Genetic risk profiles for depression and anxiety in adult and elderly cohorts

Ayse Demirkan, Brenda W.J.H. Penninx, Karin Hek, Naomi Wray, Najaf Amin, Yurii Aulchenko, Richard van Dyck, Eco de Geus, Albert Hofman, André Uitterlinden, et al.

HAL Id: hal-00549430

https://hal.science/hal-00549430

Submitted on 22 Dec 2010

Genetic risk profiles for depression and anxiety in adult and elderly cohorts

Ayşe Demirkan[1], MSc, Brenda WJH Penninx[2], PhD, Karin Hek[1], MSc, Naomi R Wray[3], PhD, Najaf Amin[1], MSc, Yurii S Aulchenko[1], PhD, Richard van Dyck[2], MD, PhD, Eco JC de Geus[4], PhD, Albert Hofman[1], MD, PhD, André G Uitterlinden [1], PhD, Jouke-Jan Hottenga[4], PhD, Willem A Nolen[5], MD, PhD, Ben Oostra[1], PhD, Patrick F Sullivan[6], MD, Gonneke Willemsen[4], PhD, Frans G Zitman[7], MD, PhD, Henning Tiemeier[1], PhD, A. Cecile J.W. Janssens[1], PhD, Dorret I Boomsma[4], PhD, Cornelia M van Duijn[1], PhD, Christel M Middeldorp MD, PhD [4,8]

[1]Erasmus University Medical Centre, Genetic Epidemiology Unit of the Departments of Epidemiology and Clinical Genetics, PO Box 2040, 3000 CA Rotterdam, The Netherlands, [2]Department of Psychiatry, VU University Medical Center, Amsterdam, the Netherlands, [3]Genetic Epidemiology Unit, Queensland Institute of Medical Research, Brisbane, Australia, [4]Department of Biological psychology, VU University Amsterdam, the Netherlands, [5]Department of Psychiatry, University Medical Center Groningen, The Netherlands [6]Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA, [7]Department of Psychiatry, Leiden University Medical Center, the Netherlands, [8]De Bascule, Academic Center for Child and Adolescent Psychiatry, Amsterdam. The Netherlands

Running title: common genetic variants in depression and anxiety

Correspondence to:

Prof. Dr. C.M. van Duijn, Genetic Epidemiology Unit Ee2200a, Department of Epidemiology, Erasmus MC University Medical Centre Rotterdam, PO Box 2040, 3000 CA, Rotterdam, The Netherlands.
Tel: +31 10 7043391; Fax: +31 10 7044657; E-mail: c.vanduijn@erasmusmc.nl

Abstract

The first generation of genome-wide association studies (GWA studies) for psychiatric disorders has led to new insights regarding the genetic architecture of these disorders. We now start to realize that a larger number of genes, each with a small contribution, are likely to explain the heritability of psychiatric diseases. The contribution of a large number of genes to complex traits can be investigated with genome-wide profiling. In a discovery sample a genetic risk profile for depression was defined based on a GWA study of 1738 adult cases and 1802 controls. The genetic risk scores were tested in two population based samples of elderly participants. The genetic risk profiles were evaluated for depression and anxiety in the Rotterdam Study cohort and the Erasmus Rucphen Family (ERF) study. The genetic risk scores were significantly associated to different measures of depression and explained up to ~0.7% of the variance in depression in Rotterdam Study and up to ~1 % in ERF study. The genetic score for depression was also significantly associated to anxiety explaining up to 2.1% in Rotterdam study. These findings suggest the presence of many genetic loci of small effect that influence both depression and anxiety. Remarkably, the predictive value of these profiles was as large in the sample of elderly participants as in the middle-aged samples.

Introduction

Genetic factors play an important role in the susceptibility to depression. A meta-analysis of twin studies on Major Depressive Disorder (MDD) estimated the heritability at 37%.[1] However, the success of studies aiming to find genes underlying the vulnerability for depression has been limited. An overview of promising results of linkage studies on MDD and neuroticism, a related personality trait, shows some overlap in regions of interest, but, so far, no single locus has been identified.[2] Candidate gene studies, mostly focusing on genes involved in neurotransmitter circuits or in reactions to stress, have also not been able to unambiguously identify a genetic variants explaining differences in the vulnerability for depression.[3, 4]

An important issue in research on the etiological factors of MDD has been the frequent comorbidity with anxiety disorders. In the National Comorbidity Survey Replication, 59% of the subjects with a lifetime diagnosis of MDD also fulfilled the criteria for a lifetime anxiety disorder diagnosis.[5] A review of twin and family studies indicated that this comorbidity might be explained by shared, mostly genetic factors.[6] Still, an overview of promising results of linkage studies of anxiety only showed one overlapping region of interest with MDD and neuroticism.[2]

The recent success of genome-wide association (GWA) studies has fuelled expectations on finding genes for MDD. One of the first GWA studies of depression showed evidence for the role of the presynaptic protein piccolo (PCLO) gene.[7, 8] Recently, this result was replicated with a p-value of $2*10^{-9}$ in a meta-analysis of the results in three population based samples, but not when five clinical samples were also included.[7, 9] However, the first GWA studies of MDD as well as those of other psychiatric phenotypes have also demonstrated that genome-wide significant findings are rare and explain a small part of heritability.[7, 10, 11] This might be due to a lack of power. The Genetic Association Information Network (GAIN)-MDD GWA study, for example including ~1700 cases and 1800 controls, had 80% statistical power to detect relative risks of 1.59, 1.40 and 1.35 with a p-value of $1*10^{-7}$, for minor allele frequencies of 0.10, 0.25

and 0.40. This is well comparable to other first generation GWA studies. However, the first results of GWA studies suggest that the strongest odds ratios may be lover than 1.2.[12] Another explanation for the scarce genome-wide significant findings could be that there is not a distinct number of genes for MDD with moderate to small risks but rather a large number of variations spread over the genome, each with small effects. Such a polygenic model predicts that the more markers are used, the better the disease is predicted and it implies that everybody carries risk variants but patients carry more than non-diseased people. We explored if a polygenetic component influences MDD implying that a large number of genetic variants are involved in explaining its heritability.

