cases) and manic disorder (5% cases). Half of the 3000 controls came from the 1958 British Birth cohort (58C) and were between 44 and 45 years of age at the time of DNA collection. The other half was selected from blood donors. Age ranged from 18 to 69 years. Analyses were carried out on observed genotypes. Ancestry principal components were available for 3919 subjects (1452 cases and 2467 controls), and these subjects were included in the current analyses.

Systematic Treatment Enhancement Program for Bipolar Disorder. STEP-BD was a US national, longitudinal public health initiative designed to examine the effectiveness of treatments and their impact on the course of BD. Over a 7-year period, 4361 participants were enrolled across 20 sites and followed for up to 2 years. From the parent STEP-BD study, 2089 participants were enrolled in a genetic substudy. Mean age was 43 years. Only non-Hispanic Caucasian individuals with European ancestry based on self-reported race and ethnicity information were included. Controls were used from the NIMH (National Institute of Mental Health) Genetics Repository and were MGS controls. This study included 1507 cases with bipolar I and bipolar II disorder and 903 controls.

Genotyping was performed using the Affymetrix GeneChip Human Mapping 500K Array Set. Beagle, version 3.1.1 was used to impute missing genotypes, with HapMap2 (Centre d’Etude du Polymorphisme Humain from Utah population, release 23, forward strand) as the reference panel.

Statistical analyses

GWA meta-analyses of personality. Genome-wide association meta-analyses were carried out as described in det al
traits measured with the NEO Five-Factor Inventory\(^2\) in GAIN-MDD controls, GAIN-MDD cases and STEP-BD cases at entry to the study. Four of the five personality scales were significantly different in MDD patients compared with controls: N scores were higher, whereas A, E and C scores were lower in depressed subjects (all P-values < 0.0001 in a t-test). As described in Barnett et al.,\(^8\) BD patients at a euthymic phase significantly differed from the population norms for all scales exhibiting higher levels of N and O and lower levels of E, C and A. The sets of SNPs with P-values below 0.1, 0.2, 0.3, 0.4 and 0.5 based on the results in the discovery sample included the number of SNPs as expected for these P-values, that is, ~200 k, 400 k, 600 k, 800 k and 1 million SNPs, respectively. The overlap with SNPs used to calculate polygenic scores in the target samples was 100% for the GAIN-MDD sample, 40% for the MDD2000 + sample, 20% for the WTCCC sample and 90% for the STEP-BD sample. The lower overlap is due to the use of HapMap 3 as a reference set in the MDD2000 + sample and to the use of observed genotypes only in the WTCCC sample.

Figure 1 presents the results of the logistic regression analysis for MDD with the personality polygenic scores based on the GWA meta-analysis as independent variables. Significant positive associations were found between MDD and polygenic N scores for all sets of SNPs. In addition, significant negative associations were found between MDD and polygenic E scores based on the sets of SNPs with P-values < 0.1 and < 0.5 and polygenic C scores based on the sets of SNPs with P-values < 0.3 to < 0.5. These effects explained 0.1% of the variance in MDD (P-value < 0.05).

Figure 2 shows results of the logistic regression analysis for BD. Polygenic E scores were significantly positively related to BD, explaining 0.1% of the variance (P-value < 0.05). The analyses were also carried out in the four samples separately to check for heterogeneity in the results across studies. None of the polygenic personality scores were significantly associated with MDD in the MDD2000 + study with the proportion of explained variance always below 0.1%. This is in contrast to the analyses of the GAIN-MDD sample in which polygenic N scores based were significantly higher in MDD cases than in controls, explaining between 0.2 and 0.4% of the variance (P-values < 0.005) (see Supplementary Figure 1a). Moreover, polygenic C scores, based on the sets of SNPs with P-values < 0.3 and < 0.5, were lower in MDD cases explaining 0.2 and 0.3% of the variance (P-values...
The genetic relationship between personality and BD has not been investigated before. With the exception of E, the absence of a genetic association between personality traits and BD is surprising given the phenotypic relationships between BD and the five personality traits found in the STEP-BD sample (Table 1) and other studies. Furthermore, the genetic association with E was positive, whereas the phenotypic association found in the total sample of STEP-BD subjects was negative. However, Barnett et al. showed in the same study sample that high N was related to a depression-prone BD course, whereas high E was related to a manic-prone BD course. Similar findings were reported by Quilty et al.

