Molecular Psychiatry (2015) 20, 735–743

www.nature.com/mp

ORIGINAL ARTICLE

The association between lower educational attainment and depression owing to shared genetic effects? Results in ~25 000 subjects

WJ Peyrot[1], SH Lee[2], Y Milaneschi[1], A Abdellaoui[3], EM Byrne[2], T Esko[4,5], EJC de Geus[3], G Hemani[2,6], JJ Hottenga[3], S Kloiber[7], DF Levinson[8], S Lucae[7], Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium (Corporate Collaborator), NG Martin[9], SE Medland[9], A Metspalu[4,5], L Milani[4,5], MM Noethen[10], JB Potash[11], M Rietschel[12], CA Rietveld[13,14], S Ripke[15], J Shi[16], Social Science Genetic Association Consortium (Corporate Collaborator), G Willemsen[3], Z Zhu[2], DI Boomsma[3], NR Wray[2] and BWJH Penninx[1]

An association between lower educational attainment (EA) and an increased risk for depression has been confirmed in various western countries. This study examines whether pleiotropic genetic effects contribute to this association. Therefore, data were analyzed from a total of 9662 major depressive disorder (MDD) cases and 14 949 controls (with no lifetime MDD diagnosis) from the Psychiatric Genomics Consortium with additional Dutch and Estonian data. The association of EA and MDD was assessed with logistic regression in 15 138 individuals indicating a significantly negative association in our sample with an odds ratio for MDD 0.78 (0.75–0.82) per standard deviation increase in EA. With data of 884 105 autosomal common single-nucleotide polymorphisms (SNPs), three methods were applied to test for pleiotropy between MDD and EA: (i) genetic profile risk scores (GPRS) derived from training data for EA (independent meta-analysis on ~120 000 subjects) and MDD (using a 10-fold leave-one-out procedure in the current sample), (ii) bivariate genomic-relationship-matrix restricted maximum likelihood (GREML) and (iii) SNP effect concordance analysis (SECA). With these methods, we found (i) that the EA-GPRS did not predict MDD status, and MDD-GPRS did not predict EA, (ii) a weak negative genetic correlation with bivariate GREML analyses, but this correlation was not consistently significant, (iii) no evidence for concordance of MDD and EA SNP effects with SECA analysis. To conclude, our study confirms an association of lower EA and MDD risk, but this association was not because of measurable pleiotropic genetic effects, which suggests that environmental factors could be involved, for example, socioeconomic status.

Molecular Psychiatry (2015) **20,** 735–743; doi:10.1038/mp.2015.50; published online 28 April 2015

INTRODUCTION

An association between lower educational attainment (EA) and increased risk for major depressive disorder (MDD) has been confirmed in various western countries. A meta-analysis of 37 studies from mainly western countries found a 3% decrease in log odds ratio (OR) for depression per additional year of education.[1] Research of the World Mental Health Survey Initiative also found that those with high educational levels are generally at lower risk for depression in high-income countries, although Japan showed an inverted association.[2] The International Consortium of Psychiatric Epidemiology found a negative correlation in the United States and the Netherlands,[3] which was confirmed in a recent study in the Netherlands.[4]

The association of lower EA and increased MDD risk could result from multiple, not necessarily independent, effects: including causal, environmental or pleiotropic genetic effects. Lower EA could lead to an increased MDD risk (social causation), for example, via stress associated with lower socioeconomic status, or via less effective coping strategies or unhealthier lifestyles among those with lower EA.[5,6] However, lower EA could also be the result of MDD vulnerability, for example, when the onset of MDD is at an early age before educational goals would have been achieved. Alternatively, a third factor could be in play impacting on both, such as personality characteristics or less developed cognitive abilities, causing lower EA and increased risk for MDD. Such a third factor could also consist of pleotropic genetic effects (or linkage

[1]Department of Psychiatry, VU University Medical Center and GGZ inGeest, Amsterdam, The Netherlands; [2]The University of Queensland, Queensland Brain Institute, Brisbane, Queensland, Australia; [3]Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; [4]Estonian Genome Center, University of Tartu, Tartu, Estonia; [5]Division of Endocrinology and Center of Basic and Translational Obesity Research, Children's Hospital Boston, Boston; Department of Genetics, Harvard Medical School, Boston; Broad Institute, Cambridge, MA, USA; [6]MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, School of Social and Community Medicine, Bristol, UK; [7]Max Planck Institute of Psychiatry, Munich, Germany; [8]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA; [9]Genetic Epidemiology Unit, QIMR Berhgofer Institute of Medical Research, Brisbane, Queensland, Australia; [10]Institute of Human Genetics, University of Bonn, Bonn, Germany; [11]Department of Psychiatry, University of Iowa, Iowa City, IA, USA; [12]Department of Genetic Epidemiology in Psychiatry Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany; [13]Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands; [14]Erasmus University Rotterdam Institute for Behavior and Biology, Erasmus University Rotterdam, Rotterdam, The Netherlands; [15]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA and [16]Biostatistics Branch, Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, MD, USA. Correspondence: Dr WJ Peyrot, Department of Psychiatry, VU University Medical Center and GGZ inGeest, Amsterdam, AJ Ernststraat 1187, Amsterdam 1081, The Netherlands.
E-mail: w.peyrot@ggzingeest.nl
Received 27 November 2014; revised 28 February 2015; accepted 24 March 2015; published online 28 April 2015

Table 1. Sample characteristics

Study (abbreviation)	N *with MDD*		*Age (years)*		*% Female*	N *with EA*		*Mean EA z-score*	
	Case	*Control*	*Mean*	*S.d.*		*Case*	*Control*	*Case*	*Control*
Bonn/Mannheim (B)	925	1282	46.9	13.3	55.7	0	0	—	—
EGCUT (E)	508	5345	48.8	20.1	53.2	446	4569	−0.06	0.01
GenRED (GE)	976	1215	47.0	16.2	56.1	0	1170	—	0.00
GSK (GS)	866	863	51.9	13.5	66.9	866	862	−0.36	0.36
MPIP (M)	337	533	48.1	13.9	54.5	0	0	—	—
NESDA/NTR-1 (N1)	1560	1123	44.5	13.2	64.3	1382	875	−0.08	0.12
NESDA/NTR-2 (N2)	236	1201	40.5	14.7	62.4	211	759	−0.28	0.09
QIMR (Q)	1432	1686	43.4	10.9	61.0	1258	1402	−0.02	0.01
RADIANT (R)	1605	1573	44.3	12.5	66.6	0	0	—	—
STAR*D (S)	1217	128	43.5	14.0	56.9	1210	128	−0.03	0.30
Overall	9662	14949	46.2	15.6	59.4	5373	9765	−0.10	0.06

Abbreviations: EA, educational attainment; EGCUT, Estonian Genome Center of the University of Tartu; GEnRED, Genetics of Recurrent Early-Onset Depression; GSK, GlaxoSmithKline; MDD, major depressive disorder; MPIP, Max Planck Institute of Psychiatry; NESDA, Netherlands Study of Depression and Anxiety; NTR, Netherlands Twin Registry; QIMR, Queensland Institute of Medical Research; STAR*D, Sequenced Treatment Alternatives to Relieve Depression; YOB, year of birth. The number of cases with a diagnosis of MDD in lifetime, controls without MDD, mean age and its s.d. and the percentage of female is displayed for the 10 cohorts separately and for the overall sample. In addition, the number of cases and controls with information on EA available and their mean EA *z*-score are displayed. The EA *z*-scores were defined as the standardized residuals of the regression of EA on sex, YOB, YOB^2 and YOB^3, and the interaction of sex with YOB, YOB^2 and YOB^3.

disequilibrium between effective variants) resulting in genetic correlation (the part of the phenotypic correlation caused by shared additive genetic effects), because EA[7] and MDD[8–10] both have a confirmed genetic basis.