The evidence for a polygenic origin has recently been examined for schizophrenia and the hypothesis of a polygenic component was directly tested using GWA data.[13] In this approach, the joint effect of multiple SNPs are tested rather than the effects of individual SNPs. These individual SNPs are not required to reach a genome-wide significance level by themselves. This approach aims to test whether the genetic disease liability reflects, at least in part, the additive effect of a large number of variants spread across the genome whose joint action may be captured in a genome-wide genetic risk score. To obtain this risk score, a discovery set is used to select SNPs based on specific p-value thresholds (e.g. 0.0001, 0.001, 0.01 etc) from a genome-wide scan for the disease of interest. In the target samples, genetic risk scores are calculated for each individual for each set of SNPs. The selected SNPs will contain false positives but if they are enriched with true associated variants with low effect size then the genetic risk score might still be significantly associated to the disease in an independent sample. The problem is to distinguish truly associated SNPs from the false positives which occur massively around liberal p-value thresholds. In the schizophrenia study, the genetic risk scores based on the multiple SNPs in the discovery sample were associated to schizophrenia in three independent samples. The variance explained by the risk scores increased as more SNPs were included, i.e. risk scores based on

SNPs that had a p-value below 0.5 in the discovery sample explained more variance in the replication sample than the risk scores based on SNPs selected at p-value below 0.1.[13] Moreover, the genetic risk scores for schizophrenia were also significantly associated with bipolar disorder assessed in two samples suggesting that the genes influencing schizophrenia and bipolar disorder partly overlap.

The current study applied the risk score approach to investigate whether a polygenic component can be detected for depression and whether this polygenic component also influences anxiety. Since the study samples differ in age, differences in the effect of the polygenic component may indicate that the genetic factors influencing depression and/or anxiety differ across the lifespan. Twin studies have already shown that the relative influence of genetic factors for depression decreases with age[14-16], but that the genes influencing depression remain the same across the years. [17] This can be investigated directly in the current study.

The discovery set consisted of the GAIN-MDD sample, including 1 738 cases and 1 802 controls[7, 18] with over 400 000 SNPs genotyped. The target sets consisted of two independent Dutch samples. The first sample was based on the Rotterdam Study cohort and consisted of 178 patients with DSM-IV defined depressive disorder and 915 controls at low liability of depression as well as 222 cases for anxiety and 290 controls at low liability for anxiety. The second target sample was the Erasmus Rucphen Family (ERF) study in which symptoms of anxiety and depression were measured in 1886 participants. The subjects in the GAIN-MDD sample and the ERF sample were around 45 years of age. The Rotterdam sample was an elderly sample with a mean age of around 70. Height and intraocular pressure (IOP), phenotypes unrelated to psychiatric disorders, were additionally investigated to examine if the association with the genetic risk scores was specific to depression and anxiety.

Material and methods

Discovery Sample

The discovery sample consisted of subjects from two large-scale longitudinal studies: the Netherlands Study of Depression and Anxiety (NESDA)[19] and the Netherlands Twin Registry (NTR).[20] The chances of success of genetic risk score analyses depend primarily on the size of the discovery or training set. If the sample size is too small, the risk profiles will be based on random noise and are not expected to explain variance in the target set. To increase the chances of success, the power of the discovery set should therefore be maximized.[21] The size of the GAIN-MDD study made it more suitable to be used as the discovery set than the Rotterdam and ERF studies, which thus supplied the target samples.

NESDA and NTR studies were approved by the Medical Ethics Committee of all participating institutes. Collection of the phenotype data and quality control of the genotype data as well as the statistical methods are described in detail elsewhere.[7, 18] In brief, inclusion criteria for MDD cases were a lifetime diagnosis of DSM-IV MDD[22] assessed with the Composite International Diagnostic Interview[23], age of 18-65 years and self-reported western European ancestry. Persons who were not fluent in Dutch and those with a primary diagnosis of schizophrenia or schizoaffective disorders, obsessive-compulsive disorder, bipolar disorder or severe substance use dependence were excluded. Inclusion criteria for control subjects were availability of biological samples and survey data with assessments of depression, anxiety and neuroticism, no report of MDD at any measurement occasion and low genetic liability for MDD based on the survey data. In addition, controls and their parents were required to have been born in the Netherlands or Western Europe. Only one control per family was selected. The cases and controls were carefully matched on age and sex.