The explained variance in MDD and BD by the polygenic scores was very modest. However, this was also true for prediction of the personality traits themselves with comparable proportions of explained variance (up to 0.4%) and P-values. Thus, the proportions of explained variance for mood disorders are at the upper limit of what could have been expected. Still, these numbers are lower than the explained variance of ~2% reported for polygenic schizophrenia scores predicting BD and polygenic MDD scores predicting anxiety disorders. This can be partly explained by a lack of power. Although the discovery sample was large with >13,000 subjects, only a few SNPs reached genome-wide significance and the deviation of the line of observed P-values from the line of expected P-values in the QQ plots is also modest. As the individual effect sizes of SNPs are small, the error of the estimates is relatively large. Therefore, the explained variance in a prediction analysis is low. This is comparable to the results of the analyses of GWA data for intelligence, which showed that the proportion of variance explained by all SNPs varied between 40 and 51%, whereas prediction analyses only explained 1% of the variance in intelligence.

An additional explanation could be the large age differences in the studies included in the discovery set, resulting in top hits in the meta-analysis that reflect genetic variants associated with stability in personality traits from adolescence through older age. It is possible that genetic risk factors influencing stability over time are less related to MDD and BD than genetic risk factors for personality that are mostly important around early adulthood, the period of onset of MDD and BD. However, longitudinal twin studies suggest that genetic influences on personality are for the largest part stable, thus without much change in genetic risk factors across time (see, for example, Kandler et al. for a study in adults and Gillespie et al. and Hopwood et al. for studies in adolescents and young adults). A strength of our study, on the other hand, is that polygenic scores were determined from GWA study results in a sample of individuals without mood disorders and are not confounded by mood state in the cases.

The STEP-BD sample and the MDD2000+ sample both included MGS controls. That is no problem as the discovery sample and target samples were not overlapping.

Analyses of the effects of the polygenic personality scores on mood disorders in the four separate studies indicated that for BD, results were consistent over studies, whereas for MDD, results were mainly driven by the GAIN-MDD study. There does not seem to be an obvious explanation for the absence of effect in the MDD2000+ sample. The overlap in
SNPs used to calculate polygenic scores was far lower in the MDD2000 + sample than in the GAIN-MDD sample because of the use of different reference sets for the imputation. HapMap3 in MDD2000 + and HapMap2 in MDD-GAIN. HapMap3 includes less SNPs but is based on more subjects. However, the overlap in SNPs was also lower in the WTCCC sample, whereas their results were similar to the results in the STEP-BD sample, which had a higher overlap in SNPs. Therefore, the low overlap in SNPs does not seem to explain the difference in results between the MDD2000 + and the MDD-GAIN sample. Given the repeatedly found genetic correlation between N and MDD, it seems most likely that the finding in MDD2000 + is a false negative finding.

Despite the low explained variance, the results of this study indicate some interesting issues regarding the etiology of MDD and especially BD. Although studies investigating the phenotypic association between personality and MDD, as well as BD suggest that both disorders are related to high N and that, in addition, BD is related to high O, the genetic association shows a different picture of shared genetic risk factors for N and MDD on the one hand and for E and BD on the other. As previous studies have already pointed to differences in personality profiles between depression-prone and manic-prone BD patients, these results imply that BD is a heterogeneous disorder with different expressions related to different, genetically influenced, personality profiles. This view is supported by the finding that an association with polymorphisms in the GABA receptor $\gamma$1 subunit gene is most significant in cases fulfilling the Research Diagnostic Criteria for schizo-affective disorder. Cases fulfilling the Research Diagnostic Criteria for Bipolar Disorder type II showed a similar allele frequency as did controls. Future studies in larger samples, for example, the Psychiatric GWAS Consortium are suited to further investigate the complex larger samples, for example, the Psychiatric GWAS (NIMH, RO1 MH059160) and matching funds from participating institutes in the European Union (EU/WLRT-2001-01254), ZonMW (Geestkracht program, 10-000-1002), NIMH (RO1 MH059160) and matching funds from participating institutes in NEU and NTR. The NTR controls in MDD2000 + were genotyped in the Genomics platform (certified service provider (CSProR) for Illumina) at the LIFE and BRAIN Center, Bonn (funded by NWO-SPI 56-464-1419). Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the NWO (480-05-003). MDD and CM Middeldorp are financially supported by the Netherlands Organization for Scientific Research (NWO), ZonMW Addiction program, grant 31160089, VENI-016-115-035 and 016-115-035 (EuroSTRESS), the Wellcome Trust (grant no. 904-61-090, 904-61-193, 480-04-004, 400-50-397, 612-10-002, NIH (RO1 MH059160) and matching funds from participating institutes in

Confidential interest

Dr Laura Bierut is listed as an inventor on a patent, ‘Markers of Addiction,’ covering the use of certain SNPs in diagnosing, prognosing and treating addiction. Dr Bierut served as a consultant to Pfizer in 2008. Paul Costa receives royalties from the NEO Inventories. The other authors declare no conflict of interest.