It is relevant to understand the mechanisms of the association between lower EA and MDD, because this can have important implications for prevention strategies of MDD and its consequences. When lower EA would increase MDD risk, the responsible mechanisms should be studied and subsequently addressed, for example, by providing psychoeducation about these mechanisms to those with lower EA. However, when shared genetic effects would link EA and MDD, no responsible mechanisms can be addressed, and prevention would be restricted to general advice to prevent MDD.

The possible impact of pleiotropic genetic effects on lower EA and increased MDD risk has not received much study. We are aware of three such studies, of which two find a substantial negative genetic correlation between EA and cross-sectional measures of depressive symptoms obtained via self-report questionnaires.[11,12] One study used DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition)-based diagnosis of MDD with a twin design, and generally supported the social causation model and found only a small genetic correlation.[5] To the best of our knowledge, no study combined DSM-IV-based diagnosis and genome-wide single-nucleotide polymorphisms (SNP) data to test for pleiotropic genetic effects between lower EA and MDD risk.

The current study was conducted to test for pleiotropic genetic effects between lower EA and MDD diagnoses in a large sample of ~25 000 subjects from the Psychiatric Genomics Consortium (PGC)[13] with additional Estonian and Dutch data. We applied the following SNP-based methods: genetic profile risk score (GPRS) analyses, bivariate genomic-relationship-matrix restricted maximum likelihood (GREML) analysis and SNP effect concordance analysis (SECA).

MATERIALS AND METHODS

Subjects

Genotype and phenotype data of 10 cohort studies were combined: eight cohorts[14–21] included in the PGC[13] plus two additional cohorts. The first additional cohort was from the Netherlands and combined additional independent data from the Netherlands Study of Depression and Anxiety[22] and the Netherlands Twin Registry[23] (NESDA/NTR-2). The second additional cohort was a population-based sample from Estonia (EGCUT).[24] The numbers of cases and controls per cohort are displayed in Table 1.

MDD cases and controls

All cases ($N = 9662$) had a DSM-IV- or ICD-10 (10th revision of the International Classification of Diseases)-based diagnosis of MDD in lifetime according to a structured diagnostic instrument. Most controls ($N = 14\,949$) were randomly selected from the population and screened for a lifetime history of MDD. A more detailed description of the PGC cohorts was given previously[13] and is summarized in Supplementary Table 1. For the NESDA/NTR-2 cohort, MDD cases were diagnosed with the DSM-IV-based CIDI (Composite International Diagnostic Interview, version 2.1), and controls scored low on various mental health screening questionnaires (NTR)[25] or had no diagnosis of a psychiatric disorder in their lifetime (NESDA). For the EGCUT (Estonian Genome Center of the University of Tartu) cohort, MDD cases were identified using ICD-10 codes F32 (depressive disorder) and/or F33 (recurrent depressive disorder), and MDD controls excluded all subjects with a lifetime ICD10 psychiatric diagnosis (category F).[24]

Educational attainment

EA was assessed in 7 of the 10 contributing cohorts (EGCUT, GenRED (Genetics of Recurrent Early-Onset Depression), GSK (GlaxoSmithKline), NESDA/NTR-1, NESDA/NTR-2, QIMR (Queensland Institute of Medical Research), and STAR*D (Sequenced Treatment Alternatives to Relieve Depression)). For NESDA/NTR-1 and NESDA/NTR-2, EA was defined as the years of education required for the highest diploma attained following the Dutch educational system. For QIMR and EGCUT, EA was defined as the US years of education required for the highest diploma attained following the ISCED (International Standard Classification of Education) classification.[7] For GSK, EA was defined as the number of years that school was attended. For STAR*D, EA was expressed in years of education. For GenRED, EA was assessed in controls only as the highest diploma attained and ranged from 1 to 5 labeling the following educational levels: (1) lower than high school, (2) high school, (3) some college, (4) bachelor degree and (5) higher than bachelor degree.

The EA measure was corrected per cohort for year of birth and sex, in line with the recent meta-analysis from the Social Science Genetic Association Consortium.[7] Thereby, the standardized residuals were obtained after regression of EA on sex, year of birth (YOB), YOB^2 and YOB^3, and the interaction of sex with YOB, YOB^2 and YOB^3. For STAR*D and GSK, YOB was not available and substituted with age. In all cohorts, EA was defined in individuals over 25 years of age only, so that they had time to

achieve their educational potential. The distribution of EA *z*-scores is displayed in Supplementary Figure 1.

Genotyping, quality control and imputation

Genotyping, quality control and imputation were performed in line with previous publications and are described in detail in the Supplementary Materials. In short, quality-controlled SNPs with a minor allele frequency >0.01 from the HapMap3 reference panel[26] were imputed and yielded information on 884 105 autosomal common SNPs. With these SNPs, the genomic relationship matrix was estimated and unrelated subjects selected (with maximum pairwise genetic relationships 0.05, which is approximately equivalent to second cousins), using the GCTA software (Brisbane, Queensland, Australia).[27] All of the subsequent genetic analyses were corrected for possible confounding cohort and genotyping effects by including a categorical covariate labeling the 10 cohorts, and within cohorts the different genotyping batches, where applicable (i.e. three batches within NESDA/NTR-2, two batches within EGCUT and two batches within QIMR). Ancestry-informative principal components were based on the genomic relationship matrix and estimated with the GCTA software.[27]

Genetic profile risk scores

Preparation of the GPRS based on EA discovery results (EA-GPRS) and MDD discovery results (MDD-GPRS) is described in detail in the Supplementary Materials. In short, the procedure from Purcell *et al.*[28] implemented in PLINK[29] was applied. The independent EA discovery results were from the recent meta-analyses on US years of schooling from the Social Science Genetics Association Consortium (SSGAC)[7] containing around 120 000 subjects. EA-GPRS analyses were not conducted for Bonn/Mannheim, GenRED and STAR*D, because no independent discovery results were available. To obtain the MDD discovery results was slightly more elaborate, because no large MDD cohort exists that is independent of PGC. Therefore, a 10-fold leave-one-cohort-out approach was followed, and the discovery results were thus based on around 8000 cases and 12 000 controls.

The GPRS were based on the same set of independent SNPs. First, the SNPs were selected with results available for all of the discovery sets. Second, this set of SNPs was pruned to a set of 76 516 independent SNPs with a maximum pairwise r^2 of 0.25 based on a sliding window of 200 SNPs with steps of 5 SNPs.[29] The EA-GPRS and MDD-GPRS were then estimated based on all SNPs up to *P*-value thresholds (P_T) in the respective discovery results of 0.001, 0.01, 0.1 and 1, respectively. Consequently, all GPRS with $P_T = 1$ were based on the exact same SNPs, but GPRS with different P_T were based on different sets of SNPs depending on the respective discovery results (see Supplementary Table 2). The GPRS were standardized per cohort to a mean of 0 and s.d. of 1 to aid interpretability of results.

Statistical analyses

The association of EA to MDD risk (phenotypic correlation) was assessed with logistic regression within EGCUT, GSK, NESDA/NTR-1 and -2, QIMR and STAR*D separately, and in the combined sample correcting for covariates labeling the cohorts.