The genotypic data used in the discovery sample were part of one of the six initial GAIN studies sponsored by the Foundation for the NIH. [20] Individual genotyping was conducted by

Perlegen Sciences (Mountain View, CA, USA) using a set of four proprietary, high-density oligonucleotide arrays. The SNPs on these arrays were selected to tag common variation in the HapMap European and Asian panels. Of the 3 820 samples sent to Perlegen, genotypes were delivered for 3 761 samples (98.5%) of which 3 540 subjects passed quality controls and were available in the final analysis dataset including 1 738 MDD cases and 1 802 controls. The SNP quality control process and the precautions against population stratification are detailed in Sullivan et al.[7] A total of 427 037 SNPs on chromosome 1 to chromosome 22 met all selection criteria and were included in the final association analyses, which were performed in PLINK.[24]

Target Samples

Rotterdam Study

The Rotterdam Study is a prospective cohort study in the district Ommoord of Rotterdam. [25] In 1990 all inhabitants aged 55 years and over were invited and 7 983 persons agreed to participate. The Medical Ethics Committee of the Erasmus Medical Center, Rotterdam, approved the study. Written informed consent was obtained from all participants.

Ascertainment of depressive symptoms and incident depressive disorders was described previously. [26] Depression and anxiety symptoms were assessed with the Center for Epidemiological Studies Depression Scale (CES-D) and Hospital Anxiety and Depression Scale (HADS). The CES-D scale consists of 20 items with scores ranging from 0 to 60. A score of 16 or higher on the CES-D is considered indicative of a depressive disorder. The HADS-Depression (HADS-D) and HADS-Anxiety (HADS-A) scales each consist of 7 items with scores ranging from 0 to 21 with higher scores indicating more symptoms of depression. These questionnaires are valid and reliable self report measures of symptoms of depression.[27, 28] The HADS was assessed during the second visit in a randomly selected subgroup of individuals (N=2 231). Depression was measured with the CES-D 3 times during the follow-up.

Among 7 983 subjects who agreed to participate, 5 974 were successfully genotyped, 524 died before depression screening and 747 did not participate in depression screening. In the remaining sample, 587 persons scoring higher than 16 on the CES-D in the third or fourth visit were invited for a semi-structured interview with the Present State Examination [29] by a clinician. In addition, general practitioner records and specialist letters were surveyed actively for the occurrence of depression. Furthermore, physicians conducted repeated interviews to assess self-reported depression in the interval period. This effort identified 178 persons with current DSM-IV defined depressive disorder (145 MDD, 15 dysthymia and 18 depression-not otherwise specified cases) and eligible genotype data. The control group consisted of 915 persons, who scored in the lowest quartile (CES-D="0") on CES-D in the third visit (N=3 879) and who did not report any depressive symptoms during the follow up.

Anxiety disorders were assessed during the fourth visit in the total sample by trained lay-interviewers who conducted a slightly adapted version of the Munich Composite International Diagnostic Interview (M-CIDI) to obtain DSM-IV diagnoses of generalized anxiety disorder (GAD), panic disorder, agoraphobia, social phobia and specific phobia. Obsessive compulsive disorder and post-traumatic stress disorder were not included. The current sample is selected from the 2 779 persons who had valid M-CIDI assessment and genotype data. Out of 2 779, 222 persons were anxiety disorder cases. The control group consists of 290 persons who did not have any anxiety disorder and scored in the lowest quartile (HADS-A=0) on the HADS-A measured in 1 322 persons of the interviewed and sample with eligible genotype data during the second visit.

Genome-wide SNP data were available from the Illumina HumanHap550K array for all cases and controls. Data were excluded based on call rate < 97.5%, sex mismatch, excess autosomal heterozygosity, and outliers identified by the clustering analysis. MACH 1.0 software (v1.0.16)[30, 31] was used to impute to ~2.54 million SNPs based on the HapMap CEU phased haplotypes (release 22). SNPs included in imputation met thresholds of MAF >= 1%, HWE p-

value $\geq 10^{-6}$ and SNP call rate >=98.0%. GWA analysis of MDD was performed with Mach2Dat (logistic regression on allele dosage) using the GRIMP interface.[30, 32] Age and sex were included as covariates in the analysis.

ERF Study

The Erasmus Rucphen Family study is part of the Genetic Research in Isolated Population program. The study population essentially consists of one extended family of descendents from 20 related couples who lived in the isolate between 1850 and 1900 and had at least 6 children baptized in the community church. The detailed information about ERF isolate can be found elsewhere.[33-35] The Medical Ethical Committee of the Erasmus Medical Center, Rotterdam approved the study and informed consent was obtained from all participants.
Symptoms of depression and anxiety were assessed using the HADS and the CES-D [27, 28] in 2385 participants who also underwent an extensive medical examination.

Data on height and IOP were collected during the medical examination. The height of participants was determined using a stadiometer and bilateral IOP measurements were performed using Goldmann applanation tonometry.[36]

Among 2 385 persons with phenotypes, high density genotype data were available for 1 886 subjects. The genotype data were available for this population on 4 different genotyping platforms (Illumina 6k, Illumina 318K, Illumina 370K and Affymetrix 250K), which were merged and ~ 2.54 x 10^6 SNPs were imputed using MACH 1.0 software (v1.0.16)[30, 31], using build 36 HapMap (release 22) CEU population as reference. Within each genotyping batch, only SNPs showing a call rate >98%, MAF>1% and HWE p–value $> 10^{-6}$ were used for imputations.

