RUNNING TITLE: EA polygenic score, SES, and depression.

Educational Attainment Polygenic Score is Associated with Depressive Symptoms via Socioeconomic Status: A Gene-Environment-Trait Correlation

Reut Avinun

Department of Psychology & Neuroscience, Duke University, Durham, NC, USA and the Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel

Corresponding Author: Reut Avinun, Ph.D., Laboratory of NeuroGenetics, Department of Psychology & Neuroscience, Duke University, Grey Building 2020 West Main St, Ste 0030 Durham, NC 27705. Email: reut.avinun@mail.huji.ac.il

1 **Funding and disclosure**

I would like to thank the participants of the Duke Neurogenetics Study and the members of the Laboratory of NeuroGenetics, especially Annchen R. Knodt, Spenser R. Radtke, and Bartholomew D. Brigidi for their assistance with data collection and analysis. I would also like to thank the head of the laboratory, Prof. Ahmad Hariri, without whom this study would not have been possible. Lastly, I would like to thank Dr. Aysu Okbay for her help with obtaining a GWAS of educational attainment that did not include UK biobank data. The DNS was supported by Duke University as well as US-National Institutes of Health grants R01DA033369 and R01DA031579. The author received support from US-National Institutes of Health grant R01AG049789 and from a fellowship from the Jerusalem Brain Community and declares no competing financial or other interests.

Abstract

Previous research has found evidence to support low educational attainment (EA) as a risk factor for depression and shown that EA and depression are genetically correlated. However, the nature of the genetic link between EA and depression remains unknown. Recently, the environment has been suggested as a mediator of genetic influences in a process termed Gene-Environment-Trait correlations (rGET). As socioeconomic status (SES) is closely related to EA and has been associated with depression, an rGET in which SES mediates an association between the genetic influences on EA and depression is possible. Summary statistics from a recent genome-wide association study of EA were used to calculate EA polygenic scores and test whether they predict depressive symptoms through SES. Two independent samples were used for the analyses: 522 non-Hispanic Caucasian university students from the Duke Neurogenetics Study (277 women, mean age 19.78±1.24 years) and 5,243 white British volunteers (2,669 women, mean age 62.30±7.41 years) from the UK biobank (UKB). Results indicated a significant mediation in the DNS (indirect effect=-.12, bootstrapped SE=.06, bootstrapped 95% CI: -.26 to -.02), wherein higher EA polygenic scores predicted higher SES, which in turn predicted lower depressive symptoms. This mediation was replicated in the UKB (indirect effect=-.07, bootstrapped SE=.01, bootstrapped 95% CI: -.091 to -.051). These findings suggest that the genetic correlates of depression may be environment-dependent and that public policy that aims to reduce socioeconomic inequalities and the adverse correlates of low SES may relieve the individual and societal burden of depression.

Keywords: Depression; Socioeconomic status (SES); Educational attainment (EA); Gene-environment-trait correlation (rGET); Gene-environment-correlation (rGE).

Depression is a major cause of disability. With a global prevalence of around 4.7% (Ferrari et al., 2013), it is predicted to become one of the three leading causes of illness by 2030 (Mathers and Loncar, 2006). Educational attainment (EA), which is often viewed as a proxy for cognitive ability and intelligence, has been linked to depression, so that the probability of experiencing depression decreases for additional years of education (Crespo et al., 2014). A recent study (Wray et al., 2018) further showed that the link between EA and depression is partly due to shared genetic influences, and by employing a genetically informed analysis (bidirectional Mendelian randomization) found evidence to support low EA as a risk factor for depression. Notably, how the genetic link between EA and depression is mediated has not been established.

Recently, I hypothesized that the environment may mediate genetic correlations between two phenotypes in a process termed gene-environment-trait correlation (Avinun, in press). This hypothesis stems from accumulating research showing passive, active, and evocative processes that lead to correlations between genetic variations and environmental measures, such as parenting and stressful life events (Avinun and Knafo, 2014; Kendler and Baker, 2007). These passive, active, and evocative processes, known as gene-environment correlations (Plomin et al., 1977; Scarr and McCartney, 1983), occur due to genetically influenced characteristics that shape the individual's environment. As the environment can in turn substantially affect various outcomes, it may act as a mediator of genetic effects and contribute to the widespread genetic correlations observed between numerous phenotypes (Bulik-Sullivan et al., 2015), including EA and depression (Wray et al., 2018).

Identifying gene-environment-depression correlations can provide modifiable targets for public policy and also demonstrate the importance of context in the discovery of the genetic variants that influence depression.