GPRS analyses. In the first method to test for pleiotropic genetic effects, we estimated the *across-trait* effects of EA-GPRS on MDD and, *vice versa*, the effects of MDD-GPRS on EA. For comparison, we also estimated the *within-trait* effects of EA-GPRS on EA and MDD-GPRS on MDD. The effects of GPRS on EA and MDD were assessed with linear and logistic regression respectively. For the full sample, the effects were assessed for the GPRS based on P_T of 0.001, 0.01, 0.1 and 1; for the individual cohorts, the effects were only assessed for the GPRS based on $P_T = 1$.

The proportions of variation explained in EA and MDD were estimated as additional measures of the impact of GPRS. For EA, this proportion was derived as the R^2 of the linear regression model including the covariates and the polygenic risk score, minus the R^2 of the model including the covariates only. For MDD, Nagelkerke's pseudo R^2 were derived and corrected for the covariates by substituting the null (or intercept) model in Nagelkerke's equation for the model including the covariates (adjusted equation in Supplementary Materials). Lee *et al.*[30] indicated that Nagelkerke's pseudo R^2 can be biased by ascertainment, when the proportion of cases in the study sample differs from the population disease frequency. Therefore, they proposed an R^2 measure that is robust against ascertainment bias and interpretable on the liability scale. This liability R^2 was obtained by rescaling Nagelkerke's R^2 for an MDD population prevalence of $K = 0.2$ (see Supplementary Materials).[30]

Bivariate GREML. The GREML mixed linear model method was used (i) to assess the proportion of variation in EA and MDD explained by genome-wide common SNPs (SNP-h^2) and (ii) to assess the pleiotropic genetic effects between MDD and EA (genetic correlation), as implemented in GCTA.[27,31,32] The MDD SNP-h^2 was expressed on the liability scale for a population prevalence of $K = 0.2$ by converting the SNP-h^2 on the observed scale (controls 0; cases 1) with Equation (23) from Lee *et al.*[33] Bivariate GREML estimates of the genetic correlation are approximately the same on the liability scale as on the observed scale,[32] which implies that (i) its value does not depend on population disease prevalence *K* and (ii) that the genetic correlation between the binary MDD status and continuous EA measure could be estimated. The genetic correlation was, first, estimated with EA information from both cases and controls. This estimate could, however, potentially be confounded by case ascertainment (which may not be education independent). Therefore, the genetic correlation was estimated a second time with EA information from controls only and MDD status from both cases and controls. The GPRS and GREML analyses were corrected for sex, the first 10 (GPRS) or 20 (GREML) principal components and covariates labeling the cohorts and genotype batches. The necessity to correct for the principal components is indicated by a significant correlation between some of the GPRS with some of the principal components (Supplementary Table 3).

SNP effect concordance analysis. In SECA (http://neurogenetics.qimrberghofer.edu.au/SECA)[34] association results are analyzed, rather than individual genotyped data, to test for concordance between two traits with respect to the SNP effects significance as well as their directions. We applied SECA on the EA meta-analyses results from the SSGAC[7] and MDD association results on our own sample.

RESULTS

The overall sample consisted of 9662 patients with MDD in lifetime and 14 949 controls with a mean age of 46.2 (s.d. 15.6) and 59.4% female; information on EA was available for 5373 cases and 9765 controls (Table 1). In all cohorts with EA information available for both cases and controls, the phenotypic associations between EA and MDD was negative, with an overall OR of 0.78 (95% confidence interval: 0.75–0.82, $P = 2.2e-31$) per standard deviation increase in EA (Figure 1). This negative association was consistent for MDD cases with known age of onset >30 years. The strongest association was found in GSK with an OR of 0.45 (95% confidence interval: 0.40–0.50). When GSK was left out of the analyses, the overall association remained significant with an OR of 0.88 (95% confidence interval: 0.84–0.92). The association was comparable in male and female (Supplementary Figure 2).

GPRS analyses

The GPRS had *within-trait* predictive effects as expected. The MDD-GPRS predicted MDD with most predictive power for the

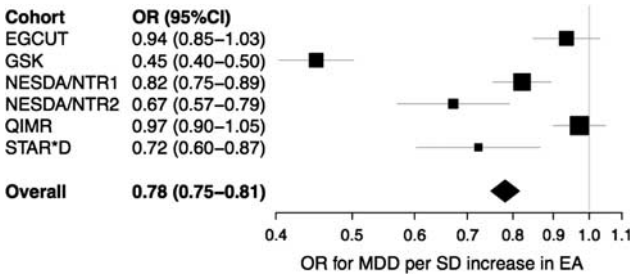

Figure 1. Forest plot of the phenotypic association between educational attainment (EA) and major depressive disorder (MDD). The odds ratio (OR) for MDD per s.d. increase in EA is displayed for the individual cohorts, as well as for the overall sample. The ORs were estimated with logistic regression of MDD on the corrected EA *z*-scores, which were defined as the standardized residuals of the regression of EA on sex, year of birth (YOB), YOB^2 and YOB^3, and the interaction of sex with YOB, YOB^2 and YOB^3.

Table 2. The effect of GPRS, based on depression (MDD-GPRS) and EA (EA-GPRS) discovery results, on MDD and EA in the overall sample (upper rows) and the separate studies (lower rows)

	Effect on MDD						*Effect on EA*			
	N		*Effect*		R^2 *(%)*		N	*β*	P-*value*	R^2 *(%)*
	Case	*Control*	*OR*	P-*value*	*NK*	*Liability*				
(A) Results in overall sample										
Effect of MDD-GPRS										
$P < 0.001$	9662	14 949	1.03	0.021	0.03	0.03	15 985	−0.01	0.188	0.01
$P < 0.01$	9662	14 949	1.04	0.010	0.04	0.04	15 985	−0.02	0.034	0.03
$P < 0.1$	9662	14 949	1.11	1.0e−11	0.28	0.29	15 985	−0.02	0.046	0.02
$P < 1$	9662	14 949	1.13	1.7e−16	0.41	0.42	15 985	−0.01	0.067	0.02
Effect of EA-GPRS										
$P < 0.001$	6544	12 324	1.02	0.167	0.02	0.02	13 477	0.05	1.2e−09	0.27
$P < 0.01$	6544	12 324	1.00	0.842	0.00	0.00	13 477	0.09	2.4e−23	0.73
$P < 0.1$	6544	12 324	0.98	0.245	0.01	0.01	13 477	0.10	2.1e−31	1.00
$P < 1$	6544	12 324	0.99	0.594	0.00	0.00	13 477	0.11	2.7e−37	1.20
(B) Results in separate studies										
Effect of MDD-GPRS (all threshold $P < 1$; letters represent separate studies)										
B	925	1282	1.14	0.003	0.54	0.54	—	—	—	—
E	508	5345	1.10	0.040	0.16	0.30	5015	−0.01	0.402	0.01
GE	976	1215	1.07	0.111	0.16	0.16	1170	−0.01	0.647	0.02
GS	866	863	1.09	0.097	0.22	0.21	1728	−0.04	0.111	0.14
M	337	533	1.05	0.461	0.08	0.09	—	—	—	—
N1	1560	1123	1.24	6.1e−08	1.50	1.50	2257	−0.02	0.373	0.04
N2	236	1201	1.19	0.019	0.70	0.98	1817	−0.02	0.330	0.05
Q	1432	1686	1.09	0.017	0.25	0.24	2660	−0.01	0.776	0.00
R	1605	1573	1.15	9.7e−05	0.65	0.63	—	—	—	—
S	1217	128	1.17	0.098	0.45	0.79	1338	0.03	0.345	0.07
Effect of EA-GPRS (all threshold $P < 1$; letters represent separate studies)										
E	508	5345	0.95	0.257	0.05	0.09	5015	0.08	1.2e−08	0.64
GS	866	863	0.96	0.358	0.07	0.06	1728	0.07	0.004	0.48
M	337	533	0.96	0.524	0.06	0.06	—	—	—	—
N1	1560	1123	0.94	0.151	0.10	0.10	2257	0.17	1.1e−15	2.79
N2	236	1201	0.95	0.512	0.05	0.08	1817	0.16	1.0e−11	2.52
Q	1432	1686	1.00	0.994	0.00	0.00	2660	0.12	1.7e−09	1.36
R	1605	1573	1.04	0.241	0.06	0.06	—	—	—	—