Since the ERF study included related individuals, GWA analyses were performed using a mixed model by "mmscore" option in GenABEL software[37] which combines the FASTA method of Abecasis et al[38], and kinship matrix estimated from genotyped SNPs.[39]

Risk Score Profiling

The score profiling method tests the association of a genetic score variable that reflects a combined effect of a number of selected SNPs with a trait. For a more detailed description of the method, we refer to Purcell et al (2009).[13] In brief, SNPs were selected using the results from the GAIN-MDD GWA studies[7] (the "discovery sample"). These sets of SNPs were used to calculate the genetic scores in the target samples. SNPs were selected on the basis of their nominal p-value ($p_{discovery}$) for association with MDD in the discovery sample. Genetic risk scores were calculated for $p_{discovery}$ thresholds ranging from 0.00001 to 1.0.

Only those SNPs were included that were directly genotyped in the discovery set (N= 427 049 SNPs). To avoid ambiguity A/T –G/C SNPs were excluded. Since an A to T or G to C change will result in the same nucleotides on the opposite strands, this change might be missed during the genotype analysis. SNPs for which the quality of imputation had an $R^2<0.95$ in target samples were also excluded. After all quality checks and exclusions, a total of 181 582 SNPs that were available in both ERF and Rotterdam study samples were selected for calculations of genetic risk scores.

For each individual in the two independent target samples, the genetic score was calculated by multiplying the number of risk alleles per SNP (0, 1 or 2) with the log odds ratio, summed over all SNPs in the considered set of SNPs.[40] We calculated individual scores for each set of SNPs using the PLINK(v1.06) software.[24]

Logistic regression models were used to test the association of the individual genetic risk scores for depressive and anxiety disorders. Linear regression models were used to test the association between genetic risk scores and the total CES-D, HADS-D and HADS-A scores as well as for height and IOP. Sex and age were used as covariates. As an alternative control for

false positivity, 10% of the non-associated cluster of SNPs with $p_{discovery} > 0.9$ (N=1569 SNPs) in the discovery set was selected and used for computing the risk profile in both target samples.

For the Rotterdam Study regression analysis were performed in SPSS 15.0 for Windows (SPSS Inc., Chicago, IL, USA). As the ERF sample includes relatives, data are not independent, which can lead to biased estimates of standard errors and test statistics if this dependency between measures is not taken into account.[41] Association analysis of genetic risk score and the traits in ERF population were performed in SOLAR (Sequential Oligogenic Linkage Analysis Routines) 4.1.5 software package (Southwest Foundation for Biomedical Research, San Antonio, Texas, USA)[42] using the "polygenic" option to adjust for pedigree kinship. Among 1 886 people both genotyped and phenotyped, 1 697 were clustered into pedigrees (using the pedigree splitting algorithm PedCut [43]) and included in the family-based analysis. The remaining persons were not included in the analysis because they were also (distantly) related. The difference of the explained variance in the null and alternative model was considered as the variance explained by the genetic score. A genetic risk score with a p-value <0.05 in the model was considered as significantly associated to the trait.

Results

Table 1 shows the descriptive data for the case-control studies of GAIN-MDD and the Rotterdam Study, and Table 2 for the ERF study. Since in the Rotterdam study subjects were ascertained on the basis of age 55 or more, the mean age was 74 years. This was higher than in the GAIN-MDD and ERF study in which the mean ages were around 45 years. Level of education was higher in the GAIN-MDD sample than in the other two samples. In the target samples, subjects diagnosed with a depressive disorder or an anxiety disorder were more often women and were older. In the discovery sample cases and controls were matched based on age and sex.

In the ERF study, CES-D, HADS-D and HADS-A scores were highly correlated (r~0.7 pair wise for all three). Figure 1A shows the variance explained by the genetic risk scores in the logistic regression analyses performed in the Rotterdam Study using depressive disorder as dependent variable. The genetic score based on the first cluster of 6 SNPs ($p_{discovery} < 0.00001$) significantly explained 0.66% of the variance in depressive disorder in the Rotterdam Study (p=0.03). This association is explained in large part by a cluster of 3 SNPs (rs2715148, rs2522833, rs2522840) in the *PCLO* gene as after removing the *PCLO* SNPs in LD, the risk score was not significantly associated to depression in the target sample anymore. More importantly, the scores based on SNPs with $p_{discovery} < 0.1$ to $p_{discovery} < 0.4$ were associated to depressive disorder in the Rotterdam Study explaining up to 0.65% of the variance, with a p< 0.05. As shown in Figure 1B, the Rotterdam Study anxiety disorder case-control sample analysis yielded the highest percentage of variance explained with the genetic risk scores from GAIN-MDD study. The risk scores based on SNPs with $p_{discovery} < 0.1$ to $p_{discovery} < 1.0$ significantly explained up to 2.1% of the variance (p=0.0025). For $p_{discovery}$ values of 0.1, 0.2 and 0.3, the percentage of variance increased from 1% to 2% when a higher number of SNPs were included in risk scoring.

Figure 2A, B and C show the linear regression results for the analysis using the

continuous scores on the CES-D, HADS-D and HADS-A in the ERF study. For CES-D, the

scores based on SNPs with $p_{discovery}$ values < 0.01 and 0.1 explained ~0.5% of the variance

(p=0.007 and p=0.008). For the HADS-D, the score based on SNPs with $p_{discovery}$<0.01

significantly explained 0.3% of the variance (p=0.03). The MDD based genetic score was also

significantly associated to anxiety measured with the HADS-A explaining up to 0.5% of the

variance (p=0.01).