Socioeconomic status (SES), which can be defined as an individual's or group's position within a social hierarchy that is determined by factors such as education, occupation, income, and wealth (Calixto and Anaya, 2014), has been shown to be genetically influenced (Hill et al., 2016; Marioni et al., 2014). In other words, genetically influenced traits affect an individual's SES. One of these traits, as has been found in a meta-analysis of longitudinal studies, is intelligence (Strenze, 2007), which is highly heritable (Plomin and Deary, 2015) and highly genetically correlated with EA (a single nucleotide polymorphism-based genetic correlation of .95; Marioni et al., 2014). Because SES has been associated with various physiological and mental disorders (e.g., Calixto and Anaya, 2014; Galobardes et al., 2004; Werner et al., 2007), including depression (Everson et al., 2002), and a genetic correlation between SES and depression has been also observed (Hill et al., 2016), a gene-environment-trait correlation in which SES mediates the genetic correlation between EA and depression, is possible.

A recent genome wide association study (GWAS) of EA (Lee et al., 2018) included about 1.1 million European-descent participants, making it one of the most powerful, and consequently prevalently used, GWASs in psychology. A polygenic score based on the summary statistics from this GWAS explained about 11% of the variance in EA. In the current study, I tested whether a polygenic score derived from the latter EA GWAS will be associated with an individual's SES, which in turn will be

associated with their depressive symptoms. Two independent samples were used: a discovery sample of 522 non-Hispanic Caucasian university students from the Duke Neurogenetics Study and a replication sample of 5,243 adult white British volunteers from the UK Biobank (UKB). As the GWAS included data from the UKB and this may bias the results, in the analyses of the UKB data I also used EA polygenic scores that were based on summary statistics from a GWAS that did not include the UKB as a discovery sample (obtained from Dr. Aysu Okbay, who is one of the authors of the original GWAS).

Materials and Methods

Participants

The discovery sample consisted of 522 self-reported non-Hispanic Caucasian participants (277 women, mean age 19.78±1.24 years) from the Duke Neurogenetics Study (DNS) who were not related and for whom there was complete data on genotypes, SES, depressive symptoms, and all covariates. Participants were recruited through posted flyers on the Duke University campus and through a Duke University listserv. All procedures were approved by the Institutional Review Board of the Duke University Medical Center, and participants provided informed consent before study initiation. All participants were free of the following study exclusions: 1) medical diagnoses of cancer, stroke, diabetes requiring insulin treatment, chronic kidney or liver disease, or lifetime history of psychotic symptoms; 2) use of psychotropic, glucocorticoid, or hypolipidemic medication; and 3) conditions affecting cerebral blood flow and metabolism (e.g., hypertension).

The replication sample consisted of 5,243 white British volunteers (2,669 women, mean age 62.30±7.41 years), who participated in the UKB's first assessment and the imaging wave, completed an online mental health questionnaire (Davis et al., 2018), and had complete genotype, SES, depressive symptoms and covariate data. The UKB (www.ukbiobank.ac.uk; Sudlow et al., 2015) includes over 500,000 participants, between the ages of 40 and 69 years, who were recruited within the UK between 2006 and 2010. The UKB study has been approved by the National Health Service Research Ethics Service (reference: 11/NW/0382), and current analyses were conducted under UKB application 28174.

Race/Ethnicity

Because self-reported race and ethnicity are not always an accurate reflection of genetic ancestry, an analysis of identity by state of whole-genome SNPs in the DNS was performed in PLINK (Purcell et al., 2007). Before running the multidimensional scaling components analysis, SNPs were pruned for high LD (r^2>0.1), and the following were removed: C/G and A/T SNPs, SNPs with a missing rate >.05 or a minor allele frequency <.01, SNPs that did not pass the Hardy-Weinberg equilibrium test (p<1e-6), sex chromosomes, and regions with long range LD (the MHC and 23 additional regions; Price et al., 2008). The first two multidimensional scaling components computed for the non-Hispanic Caucasian subgroup, as determined by both self-reports and the multidimensional scaling components of the entire mixed race/ethnicity DNS sample, were used as covariates in analyses of data from the DNS. The decision to use only the first two components was based on an examination of a scree plot of the variance explained by each component. For

analyses of data from the UKB, only those who were 'white British' based on both self-identification and a genetic principal components analysis were included. Additionally, the first 10 multidimensional scaling components received from the UKB's data repository (unique data identifiers: 22009-0.1-22009-0.10) were included as covariates as previously done (e.g., Avinun and Hariri, 2019; Whalley et al., 2016). Further details on the computation of the multidimensional scaling components can be found elsewhere (http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf).