Abbreviations: B, Bonn/Mannheim; E, EGCUT; EA, educational attainment; EGCUT, Estonian Genome Center of the University of Tartu; GE, GenRED; GEnRED, GAIN-MDD cohort and the Genetics of Recurrent Early-Onset Depression; GPRS, genetic profile risk scores; GS, GSK; GSK, GlaxoSmithKline; M, MPIP; MDD, major depressive disorder; MPIP, Max Planck Institute of Psychiatry; N1, NESDA/NTR-1; N2, NESDA/NTR-2; NESDA, Netherlands Study of Depression and Anxiety; NTR, Netherlands Twin Registry; OR, odds ratio; Q, QIMR; QIMR, Queensland Institute of Medical Research; R, RADIANT MDD study; S, STAR*D; SNP, single-nucleotide polymorphism. The impact of the GPRS, based on EA discovery results (EA-GPRS) and MDD discovery results (MDD-GPRS), on target MDD and on target EA were estimated with, respectively, logistic and linear regression, while including sex, the first 10 principal components and covariates labeling the cohorts and genotype batches. The impact on MDD was, in addition, estimated as Nagelkerke's R^2 and the R^2 on the liability scale;[30] the impact on EA as the standard R^2 of linear regression. On the overall sample, the effects were estimated for GPRS on different sets of SNPs with different thresholds of significance in the discovery set ($P < 0.001$; $P < 0.01$; $P < 0.1$; $P < 1$) (panel A). The impact on MDD and EA in the separate cohorts was, subsequently, estimated for the polygenic risk scores based on all SNPs (threshold $P < 1$) (panel B). The number of individuals included in the analyses is displayed: note that individuals from B, GE and S are excluded from the analyses with polygenic risk scores based on EA discovery results, because these cohorts were (partly) included in the discovery phase.

polygenic risk score including all SNPs ($P_T = 1$), with an OR of 1.13 ($P = 1.7e-16$) and an R^2 of 0.4% on the liability scale (Table 2a). The EA-GPRS predicted EA also in the expected direction, again with most predictive power for GPRS including all SNPs, with a β of 0.11 ($P = 2.7e-37$) and an R^2 of 1.2% (Table 2a). However, we found no significant *across-trait* prediction: the MDD-GPRS did not predict EA ($\beta = -0.01$ $P = 6.7e-2$) and the EA-GPRS did not predict MDD (OR = 0.99 $P = 5.9e-1$; Table 2a). Secondary analyses, performed within all cohorts separately, indicated that the within-trait predictive effects were consistent in all cohorts, and that the lack of across-trait predictive power was also consistent for all cohorts (Table 2b). In addition, no correlation was found between the MDD-GPRS and the EA-GPRS themselves ($P_T = 1$; correlation coefficient of 0.006, $P = 0.413$). In additional analyses, across-trait predictive effects on MDD were tested for GPRS based on the SSGAC EA outcome tagging College completion (College-GPRS).[7] College completion distinguishes more in the extreme end of the EA distribution, and has a confirmed genetic basis.[7] However, no predictive effects of the College-GPRS on MDD were found (OR = 0.99, $P = 0.74$ for $P_T = 1$; Supplementary Table 4).

GREML analyses

GREML analyses in the overall study sample generated an estimate of MDD SNP-h^2 of 0.173 (s.e. = 0.017, $P < 1e-16$) on the liability scale ($K = 0.2$); this finding was not solely driven by one of the individual cohorts, because the MDD SNP-h^2 was estimated at consistent values when one cohort was left out at the time (Table 3). The MDD SNP-h^2 was larger when expressed on the liability scale (0.173) compared with that on the observed scale (0.126), with a larger SNP-h^2 for larger values of disease frequency

Table 3. Bivariate GREML analyses of MDD and EA

Cohorts included	*MDD SNP-h²*					*EA SNP-h²*				*Genetic correlation*		
	Case	*Control*	*Est*	*S.e.*	P-*value*	N	*Est*	*S.e.*	P-*value*	*Est*	*S.e.*	P-*value*
(A) EA in both cases and controls												
All	9662	14 949	0.173	0.017	< 1e–16	15 985	0.124	0.019	2.8e–11	− 0.253	0.087	0.004
All excluding B	8737	13 667	0.155	0.019	1.1e–16	15 985	0.124	0.019	3.1e–11	− 0.249	0.096	0.009
All excluding E	9154	9604	0.207	0.022	< 1e–16	10 970	0.133	0.026	4.5e–07	− 0.265	0.102	0.009
All excluding GE	8686	13 734	0.183	0.019	< 1e–16	14 815	0.130	0.020	9.0e–11	− 0.345	0.091	1.4e–04
All excluding GS	8796	14 086	0.172	0.019	< 1e–16	14 257	0.120	0.021	7.6e–09	− 0.231	0.099	0.020
All excluding M	9325	14 416	0.177	0.018	< 1e–16	15 985	0.124	0.019	3.9e–11	− 0.208	0.088	0.018
All excluding N1	8102	13 826	0.150	0.020	2.1e–14	13 728	0.120	0.022	3.0e–08	− 0.239	0.109	0.028
All excluding N2	9426	13 748	0.174	0.018	< 1e–16	14168	0.118	0.021	1.4e–08	− 0.294	0.097	0.002
All excluding Q	8230	13 263	0.199	0.020	< 1e–16	13 325	0.133	0.022	1.6e–09	− 0.237	0.091	0.009
All excluding R	8057	13 376	0.161	0.020	1.0e–15	15 985	0.124	0.019	3.6e–11	− 0.206	0.097	0.033
All excluding S	8445	14 821	0.187	0.019	< 1e–16	14 647	0.126	0.020	6.3e–10	− 0.294	0.092	0.001
(B) EA in controls only												
All	9662	14 949	0.173	0.017	< 1e–16	9765	0.144	0.030	1.5e–06	− 0.110	0.105	0.298
All excluding B	8737	13 667	0.156	0.019	1.1e–16	9765	0.144	0.030	1.6e–06	− 0.113	0.115	0.327
All excluding E	9154	9604	0.208	0.022	< 1e–16	5196	0.177	0.054	0.001	0.004	0.130	0.972
All excluding GE	8686	13 734	0.183	0.019	< 1e–16	8595	0.159	0.034	2.6e–06	− 0.187	0.110	0.087
All excluding GS	8796	14 086	0.172	0.019	< 1e–16	8903	0.133	0.033	4.4e–05	− 0.120	0.119	0.310
All excluding M	9325	14 416	0.178	0.018	< 1e–16	9765	0.144	0.030	1.4e–06	− 0.065	0.105	0.537
All excluding N1	8102	13 826	0.151	0.020	1.7e–14	8890	0.144	0.033	9.4e–06	− 0.112	0.124	0.369
All excluding N2	9426	13 748	0.174	0.018	< 1e–16	9006	0.132	0.032	4.0e–05	− 0.176	0.117	0.133
All excluding Q	8230	13 263	0.200	0.020	< 1e–16	8363	0.143	0.035	3.6e–05	− 0.148	0.114	0.196
All excluding R	8057	13 376	0.161	0.020	1.0e–15	9765	0.144	0.030	1.4e–06	− 0.059	0.116	0.610
All excluding S	8445	14 821	0.188	0.019	< 1e–16	9637	0.140	0.030	3.3e–06	−0.084	0.107	0.434