To examine whether these results were due to chance, we tested whether the MDD based

genetic risk score predict also variation in height and IOP measured in ERF. Heritability of IOP

was 35% and 86% for height [36, 44] and none of the traits was correlated to depression or anxiety

in the ERF study. The genetic risk score for MDD failed to predict IOP and height (Figure 3A

and B) suggesting that this relation is specific to depression and anxiety. Moreover, a genetic

score of SNPs with $p_{discovery}$ ≤ 0.9 in the discovery set did not show significant association to any

of the phenotypes in the target samples (data not shown).

Discussion

The aim of this study was to investigate the genetic architecture of depression and the potential

overlap in genetic risk factors with anxiety. Due to the availability of an elderly cohort, it was

also possible to examine whether the genetic factors influencing anxiety and depression change

across the life span. Using genetic risk scores derived from the association results of the GAIN-

MDD study in two independent target samples we evaluated the evidence for a genome-wide

signature for several measures of depression and anxiety used as outcome variables in the target

samples. For depression, either diagnosed according to the DSM-IV or measured with the CES-D

or HADS-D, we could explain up to ~1% of the variance with the genetic risk scoring approach.

Moreover, the genetic risk scores for depression were also associated to anxiety explaining up to

2.1% of the variance when approximately half of the genome-wide SNP data were included in the score. The explained variance was highest in the elderly sample indicating that the genetic factors influencing anxiety and depression hardly change with age. No significant results were found for the control variables height and IOP, implying that the association of the genetic risk score with depression and anxiety does not reflect chance alone. Overall, these findings suggest the presence of many loci, each with a small effect influencing depression as well as anxiety.

We checked whether our results were only due to SNPs in high LD segregating together. We performed a strict LD pruning (200 SNPs sliding window with $r^2_{snp\text{-}snp}$ threshold of 0.25). Considering the CES-D scale in ERF study, percentage of variance explained by the risk scores based on SNPs with $p_{discovery}$<0.01 dropped slightly after LD pruning from 0.52 to 0.49 but remained significant (p=0.01) whereas a less strict pruning with an $r^2_{snp\text{-}snp}$ threshold of 0.50 improved the percentage of explained variance to 0.62 (p=0.0003). Results with HADS-A and HADS-D scales were similar which shows that LD pruning itself does not add a major difference to the method (data not shown). Excluding the SNPs with minor allele frequency < 0.05 did not change the explained variance. This finding suggest the common disease common variant hypothesis is explaining MDD heritability, on the other hand, the power to detect the effect of rare variants in the discovery and target sets, was low and such rare variants may be detected by other approaches such as linkage or deep sequencing.

Our results are in agreement with the results from the International Schizophrenia Consortium (ISC)[13] that pointed out a polygenic component influencing schizophrenia as well as bipolar disorder. There was a somewhat higher amount of explained variance for schizophrenia (3.2% compared to %1). This may be due to power issues such as differences in sample size (~3300 cases for ISC vs ~1800 cases for GAIN-MDD), MDD being a common disease with clear non-genetic influence due to life events, and lower heritability compared to schizophrenia (~40% vs ~ 80%).

The percentage of explained variance in anxiety (2.1%) supports the idea of shared genetic background between these disorders. This has already been suggested by twin studies [45] and is confirmed by our results. The trend of an increase in R^2 for anxiety with different $p_{discovery}$ thresholds is different from the trend that we observe in depression, pointing out that the effect sizes are different, but the direction of effect is the same. We would like to stress that difference in explained variance between the target samples can evenly well be explained by chance. Moreover, it is important to note that 70% of the GAIN-MDD cases had a co-morbid lifetime anxiety diagnosis. However, this high co-morbidity is exactly what is expected if two disorders are influenced by similar genes and diagnoses are not mutually exclusive. Future research, preferably with a more balanced proportion of pure depressed and co-morbid cases, can shed more light on the overlap in genetic factors influencing anxiety and depression.

A limitation of the current study was that there were some differences between the discovery set and the target samples. Different instruments were used to measure depression and anxiety. In the GAIN-MDD Study, the CIDI was used to diagnose MDD and anxiety disorders, while in the Rotterdam study, the PSE was used. However, both instruments aim to make diagnoses according to the DSM-IV criteria and have adequate agreement for overall syndromes. [46] In the ERF population, symptoms of depression and anxiety were measured using the CES-D and HADS. Several validation studies on various types of patients using different diagnostic tools have shown that HADS performs well in assessing the symptom severity and case status of anxiety disorders and depression in both somatic, psychiatric and primary care patients and in the general population. [47] The HADS-D subscale has shown high sensitivity (~0.9) and specificity (~0.7) for MDD as diagnosed by DSM-IV in various studies. [48] The CES-D scale was found to be satisfactory in a semi-clinical sample of the elderly and in general population (sensitivity=0.9 and specificity =0.6) for life-time MDD and also performed excellent for one month of prevalence of MDD as diagnosed by DSM-IV (sensitivity=1.0 and specificity =0.9) among

elderly Dutch. [49-51] Considering the HADS-A subscale, the sensitivity and specificity for DSM-IV GAD was reported to be 0.9 and 0.8 respectively. [52] In addition, the discovery set in the current study included lifetime MDD cases, whereas the Rotterdam study recorded depressive disorders during a 9 year follow-up rather than lifetime MDD. Similarly, CES-D and HADS measure depressive and anxious symptoms in the last week. This means that subjects in the control groups in the target samples may be non-current but life-time MDD or anxiety cases. To summarize, although the measurements of anxiety and depression used in the three study samples are definitely related to each other, the fact that they are not entirely similar implies some heterogeneity, biasing the results towards the null hypothesis.