Socioeconomic status

In the DNS, SES was assessed using the "social ladder" instrument (Adler et al., 2000), which asks participants to rank themselves relative to other people in the United States (or their origin country) on a scale from 0–10, with people who are best off in terms of money, education, and respected jobs, at the top (10) and people who are worst off at the bottom (0). In the UKB, SES was assessed based on the report of average household income before tax, coded as: 1 - Less than 18,000; 2 - 18,000 to 31,000; 3 - 31,000 to 52,000; 4 - 52,000 to 100,000; and 5 - Greater than 100,000. The reports made during the first assessment (i.e., before the evaluation of depressive symptoms), between 2006 and 2010, were used.

Depressive symptoms

In the DNS, the 20-item Center for Epidemiologic Studies Depression Scale (CES-D) was used to asses depressive symptoms in the past week (Radloff, 1977). All items

were summed to create a total depressive symptoms score. In the UKB, the Patient Health Questionnaire 9-question version (PHQ-9) was used to asses depressive symptoms in the past 2 weeks (Kroenke et al., 2001). The participants answered these questions during a follow-up between 2016 and 2017. All items were summed to create a total depressive symptoms score.

Genotyping

In the DNS, DNA was isolated from saliva using Oragene DNA self-collection kits (DNA Genotek) customized for 23andMe (www.23andme.com). DNA extraction and genotyping were performed through 23andMe by the National Genetics Institute (NGI), a CLIA-certified clinical laboratory and subsidiary of Laboratory Corporation of America. One of two different Illumina arrays with custom content was used to provide genome-wide SNP data, the HumanOmniExpress (N=328) or HumanOmniExpress-24 (N=194; Do et al., 2011; Eriksson et al., 2010; Tung et al., 2011). In the UKB, samples were genotyped using either the UK BiLEVE (N=501) or the UKB axiom (N=4,742) array. Details regarding the UKB's quality control can be found elsewhere (Bycroft et al., 2017).

Quality control and polygenic scoring

For genetic data from both the DNS and UKB, PLINK v1.90 (Purcell et al., 2007) was used to apply quality control cutoffs and exclude SNPs or individuals based on the following criteria: missing genotype rate per individual >.10, missing rate per SNP >.10, minor allele frequency <.01, and Hardy-Weinberg equilibrium p<1e-6. Additionally, in the UKB, quality control variables that were provided with the

dataset were used to exclude participants based on a sex mismatch (genetic sex different from reported sex), a genetic relationship to another participant, outliers for heterozygosity or missingness (unique Data Identifier 22010-0.0), and UKBiLEVE genotype quality control for samples (unique Data Identifiers 22050-0.0-22052-0.0).

Polygenic scores were calculated using PLINK's (Purcell et al., 2007) "--score" command based on published SNP-level summary statistics from the most recent EA GWAS (Lee et al., 2018). Published summary statistics do not include the data from 23andMe per the requirements of this company. SNPs from the GWAS of EA were matched with SNPs from the DNS and the UKB. For each SNP the number of the alleles (0, 1, or 2) associated with EA was multiplied by the effect estimated in the GWAS. The polygenic score for each individual was an average of weighted EA-associated alleles. All SNPs matched with SNPs from the DNS and UKB were used regardless of effect size and significance in the original GWAS, as previously recommended and shown to be effective (Dudbridge, 2013; Ware et al., 2017).

Statistical analysis

The PROCESS SPSS macro, version 3.1 (Hayes, 2017), was used to conduct the mediation analyses in SPSS version 26. Participants' sex (coded as 0=males, 1=females), age, and genetic principal components (two for the DNS and 10 for the UK biobank) were entered as covariates in all analyses. In the mediation analyses, bias-corrected bootstrapping (set to 5,000) was used to allow for non-symmetric 95% confidence intervals (CIs). Specifically, indirect effects are likely to have a non-normal distribution, and consequently the use of non-symmetric CIs for the determination of significance is recommended (MacKinnon et al., 2004). However,

bias-corrected bootstrapping also has its faults (Hayes and Scharkow, 2013) and, consequently, as supportive evidence for the indirect effect, I also present the test of joint significance, which examines whether the *a path* (EA polygenic scores to SES) and the *b path* (SES to depressive symptoms, while controlling for the EA polygenic scores) are significant. The EA polygenic scores were standardized (i.e., M=0, SD=1) in SPSS to make interpretability easier. The mediation was first tested in the DNS, and then a replication was tested in the UKB. As a validation of the indirect effect in the UKB, it was also tested with EA polygenic scores that were not based on a GWAS that included the UKB. Notably, as these polygenic scores are based on a smaller sample GWAS, they are weaker predictors of EA. Additionally, to further test the robustness of the effect, in the UKB it was possible to analyze the longitudinal data while excluding those who reported on ever seeing a general physician (N=1,843) or a psychiatrist (N=501) "for nerves, anxiety, tension or depression", at the first assessment.