Abbreviations: B, Bonn/Mannheim; E, EGCUT; EA, educational attainment; EGCUT, Estonian Genome Center of the University of Tartu; GE, GenRED; GEnRED, GAIN-MDD cohort and the Genetics of Recurrent Early-Onset Depression; GPRS, genetic profile risk scores; GREML, genomic-relationship-matrix restricted maximum likelihood; GS, GSK; GSK, GlaxoSmithKline; M, MPIP; MDD, major depressive disorder; MPIP, Max Planck Institute of Psychiatry; N1, NESDA/NTR-1; N2, NESDA/NTR-2; NESDA, Netherlands Study of Depression and Anxiety; NTR, Netherlands Twin Registry; OR, odds ratio; Q, QIMR; QIMR, Queensland Institute of Medical Research; R, RADIANT MDD study; S, STAR*D; SNP, single-nucleotide polymorphism. Bivariate GREML was performed to estimate the proportion of variation explained in MDD (MDD SNP-h^2) and EA (EA SNP-h^2) by genome-wide common SNPs, as well as the genetic correlation between MDD and EA. First, EA information of both cases and controls was analyzed (first line upper panel), but the correlation thus found should be interpreted with caution, because it could be biased (see Materials and methods). Therefore, analyses were repeated taking EA information of controls only into account (first line lower panel). To estimate the robustness of these findings, the analyses were repeated leaving one cohort at the time (additional lines in upper and lower panel). The MDD SNP-h^2 was expressed on the liability scale assuming a population prevalence K of 20% (see Supplementary Table 5 for comparison of the MDD SNP-h^2 for different values of K with the SNP-h^2 on the observed scale). MDD was available for all 10 cohorts; EA was only available for cohorts E, GE, GS, N1, N2, Q and S. The analyses were corrected for sex, the first 20 principal components and a categorical covariate labeling the cohorts and genotype batches. [a]Explained in MDD (MDD SNP-h^2) and EA (EA SNP-h^2) by genome-wide common SNPs, as well as the genetic correlation between MDD and EA.

(as expected from Equation (23) from Lee *et al*.;[33] Supplementary Table 5). The EA SNP-h^2 was estimated at 0.124 (s.e. = 0.019, $P = 2.8e-11$) when EA information in both cases and controls was taken into account (Table 3a), and at 0.144 (s.e. = 0.030, $P = 1.5e-6$) when EA information of controls only was used (Table 3b). Again, these estimates were not solely driven by one of the individual cohorts (Table 3). The genetic correlation between MDD and EA was estimated at −0.253 (s.e. = 0.087, $P = 0.004$) when EA information of both cases and controls was taken into account (Table 3a). As a correlation between genetic and environmental factors is likely to be partitioned into the genetic variance and covariance components, we explored the robustness of this estimate by limiting EA to be measured only in controls. When taking into account EA of controls only and MDD status from cases and controls, the genetic correlation dropped considerably and was no longer significantly different from 0 with an estimate of −0.110 (s.e. = 0.105, $P = 0.298$; Table 3b). In *post hoc* analyses, we tested if EA moderated the polygenic effects on MDD, but found no such evidence with neither GPRS nor GREML analyses (Supplementary Materials).

SNP effect concordance analysis

SECA showed no evidence for genetic correlation. The primary SECA test divided the SNPs in 144 subsets based on significance of association with MDD and EA smaller than, respectively, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0. Not a single of these subsets contained a larger number of SNPs than expected by chance, that is, no concordance was found with respect to the MDD and EA SNP effect *significances*. When comparing the directions of SNP effects, only 4 of the 144 subsets showed nominally correlated directions of effect, which is not more than expected by chance (permuted empirical *P*-value 0.244), indicating no concordance with respect to the MDD and EA SNP effect *directions*.

DISCUSSION

This study tested the existence of pleiotropic genetic effects (genetic correlation) between MDD and lower EA on individual genotype data from a large sample of ~25 000 subjects from western countries. To start, a strong negative phenotypic association was found with an OR for MDD of 0.78 per s.d. increase in EA, which is in line with findings from a meta-analysis of 37 studies from mainly western countries by Lorant *et al.*[1] Our first test for genetic correlation was negative with no across-trait predictive power of the GPRS: EA-GPRS did not predict MDD and MDD-GPRS did not predict EA. In the second test for genetic correlation, GREML analyses did not show consistent evidence for genetic correlation. The third test, SECA, also showed no evidence

for concordance of EA and MDD SNP effects with respect to their significance or direction.

The GPRS in our study had within-trait predictive power in line with previous findings,[7,13] and were based on an independent EA discovery sample from the SSGAC[7] of ~120 000 subjects and independent MDD leave-one-cohort-out discovery samples of ~8000 cases and 12000 controls. These numbers seem adequate, but the discovery sets would ideally have been even larger, because most predictive power was still found for the GPRS including all SNPs ($P_T = 1$), indicating that true effect SNPs were associated in the discovery sample with *P*-values close to 1.[28] Nevertheless, Dudbridge power calculations suggested that the EA-GPRS were well powered to predict MDD when the genetic correlation would have been around − 0.2 (Supplementary Figure 3).[35] Our GPRS results, therefore, indicate that a large genetic correlation between EA and MDD is unlikely, but could not exclude a small genetic correlation of around − 0.1.

We performed GREML analyses to estimate the MDD SNP-h^2, EA SNP-h^2 and genetic correlation. The MDD SNP-h^2 found (0.17) was considerably smaller than the one previously found by Lubke *et al.* (0.32),[10] which could well be because of the actual differences in SNP-h^2 across cohorts; the sample of Lubke was included in the current study as NESDA/NTR-1 and indeed had the largest contribution to the overall SNP-h^2 of all cohorts (Table 3). The EA SNP-h^2 (0.14 in controls only) was of the same magnitude (< 2 s.e. difference) as the SNP-h^2 found by Rietveld *et al.* (0.2).[7] The GREML estimate of the genetic correlation was somewhat complicated to interpret. A significant negative genetic correlation was found (−0.25, $P = 0.004$) when EA information of both cases and controls was taken into account, but we fear this finding could be biased particularly in the context of genotype and environment correlation. In fact, when taking EA information of only controls into account, the estimate of genetic correlation dropped considerably and was no longer significant (−0.11, $P = 0.30$). However, we note that this estimate was conservative as it reduced variation in EA, and we note the negative point estimate and high standard error, showing that this analyses was underpowered to draw definitive inference. Taken together, the GREML analyses could be in line with a small genetic correlation of around − 0.1. In addition to the two methods based on individual-level genotype data, we also performed analyses on GWAS summary statistics with the recently published SECA method[34] and found no evidence for genetic correlation.