Another point involves the difference in gender ratios between discovery and target samples. In the discovery sample, the cases and controls were carefully matched on age and sex. Meta-analysis of twin studies suggests that genetic factors that influence depression are mostly shared between men and women [1, 53, 54] Sex was also used as a covariate when predicting depression or anxiety in the target samples. Thus, it seems unlikely that the gender ratio may have a major impact in the replication of the findings.

There was also heterogeneity in education level as a measure of socio-economical status. In spite of these differences, we still found a significant effect of the genetic risk score suggesting that the effects of the risk scores are actually even stronger. In both the ISC study and the current studies the low variance explained compared to the heritability of the disorders will also reflect that the analyses did not include the X chromosome, that gene-gene or gene-environment interactions are not considered and that the current generation of genotyping platforms do not fully tag genomic variance. [55]

This study is the second study showing direct evidence for a polygenic component influencing the susceptibility for a psychiatric disorder as well as overlap in genetic risk factors with another psychiatric condition. In addition, the current study suggests that the genetic factors

influencing anxiety and depression hardly change with age. The results imply that causal SNPs or the SNPs in LD with such variants do exist, but have lower effect sizes than the first generation of genome-wide association studies on psychiatric disorders was powered to detect. This provides optimism that variants associated at genome-wide levels of significance will be detectable as sample sizes increase in the next generation of GWA studies and their meta/mega-analyses. Moreover, it confirms that genome-wide profiling is a useful approach to investigate the genetic architecture of disorders, i.e. similarities and differences in genetic factors influencing several disorders or influencing the same disorder across the lifespan.

Conflict of interest

None of the authors declare conflict of interest

REFERENCES

1   Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000; **157**(10)**:** 1552-1562.

2   Middeldorp C, Boomsma D. Genetics and Psychopathology. In: Berntson GG CJ (ed). *Handbook of Neuroscience for Behavioral Sciences*. Wiley: New York, 2009.

3   Levinson DF. The genetics of depression: a review. *Biol Psychiatry* 2006; **60**(2)**:** 84-92.

4   Lopez-Leon S, Janssens AC, Gonzalez-Zuloeta Ladd AM, Del-Favero J, Claes SJ, Oostra BA *et al.* Meta-analyses of genetic studies on major depressive disorder. *Mol Psychiatry* 2008; **13**(8)**:** 772-785.

5   Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Jama* 2003; **289**(23)**:** 3095-3105.

6   Middeldorp CM, Cath DC, Van Dyck R, Boomsma DI. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychol Med* 2005; **35**(5)**:** 611-624.

7   Sullivan PF, de Geus EJ, Willemsen G, James MR, Smit JH, Zandbelt T *et al.* Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry* 2009; **14**(4)**:** 359-375.

8   Bochdanovits Z, Verhage M, Smit AB, de Geus EJ, Posthuma D, Boomsma DI *et al.* Joint reanalysis of 29 correlated SNPs supports the role of PCLO/Piccolo as a causal risk factor for major depressive disorder. *Mol Psychiatry* 2009; **14**(7)**:** 650-652.

9   Hek K, Mulder CL, Luijendijk HJ, van Duijn CM, Hofman A, Uitterlinden AG *et al.* The PCLO gene and depressive disorders: replication in a population-based study. *Hum Mol Genet* 2009.

10   Psychiatric GCCC, Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV *et al.* Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry* 2009; **166**(5)**:** 540-556.

11   Muglia P, Tozzi F, Galwey NW, Francks C, Upmanyu R, Kong XQ *et al.* Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry* 2008.

12   Craddock N, O'Donovan MC, Owen MJ. Genome-wide association studies in psychiatry: lessons from early studies of non-psychiatric and psychiatric phenotypes. *Mol Psychiatry* 2008; **13**(7)**:** 649-653.

13   International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**(7256)**:** 748-752.

14   Baldwin RC, Tomenson B. Depression in later life. A comparison of symptoms and risk factors in early and late onset cases. *Br J Psychiatry* 1995; **167**(5)**:** 649-652.

15   Jansson M, Gatz M, Berg S, Johansson B, Malmberg B, McClearn GE *et al.* Gender differences in heritability of depressive symptoms in the elderly. *Psychol Med* 2004; **34**(3)**:** 471-479.

16   Johnson W, McGue M, Gaist D, Vaupel JW, Christensen K. Frequency and heritability of depression symptomatology in the second half of life: evidence from Danish twins over 45. *Psychol Med* 2002; **32**(7)**:** 1175-1185.

17   Gillespie NA, Kirk KM, Evans DM, Heath AC, Hickie IB, Martin NG. Do the genetic or environmental determinants of anxiety and depression change with age? A longitudinal study of Australian twins. *Twin Res* 2004; **7**(1)**:** 39-53.

18   Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D *et al.* Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet* 2008; **16**(3)**:** 335-342.

19   Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res* 2008; **17**(3)**:** 121-140.

20   Boomsma DI, de Geus EJ, Vink JM, Stubbe JH, Distel MA, Hottenga JJ *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* 2006; **9**(6)**:** 849-857.