Acknowledgements. This study makes use of data generated by the Wellcome Trust Case–Control Consortium. A full list of the investigators who contributed to the generation of the WTCCC data is available from http://www.wtccc.org.uk. Funding for the WTCCC project was provided by the Wellcome Trust under award 076113. NIMSDC/NTF: Funding support was provided by the Netherlands Scientific Research (904-61-090, 904-61-193, 480-04-004, 400-05-017, 912-109-20) Centre for Medical Systems Biology (NWO Genomics), the Neuroscience Campus Amsterdam (NCA) 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 the National Health and Medical Research Council (NHMRC) Fellowship Scheme. NRWay and DR Nyholt are supported by the Australian Research Council Future Fellowship Scheme. The ERF study was supported by grants from the Netherlands Organization for Scientific Research (NWO), Erasmus MC and the Netherlands Genomics Initiative (NGI)-sponsored Center of Medical Systems Biology (CMSB). Funding support for the Study of Addiction: Genetics and Environment through the National Institutes of Health and Health Initiative (GEI) (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA08401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA09392) and the Family Study of Cocaine Dependence (FSCD; R01 DA013423, R01 TA019963). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse and the NIH contract ‘High throughput genotyping for studying the genetic contributions to human disease’ (HSIN268200782086C). The Collaborative Study on the Genetics of Alcoholism (COGA) Principal Investigators: B Porjesz, V Hesselbrock, H Edenberg, L Bierut. Includes 10 different centers: University of Connecticut (V. Hesselbrock); Indiana University (H. Edenberg, J Nurnberger Jr., T Foroud); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz); Washington University in St Louis (L. Bierut, A. Goate, J. Rice, K. Bucholz); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield); Southwest Foundation (L. Almasy), Howard University (R. Taylor) and Virginia Commonwealth University (D. Dick). A Parsian and M. Reilly are the NIAAA Staff Colleagues. We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, currently a consultant with COGA, P. Michael Conneally, Raymond Crowe and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA08401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). The Collaborative Genetic Study of Nicotine Dependence (COGEND) project is a collaborative research group and part of the NIDA Genetics Consortium. Lead investigators directing data collection are Laura Bierut, Naomi Breslau, Dorothy Hatsukami and Eric Johnson. We thank Heidi Kromrei and Tracey Richmond for their assistance in data collection. In memory of Theodore Reich, founding Principal Investigator of COGEND, we are indebted to his leadership in the establishment and nurturing of COGEND and acknowledge with great admiration his seminal scientific contributions to the field. COGEND is supported by the NIH grant P01CA89392 from the National Cancer Institute. SardiNIA: We acknowledge support from the Intramural Research Program of the NIH, National Institute on Aging. Funding was provided by the National Institute on Aging, NIH contract NO1-AG-1-2108 to the SardiNIA (ProgeNIA) team. HBCS: We acknowledge financial support from the Academy of Finland (grant no. 120315 and 129297) to EW, 1129457 and 1216965 to KR, 120386 and 125876 to JGE), the European Science Foundation (EuroSTRESS), the Wellcome Trust (grant no. 89061/Z/09/Z and 89062/Z/09/Z)
and the Signe and Ane Gyllenberg foundation. NAG/IRPG: This study is supported by NIH grants DA12854 (to PAFM), AA07728, AA07580, AA11998, AA13320 and AA13321 (to ACH); and grants from the Australian National Health and Medical Research Council; MLP is supported by DA019951. LBC36: We thank David Liewald and Paul Redmond for technical assistance; the study Secretary Paula Davies; Alan Gow, Michelle Taylor, Janie Corley, Caroline Brett and Caroline Cameron for data collection and data entry; nurses and staff at the Wellcome Trust Clinical Research Facility, where subjects were tested and at the genotyping was performed; staff at theLothian Health Board and staff at the SCRE Centre, University of Glasgow. The research was supported by a program grant from Research INTO Ageing. The research continues with program grants from Help the Aged/Age Concern (The Disconnected Mind). GWAS funding awarded by the Biotechnology and Biological Sciences Research Council (BBSRC) to UD and AT. ML is a Royal Society ofEdinburgh/Lloyds TSB Foundation for Scotland Personal Research Fellow. The study was conducted within the University ofEdinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, supported by the (BBSRC), Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC) and Medical Research Council (MRC), as part of the cross-council Lifelong Health and Wellbeing Initiative. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/). The ECDF is partially supported by the work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.edikt.org.uk). The ECDF is partially supported by the Economic and Social Research Council (ESRC) and Medical Research Council (BBSRC), Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC) and Medical Research Council (MRC), as part of the cross-council Lifelong Health and Wellbeing Initiative. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.edikt.org.uk). The ECDF is partially supported by the
Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)