To the best of our knowledge, only three previous studies tested for a genetic correlation between MDD and EA. López-León *et al.*[11] used a family-based approach in 2383 subjects to find a negative genetic correlation of − 0.65 and − 0.50 between EA and self-reports of depressive symptoms based on, respectively, the Center for Epidemiologic Studies Depression Scale (CES-D) and the Hospital Anxiety and Depression Scale (HADS-D). Boardman *et al.*[12] also used cross-sectional CES-D assessments and found a genetic correlation of − 0.7 with GREML analyses. Mezuk *et al.*[5] used a twin design with depression assessed with the DSM-IV-based Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders (SCID-I), and their study generally supported social causation as cause for the link between lower EA and increased MDD risk, and found only a small genetic correlation of − 0.22.[5] The studies of López-León *et al.*[11] and Boardman *et al.*[12] contrast our finding of no, or at most a small, genetic correlation, but this could be because they tested symptom reports of depressive state at a specific point in time, whereas our study tested the presence of a more clinical construct: DSM-IV- or ICD-10-based lifetime diagnosis of MDD. Indeed, our results appear in line with the findings from Mezuk *et al.*,[5] who also used DSM-IV-based diagnoses of MDD. Furthermore, we found that the association between lower EA and MDD remained when cases with an age of first MDD onset > 30 were taken into account exclusively. This indicates that it is unlikely that MDD directly causes a lowering of EA, as it can be assumed that one reaches his or her education potential before the age of 30 years, which is in line with the suggested social causation by Mezuk *et al.*[5]

The finding that there is no, or at most a small, genetic correlation between lower EA and MDD is relevant, because this implies that non-genetic factors have an important role, and that underlying mechanisms may possibly be accessible to interventions. For example, when the social causation model would be studied in more detail, this could potentially lead to underlying clues on how lower socioeconomic status could contribute to vulnerability for MDD, or alternatively how higher socioeconomic status may buffer against vulnerability for MDD. For instance, lower socioeconomic status has shown to be associated with less healthy life styles (less physical exercise, more smoking, higher body mass index and more use of alcohol),[36,37] less adequate medical treatment seeking behavior,[38] less knowledge about MDD[39] and higher vulnerability to experience stressful life events.[40] These factors could all contribute to increased MDD risk. However, future research should be conducted to elucidate the most important underlying mechanism as these may hint to either public or personal actions to best prevent MDD among individuals with lower EA. Yet another mechanism underlying the link between lower EA and MDD could possibly be found in a third factor other than genetic effects, such as a certain personality characteristic or less developed cognitive abilities, that causes both lower EA and increased MDD risk.

Our study has several strengths, but also some limitations. First, our study is one of the first and largest studies to test for pleiotropy between lower EA and MDD, and we used individual-level genotype data. In addition, we used clinically relevant DSM-IV- and ICD-10-based diagnoses of MDD. Furthermore, we applied three distinct methods that essentially lead to the same conclusion. A limitation of our study is that the discovery samples of the polygenic risk score analyses were not optimally sized with maximum predictive power of the GPRS including all SNPs ($P_T = 1$). However, this is a limitation of most current genetic studies, and we feel our discovery samples were adequately powered given the availability of relevant genetic cohorts up to date. Furthermore, the genetic basis of MDD is strong enough to study pleiotropy, as has been indicated in a previous work from the PGC that indicate a genetic correlation between MDD schizophrenia (0.43 ± 0.06) and MDD bipolar disorder (0.47 ± 0.06) with both GREML[41] and GPRS analyses.[42] Another limitation is that we could have missed pleiotropic effects among rare SNPs with an MAF < 0.01. This limitation could be addressed with a family or twin study, but it would be surprising when SNPs with MAF < 0.01 would have large pleiotropic effects, while SNPs with MAF > 0.01 show no such evidence.

To conclude, we did confirm a negative phenotypic association between MDD and EA, but found no evidence that this association is due to genetic factors, which indicates that a large genetic correlation between lower EA and MDD is unlikely, but does not exclude a small genetic correlation of around − 0.1. Understanding of the possible pathways between lower EA and MDD risk requires further research including twin analyses for an additional estimate of the upper bound of the genetic correlation. Nevertheless, we believe that the finding of the absence of large pleiotropic genetic effects underlying the established correlation of lower EA with increased MDD risk may be relevant, as it points to non-genetic mechanisms that may be accessible to interventions aimed at breaking this deleterious link.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

NRW and SHL supported by the Australian Research Council (FT0991360 and DE130100614) and the National Health and Medical Research Council (613608, 1011506 and 1047956). The PGC is supported by National Institute of Mental Health (NIMH) Grant U01 MH085520. Statistical analyses were carried out on the Genetic Cluster Computer (see Web Resources), which is financially supported by the Netherlands Scientific Organization. The Bonn/Mannheim (BoMa) GWAS was supported by the German Federal Ministry of Education and Research, within the context of the National Genome Research Network 2 (NGFN2), the National Genome Research Network plus (NGFNplus) and the Integrated Genome Research Network (IG) MooDS (Grant 01GS08144 to S Cichon and MMN, and Grant 01GS08147 to MR). The work at deCODE was funded by European Union Grants LSHM-CT-2006-037761 (Project SGENE), PIAP-GA-2008-218251 (Project PsychGene) and HEALTH-F2-2009-223423 (Project PsychCNVs). GenPod was funded by the Medical Research Council (UK) and supported by the Mental Health Research Network. Genotyping of the GenPod sample was funded by the Innovative Medicines Initiative Joint Undertaking under Grant Agreement number 115008 (NEWMEDS). The GenRED GWAS project was supported by NIMH R01 Grants MH061686 (to DFL), MH059542 (to WHC), MH075131 (to WBL), MH059552 (to JBP), MH059541 (to WAS) and MH060912 (to MMW). We acknowledge the contributions of Dr George S Zubenko and Dr Wendy N Zubenko, Department of Psychiatry, University of Pittsburgh School of Medicine, to the GenRED I project. The NIMH Cell Repository at Rutgers University and the NIMH Center for Collaborative Genetic Studies on Mental Disorders made essential contributions to this project. Genotyping was carried out by the Broad Institute Center for Genotyping and Analysis with support from Grant U54 RR020278 (which partially subsidized the genotyping of the GenRED cases). Collection and quality control analyses of the control data set were supported by grants from NIMH and the National Alliance for Research on Schizophrenia and Depression. We are grateful to Knowledge Networks (Menlo Park, CA, USA) for assistance in collecting the control data set. We express our profound appreciation to the families who participated in this project, and to the many clinicians who facilitated the referral of participants to the study. The Depression Genes and Networks ARRA grant was funded by RC2MH089916. Funding for the Harvard i2b2 sample was provided by a subcontract to RH Perlis and JW Smoller as part of the i2b2 Center (Informatics for Integrating Biology and the Bedside), an NIH-funded National Center for Biomedical Computing based at Partners HealthCare System (U54LM008748, PI: IS Kohane), and by an NIMH Grant (to RH Perlis; MH086026). Max Planck Institute of Psychiatry MARS study was supported by the BMBF Program Molecular Diagnostics: Validation of Biomarkers for Diagnosis and Outcome in Major Depression (01ES0811). Genotyping was supported by the Bavarian Ministry of Commerce, and the Federal Ministry of Education and Research (BMBF) in the framework of the National Genome Research Network (NGFN2 and NGFN-Plus, FKZ 01GS0481 and 01GS08145). The Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR) contributed to GAIN-MDD and to MDD2000. This study was funded by: the Netherlands Organization for Scientific Research (MagW/ZonMW Grants 904-61-090, 985-10-002, 904-61-193, 480-04-004, 400-05-717, 912-100-20; Spinozapremie 56-464-14192; Geestkracht program Grant 10-000-1002); the Center for Medical Systems Biology (NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure, VU University's Institutes for Health and Care Research and Neuroscience Campus Amsterdam, NBIC/BioAssist/RK (2008.024); the European Science Foundation (EU/QLRT-2001-01254); the European Community's Seventh Framework Program (FP7/2007-2013); ENGAGE (HEALTH-F4-2007-201413); and the European Science Council (ERC, 230374). Genotyping was funded, in part, by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). The PsyCoLaus study was supported by grants from the Swiss National Science Foundation (nos. 3200B0-105993, 3200B0-118′308, 33CSC0-122661) and from GlaxoSmithKline (Psychiatry Center of Excellence for Drug Discovery and Genetics Division, Drug Discovery—Verona, R&D). We express our gratitude to the Lausanne inhabitants who volunteered to participate in the PsyCoLaus study. We also thank V Mooser, G Weaber and P Vollenweider who initiated the CoLaus project. Funding for the QIMR samples was provided by the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496675, 496739, 552485, 552498, 613602, 613608, 613674, 619667), the Australian Research Council (FT0991360, FT0991022), the FP-5 GenomEUtwin Project (QLG2-CT- 2002-01254) and the US National Institutes of Health (AA07535, AA10248, AA13320, AA13321, AA13326, AA14041, MH66206, DA12854, DA019951), and the Center for Inherited Disease Research (Baltimore, MD, USA). We thank the twins and their families registered at the Australian Twin Registry for their participation in the many studies that have contributed to this research. RADIANT was funded by: a joint grant from the UK Medical Research Council and GlaxoSmithKline (G0701420); the National Institute for Health Research (NIHR) Specialist Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and the Institute of Psychiatry, King's College London; and the UK Medical Research Council (G0000647). The GENDEP study was funded by a European Commission Framework 6 grant, EC Contract Ref.: LSHB-CT-2003-503428. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (Grant nos. 01ZZ9603, 01ZZ0103 and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (Grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG. SHIP-LEGEND is funded by the German Research Foundation (DFG: GR 1912/5-1). Genotyping of STAR*D was supported by an NIMH Grant to SP Hamilton (MH072802). STAR*D was funded by the National Institute of Mental Health (Contract No. N01MH90003) to the University of Texas Southwestern Medical Center at Dallas (AJ Rush, principal investigator). The TwinGene study was supported by the Swedish Ministry for Higher Education, the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), the Swedish Foundation for Strategic Research and the US National Institutes of Health (U01 DK066134). 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 data is available from http://www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under awards 076113 and 085475. EGCUT received financing from FP7 Grant 313010 and from European Regional Development Fund, road map Grant No. 3.2.0304.11-0312 and grant 'Center of Excellence in Genomics' (EXCEGEN). EGCUT studies were covered also by targeted financing from Estonian Government (IUT24-6, IUT20-60) and CTG grant (SP1GVAR-ENG) from Development Fund of the University of Tartu. We acknowledge EGCUT technical personnel, especially Mr V Soo and S Smit and Mari-Liis Tammesoo. EMB is supported by the National Health and Medical Research Early Career Fellowship Scheme.