21   Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 2007; **17**(10)**:** 1520-1528.

22   Association AP. *Diagnostic and statistical manual of mental disorders 4th edition*. 4 edn. APA, 1994

23   World Health Organization. *Composite International Diagnostic Interview (version 2.1)* WHO: Geneva, 1992.

24   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**(3)**:** 559-575.

25   Hofman A, Breteler MM, van Duijn CM, Janssen HL, Krestin GP, Kuipers EJ *et al.* The Rotterdam Study: 2010 objectives and design update. *European journal of epidemiology* 2009.

26   Luijendijk HJ, van den Berg JF, Dekker MJ, van Tuijl HR, Otte W, Smit F *et al.*
      Incidence and recurrence of late-life depression. *Archives of general psychiatry* 2008;
      65(12)**:** 1394-1401.

27   Radloff. The CES-D scale: a self report depression scale for research in theb general
      population. *Appl Pshycol Measurement* 1977; **3:** 385-401.

28   Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*
      1983; **67**(6)**:** 361-370.

29   Wing J, Cooper, JE and Sartorius, N. Measurement and Classification of Psychiatric
      Symptoms. An Instruction Manual for the PSE and CATEGO Programme. Cambridge
      University Press: Cambridge, 1974.

30   Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Proceedings
      of the American Society of Human Genetics*; 9–13 October 2006 2006; New Orleans,
      Louisiana.

31   Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A. A comprehensive
      evaluation of SNP genotype imputation. *Hum Genet* 2009; **125**(2)**:** 163-171.

32   Estrada K, Abuseiris A, Grosveld FG, Uitterlinden AG, Knoch TA, Rivadeneira F.
      GRIMP: A web- and grid-based tool for high-speed analysis of large-scale genome-wide
      association using imputed data. *Bioinformatics* 2009.

33   Lopez-Leon S, Chi Choy W, Aulchenko YS, Claes SJ, Oostra BA, Mackenbach JP *et al.*
      Genetic Factors Influence the Clustering of Depression among Individuals with Lower
      Socioeconomic Status. *PLoS ONE* 2009; **4**(3)**:** e5069.

34   Santos R, Zillikens M, Rivadeneira F, Pols H, Oostra B, van Duijn C *et al.* Heritability of
      fasting glucose levels in a young genetically isolated population. *Diabetologia* 2006;
      49(4)**:** 667-672.

35   Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N *et al.*
      Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet*
      2004; **12**(7)**:** 527-534.

36   van Koolwijk LME, Despriet DDG, van Duijn CM, Pardo Cortes LM, Vingerling JR,
      Aulchenko YS *et al.* Genetic Contributions to Glaucoma: Heritability of Intraocular
      Pressure, Retinal Nerve Fiber Layer Thickness, and Optic Disc Morphology. *Investigative
      Ophthalmology Visual Science* 2007; **48**(8)**:** 3669-3676.

37   Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-
      wide association analysis. *Bioinformatics* 2007; **23**(10)**:** 1294-1296.

38   Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense
      genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**(1)**:** 97-101.

39   Amin N, van Duijn CM, Aulchenko YS. A genomic background based method for
      association analysis in related individuals. *PLoS One* 2007; **2**(12)**:** e1274.

40   Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet* 2009; **18**(18)**:** 3525-3531.

41   Rebollo I, de Moor MH, Dolan CV, Boomsma DI. Phenotypic factor analysis of family data: correction of the bias due to dependency. *Twin Res Hum Genet* 2006; **9**(3)**:** 367-376.

42   Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**(5)**:** 1198-1211.

43   Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS. An approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet* 2008; **16**(7)**:** 854-860.

44   Axenovich TI, Zorkoltseva IV, Belonogova NM, Struchalin MV, Kirichenko AV, Kayser M *et al.* Linkage analysis of adult height in a large pedigree from a Dutch genetically isolated population. *Hum Genet* 2009.

45   Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry* 2003; **60**(9)**:** 929-937.

46   Andrews G, Peters L. The psychometric properties of the Composite International Diagnostic Interview. *Soc Psychiatry Psychiatr Epidemiol* 1998; **33**(2)**:** 80-88.

47   Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res* 2002; **52**(2)**:** 69-77.

48   Silverstone PH. Concise assessment for depression (CAD): a brief screening approach to depression in the medically ill. *J Psychosom Res* 1996; **41**(2)**:** 161-170.

49   Beekman AT, Deeg DJ, Van Limbeek J, Braam AW, De Vries MZ, Van Tilburg W. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. *Psychol Med* 1997; **27**(1)**:** 231-235.

50   Donker T, van Straten A, Marks I, Cuijpers P. Brief self-rated screening for depression on the Internet. *J Affect Disord* 2009.

51   Haringsma R, Engels GI, Beekman AT, Spinhoven P. The criterion validity of the Center for Epidemiological Studies Depression Scale (CES-D) in a sample of self-referred elders with depressive symptomatology. *Int J Geriatr Psychiatry* 2004; **19**(6)**:** 558-563.

52   Olsson I, Mykletun A, Dahl AA. The Hospital Anxiety and Depression Rating Scale: a cross-sectional study of psychometrics and case finding abilities in general practice. *BMC Psychiatry* 2005; **5:** 46.

53   Kendler KS, Gatz M, Gardner CO, Pedersen NL. A Swedish national twin study of lifetime major depression. *Am J Psychiatry* 2006; **163**(1)**:** 109-114.