MEMBERS OF MAJOR DEPRESSIVE DISORDER WORKING GROUP OF THE PSYCHIATRIC GWAS CONSORTIUM

The following people are not listed as coauthors on this manuscript, but did contribute to the Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium,[13] from which individual-level genotype and phenotype data was used. We are grateful for their contribution to the Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. The views presented in the present paper may not reflect the opinions of the individuals listed below. We thank: CM Lewis, SP Hamilton, MM Weissman, G Breen, DH Blackwood, S Cichon, AC Heath, F Holsboer, PA Madden, P McGuffin, P Muglia, ML Pergadia, D Lin, B Müller-Myhsok, S Steinberg, HJ Grabe, P Lichtenstein, P Magnusson, RH Perlis, M Preisig, JW Smoller, K Stefansson, R Uher, Z Kutalik, KE Tansey, A Teumer, A Viktorin, MR Barnes, T Bettecken, EB Binder, R Breuer, VM Castro, SE Churchill, WH Coryell, N Craddock, IW Craig, D Czamara, F Degenhardt, AE Farmer, M Fava, J Frank, VS Gainer, PJ Gallagher, SD Gordon, S Goryachev, M Gross, M Guipponi, AK Henders, S Herms, IB Hickie, S Hoefels, W Hoogendijk, DV Iosifescu, M Ising, I Jones, L Jones, T Jung-Ying, JA Knowles, IS Kohane, MA Kohli, A Korszun, M Landen, WB Lawson, G Lewis, D Macintyre, W Maier, M Mattheisen, PJ McGrath, A McIntosh, A McLean, CM Middeldorp, L Middleton, GM Montgomery, SN Murphy, M Nauck, WA Nolen, DR Nyholt, M O'Donovan, H Oskarsson, N Pedersen, WA Scheftner, A Schulz, TG Schulze, SI Shyn, E Sigurdsson, SL Slager, JH Smit, H Stefansson, M Steffens, T Thorgeirsson, F Tozzi, J Treutlein, M Uhr, EJ van den Oord, G Van Grootheest, H Völzke, JB Weilburg, G Willemsen, FG Zitman, B Neale, M Daly and PF Sullivan.

MEMBERS OF SOCIAL SCIENCE GENETIC ASSOCIATION CONSORTIUM

The following people who are not listed as coauthors on this manuscript contributed to the original GWAS meta-analysis on educational attainment,[7] on which the present paper is based. Data access has been granted under section 4 of the Data Sharing Agreement of the Social Science Genetic Association Consortium (SSGAC). The views presented in the present paper may not reflect the opinions of the individuals listed below. We are grateful

to the authors of Rietveld *et al*.[7] for providing the meta-analysis data. We thank: Arpana Agrawal, Eva Albrecht, Behrooz Z Alizadeh, Jüri Allik, Najaf Amin, John R Attia, Stefania Bandinelli, John Barnard, François Bastardot, Sebastian E Baumeister, Jonathan Beauchamp, Daniel J Benjamin, Kelly S Benke, David A Bennett, Klaus Berger, Lawrence F Bielak, Laura J Bierut, Jeffrey A Boatman, Patricia A Boyle, Ute Bültmann, Harry Campbell, David Cesarini, Christopher F Chabris, Lynn Cherkas, Mina K Chung, Dalton Conley, Francesco Cucca, George Davey-Smith, Gail Davies, Mariza de Andrade, Philip L. De Jager, Christiaan de Leeuw, Jan-Emmanuel De Neve, Ian J. Deary, George V Dedoussis, Panos Deloukas, Jaime Derringer, Maria Dimitriou, Gudny Eiriksdottir, Niina Eklund, Martin F Elderson, Johan G Eriksson, Daniel S Evans, David M Evans, Jessica D Faul, Rudolf Fehrmann, Luigi Ferrucci, Krista Fischer, Lude Franke, Melissa E Garcia, Christian Gieger, Håkon K Gjessing, Patrick JF Groenen, Henrik Grönberg, Vilmundur Gudnason, Sara Hägg, Per Hall, Jennifer R Harris, Juliette M Harris, Tamara B Harris, Nicholas D Hastie, Caroline Hayward, Andrew C Heath, Dena G Hernandez, Wolgang Hoffmann, Adriaan Hofman, Albert Hofman, Rolf Holle, Elizabeth G Holliday, Christina Holzapfel, William G Iacono, Carla A Ibrahim-Verbaas, Thomas Illig, Erik Ingelsson, Bo Jacobsson, Marjo-Riitta Järvelin, Min A Jhun, Magnus Johannesson, Peter K Joshi, Astanand Jugessur, Marika Kaakinen, Mika Kähönen, Stavroula Kanoni, Jaakkko Kaprio, Sharon LR Kardia, Juha Karjalainen, Robert M Kirkpatrick, Philipp D Koellinger, Ivana Kolcic, Matthew Kowgier, Kati Kristiansson, Robert F Krueger, Zóltan Kutalik, Jari Lahti, David Laibson, Antti Latvala, Lenore J Launer, Debbie A Lawlor, Terho Lethimäki, Jingmei Li, Paul Lichtenstein, Peter K Lichtner, David C Liewald, Peng Lin, Penelope A Lind, Yongmei Liu, Kurt Lohman, Marisa Loitfelder, Pamela A Madden, Patrick KE Magnusson, Tomi E Mäkinen, Pedro Marques Vidal, Nicolas W Martin, Marco Masala, Matt McGue, George McMahon, Osorio Meirelles, Michelle N Meyer, Andreas Mielck, Lili Milani, Michael B Miller, Grant W Montgomery, Sutapa Mukherjee, Ronny Myhre, Marja-Liisa Nuotio, Dale R Nyholt, Christopher J Oldmeadow, Ben A Oostra, Lyle J Palmer, Aarno Palotie, Markus Perola, Katja E Petrovic, Patricia A Peyser, Ozren Polašek, Danielle Posthuma, Martin Preisig, Lydia Quaye, Katri Räikkönen, Olli T Raitakari, Anu Realo, Eva Reinmaa, John P Rice, Susan M Ring, Samuli Ripatti, Fernando Rivadeneira, Thais S Rizzi, Igor Rudan, Aldo Rustichini, Veikko Salomaa, Antti-Pekka Sarin, David Schlessinger, Helena Schmidt, Reinhold Schmidt, Rodney J Scott, Konstantin Shakhbazov, Albert V Smith, Jennifer A Smith, Harold Snieder, Beate St Pourcain, John M Starr, Jae Hoon Sul, Ida Surakka, Rauli Svento, Toshiko Tanaka, Antonio Terracciano, Alexander Teumer, A Roy Thurik, Henning Tiemeier, Nicholas J Timpson, André G Uitterlinden, Matthijs JHM van der Loos, Cornelia M van Duijn, Frank JA van Rooij, David R Van Wagoner, Erkki Vartiainen, Jorma Viikari, Peter M Visscher, Veronique Vitart, Peter K Vollenweider, Henry Völzke, Judith M Vonk, Gérard Waeber, David R Weir, Jürgen Wellmann, Harm-Jan Westra, H-Erich Wichmann, Elisabeth Widen, James F Wilson, Alan F Wright, Jian Yang, Lei Yu, Wei Zhao.