54   Middeldorp CM, Birley AJ, Cath DC, Gillespie NA, Willemsen G, Statham DJ *et al.* Familial clustering of major depression and anxiety disorders in Australian and Dutch twins and siblings. *Twin Res Hum Genet* 2005; **8**(6)**:** 609-615.

55   Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009; **461**(7261)**:** 272-276.

Table 1

Descriptive data of case-control samples

	GAIN-MDD		Rotterdam Study DD		Rotterdam Study AD	
	Cases (N=1 738)	Controls (N=1 802)	Cases (N=178)	Controls (N=915)	Cases (N=222)	Controls (N=290)
Age, mean (s.d.)	42.6 (12.6)	45.1 (14.1)	67.7 (6.8)*	64.8 (6.5)*	75.4 (5.82)	74.6 (5.32)
Women (%)	69.6	62	75.7	43.5	78	43
Education (%)						
Elementary	7.8	5.7	41.8	22.7	37.2	19.4
Intermediate	62	56.3	53.7	64.5	56.4	64.6
Higher	32.2	38.1	4.5	12.8	6.4	16.0
Antidepressant medication (%)	34.5	0	12.4*	0*	12.1	2
Co-morbid AD / MDD (%)	69.9	0	35.8**	3.6***	10.8	0.7

AD: Anxiety disorder, DD: Depressive Disorder, *Recorded at baseline, **Data available in 108 out of 178 cases, ***Data available in 701 of 915 controls.

Table 2

Descriptive data of ERF study

	Mean (sd)
CES-D	10.6 (9.6)
HADS-D	5.9 (4.3)
HADS-A	6.7(4.5)
Age, mean (sd)	48.2 (14.7)
Women(%)	57.4
Education(%)	
Elementary	30.8
Intermediate	63.8
Higher	5.4
Antidepressant medication (%)	5.0

CES-D: Center for Epidemiological Studies Depression Scale. HADS-D: Hospital Anxiety and Depression Scale-Depression subscale. HADS-A: Hospital Anxiety and Depression Scale-Anxiety subscale.

Figure 1.

Percentage of variance explained by genetic risk scores in Rotterdam Study.

Percentage of variance represented as difference in Nagelkerke R^2 after adjustment for age and sex. A, Analyses based on comparison of DD persons (n=178) to persons scoring in the lowest quartile of CES-D (CES-D="0") scale and who did not report any depressive complaints during the follow-up (n=915). B, Analyses based on comparison of persons with anxiety disorder (n=222) to persons scoring in the lowest quartile of HADS-A (HADS-A="0") scale and who did not report any depression or anxiety symptoms during the follow-up (n=290). * p value <0.05, ** p value <0.001

Figure 2.

Percentage of variance explained in depression and anxiety symptoms in ERF study by the genetic risk scores.
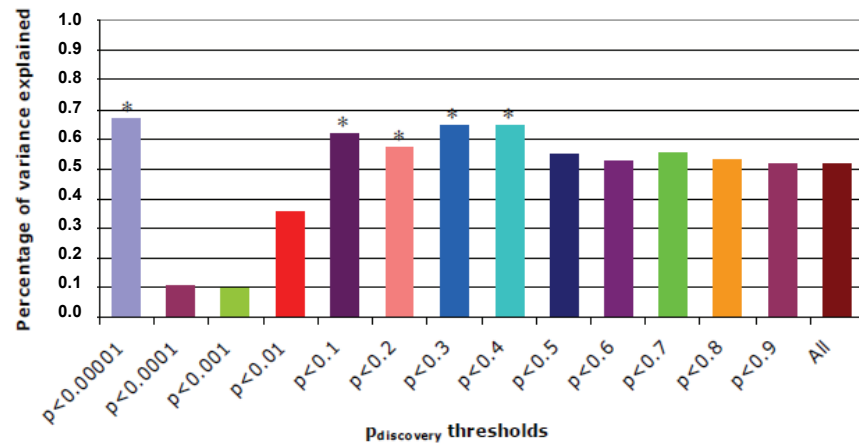
Percentage of variance represented as difference in R^2 after adjusting for age sex and family relations. A, B and C, Analyses of continuous scales. * p value <0.05, ** p value <0.01

Figure 3.

Predicting height and intraocular pressure in ERF study.

Linear regression analysis of height and Intraocular Pressure. Percentage of variance represented as difference in R^2 after adjusting for age sex and family relations.

A. Rotterdam Study-Depressive Disorder

B. Rotterdam Study-Anxiety Disorder

A. ERF Study CES-D scale

B. ERF Study HADS-D scale

C. ERF Study HADS-A scale

A. ERF Study-Adult height

Percentage of variance explained (y-axis: 0.0 to 1.0)

$p_{discovery}$ thresholds (x-axis): p<0.00001, p<0.0001, p<0.001, p<0.01, p<0.1, p<0.2, p<0.3, p<0.4, p<0.5, p<0.6, p<0.7, p<0.8, p<0.9, All

B. ERF Study- Intraocular pressure

Percentage of variance explained (y-axis: 0.0 to 1.0)

$p_{discovery}$ thresholds (x-axis): p<0.00001, p<0.0001, p<0.001, p<0.01, p<0.1, p<0.2, p<0.3, p<0.4, p<0.5, p<0.6, p<0.7, p<0.8, p<0.9, All