REFERENCES

1 Lorant V, Deliège D, Eaton W, Robert A, Ansseau M. Socioeconomic inequalities in depression: a meta-analysis. *Am J Epidemiol* 2003; **157**: 98–112.
2 Bromet E, Andrade LH, Hwang I, Sampson Na, Alonso J, de Girolamo G, *et al.* Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med* 2011; **9**: 90–106.
3 Andrade L, Caraveo-Anduaga JJ, Berglund P, Bijl RV, De Graaf R, Vollebergh W, *et al.* The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *Int J Methods Psychiatr Res* 2003; **12**: 3–21.
4 de Graaf R, ten Have M, van Gool C, van Dorsselaer S. Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc Psychiatry Psychiatr Epidemiol* 2012; **47**: 203–213.
5 Mezuk B, Myers JM, Kendler KS. Integrating social science and behavioral genetics: testing the origin of socioeconomic disparities in depression using a genetically informed design. *Am J Public Health* 2013; **103**: S145–S151.
6 Turner RJ, Lloyd DA. The stress process and the social distribution of depression. *J Health Soc Behav* 1999; **40**: 374–404.
7 Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 2013; **340**: 1467–1471.
8 Dongen van J, Slagboom PE, Draisma HHM, Martin NG, Boomsma DI. The continuing value of twin studies in the omics era. *Nat Rev Genet* 2012; **13**: 640–653.
9 Demirkan A, BWJH Penninx, Hek K, Wray NR, Amin N, Aulchenko YS, *et al.* Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry* 2011; **16**: 773–783.
10 Lubke GH, Hottenga JJ, Walters R, Laurin C, De Geus EJC, Willemsen G, *et al.* Estimating the Genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biol Psychiatry* 2012; **72**: 707–709.
11 López-León S, Choy W, Aulchenko Y. Genetic factors influence the clustering of depression among individuals with lower socioeconomic status. *PLoS One* 2009; **4**: 1–5.
12 Boardman JD, Domingue BW, Daw J. What can genes tell us about the relationship between education and health? *Soc Sci Med* 2014; **127**: 171–180.
13 Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 2013; **18**: 497–511.
14 Rietschel M, Mattheisen M, Frank J, Treutlein J, Degenhardt F, Breuer R, *et al.* Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. *Biol Psychiatry* 2010; **68**: 578–585.
15 Sullivan PF, De Geus EJC, 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**: 359–375.
16 Shi J, Potash JB, Knowles JA, Weissman MM, Coryell W, Scheftner WA, *et al.* Genome-wide association study of recurrent early-onset major depressive disorder. *Mol Psychiatry* 2011; **16**: 193–201.
17 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* 2010; **15**: 589–601.
18 Ising M, Lucae S, Binder EB, Bettecken T, Uhr M, Ripke S, *et al.* A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch Gen Psychiatry* 2009; **66**: 966–975.
19 Wray NR, Pergadia ML, Blackwood DHR, Penninx BWJH, Gordon SD, Nyholt DR, *et al.* Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry* 2012; **17**: 36–48.
20 Lewis CM, Ng MY, Butler AW, Cohen-Woods S, Uher R, Pirlo K, *et al.* Genome-wide association study of major recurrent depression in the U.K. population. *Am J Psychiatry* 2010; **167**: 949–957.
21 Shyn SI, Shi J, Kraft JB, Potash JB, Knowles Ja, Weissman MM, *et al.* Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Mol Psychiatry* 2011; **16**: 202–215.
22 Penninx BWJH Beekman ATF, 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**: 121–140.
23 Boomsma DI, de Geus EJC, Vink JM, Stubbe JH, Distel Ma, Hottenga J-J, *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* 2006; **9**: 849–857.
24 Leitsalu L, Haller T, Esko T, Tammesoo M-L, Alavere H, Snieder H, *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol* 2014 E-pub ahead of print: 1–11.
25 Boomsma D, Beem A, van den Berg M, Dolan C, Koopmans J, Vink J, *et al.* Netherlands twin family study of anxious depression (NETSAD). *Twin Res* 2012; **3**: 323–334.
26 Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**: 52–58.
27 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.
28 Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
29 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
30 Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* 2012; **36**: 214–224.

31 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, *et al*. Common SNPs explain a large proportion of heritability for human height. *Nat Genet* 2010; **42**: 565–569.
32 Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 2012; **28**: 2540–2542.
33 Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 2011; **88**: 294–305.
34 Nyholt DR. SECA: SNP effect concordance analysis using genome-wide association summary results. *Bioinformatics* 2014; **30**: 2086–2088.
35 Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013; **9**: e1003348.
36 Pampel FC, Krueger PM, Denney JT. Socioeconomic disparities in health behaviors. *Annu Rev Sociol* 2010; **36**: 349–370.
37 Cutler DM, Lleras-Muney A. Understanding differences in health behaviors by education. *J Health Econ* 2010; **29**: 1–28.
38 Fletcher JM, Frisvold DE. Higher education and health investments: does more schooling affect preventive health care use? *J Hum Cap* 2009; **3**: 144–176.
39 Johnston DW, Lordan G, Shields MA, Suziedelyte A. Education and health knowledge: evidence from UK compulsory schooling reform. *Soc Sci Med* 2015; **127**: 92–100.
40 Lantz PM, House JS, Mero RP, Williams DR. Stress, life events, and socioeconomic disparities in health: results from the Americans' Changing Lives Study. *J Health Soc Behav* 2005; **46**: 274–288.
41 Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, *et al*. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**: 984–994.
42 Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**: 1371–1379.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)