Results

Descriptive statistics

In the DNS, the SES measure ranged between 2 and 10 (M=7.34, SD=1.43) and depressive symptoms ranged between 0 and 43 (M=8.94, SD=7.13). In the UKB, the SES measure ranged between 1 and 5 (M=2.92, SD=1.11), and depressive symptoms, estimated about 6 years later, ranged between 0 and 27 (M=2.50, SD=3.43).

EA polygenic scores and SES (a path) in the DNS

The EA polygenic scores were significantly associated with SES (b=.20, SE=.06, p=.0016; R^2=0.018), so that higher scores predicted higher SES. Of the covariates, age and sex were significantly associated with SES, so that older participants (b=.13, SE=.05, p=.008) and men (b=-.45, SE=.12, p=.0003) were characterized by higher SES.

SES and depressive symptoms (b path) in the DNS

With the EA polygenic scores in the model, SES significantly and negatively predicted depressive symptoms (b=-.61, SE=.22, p=.007; R^2=0.014), such that higher SES predicted lower depressive symptoms. Of the covariates, age was significantly associated with depressive symptoms, so that being younger was associated with higher depressive symptoms (b=-.53, SE=.25, p=.037).

EA polygenic scores and depressive symptoms in the DNS

The EA polygenic scores did not significantly predict depressive symptoms (b=-.11, SE=.32, p=.74). Notably, however, the significance of a direct path from X (EA polygenic scores) to Y (depressive symptoms) or the 'total effect' (the 'c' path), is not a prerequisite for the testing of a mediation/indirect effect (Hayes, 2009; MacKinnon et al., 2000; Rucker et al., 2011), which was the main interest of the current study.

Indirect Effects in the DNS

The indirect path (*a*b*), EA polygenic scores to depressive symptoms via SES was significant as indicated by the bias corrected bootstrapped 95% CI not including zero (Figure 1a; indirect effect=-.12, bootstrapped SE=.06, bootstrapped 95% CI: -.26 to -.02).

Indirect Effects in the UBK

The *a path*, from the EA polygenic scores to SES, and the *b path*, from SES to depressive symptoms while controlling for EA polygenic scores, were significant (*a path*: b=.17, SE=.01, p<.0001, R^2=0.022; *b path*: b=-.42, SE=.04, p<.0001, R^2=0.016). The indirect path also replicated (Figure 1b; indirect effect=-.07, bootstrapped SE=.01, bootstrapped 95% CI: -.091 to -.051). Similar results were obtained with the EA polygenic scores that were based on a GWAS that did not include the UKB as a discovery sample (*a path*: b=.10, SE=.01, p<.0001, R^2=0.008; *b path*: b=-.43, SE=.04, p<.0001, R^2=0.017; indirect effect=-.04, bootstrapped SE=.008, bootstrapped 95% CI: -.06 to -.03). An analysis that excluded participants who, at the first assessment, reported on ever seeing a professional for nerves or depression (leaving 3,447 participants), and that relied on the EA polygenic scores that were based on a GWAS that excluded the UKB, further supported a causal mediation, in which higher EA polygenic scores predicted higher SES, which in turn predicted lower depressive symptoms (*a path*: b=.08, SE=.02, p<.0001, R^2=0.005; *b path*: b=-.15, SE=.04, p=.0003, R^2=0.004; indirect effect=-.012, bootstrapped SE=.004, bootstrapped 95% CI: -.022 to -.005).

Discussion

The results of the current study show that EA polygenic scores are associated with depressive symptoms partly through SES, such that individuals with higher EA polygenic scores, are more likely to be of higher SES, and in turn less likely to experience depressive symptoms. This indirect effect was found in two independent

samples with different characteristics and measures, demonstrating the robustness of the associations. Notably, in the UKB the indirect effect was tested longitudinally, with data on SES that was collected about 6 years before the assessment of depressive symptoms. A supplementary analysis that excluded participants who reported ever seeing a professional for nerves or depression at time point 1, was also significant, further supporting a causal mediation.

The found mediation supports the gene-environment-trait correlations hypothesis (rGET; Avinun, in press), which suggests that certain genetic correlations may be mediated, at least in part, by the environment, i.e., an environmentally mediated pleiotropy. The found EA polygenic scores→SES→depressive symptoms mediation stresses the importance of context in genetic studies of depression. In other words, the current results suggest that, for example, a GWAS of depression that relies mostly on participants with a higher SES, may miss the genetic influences that contribute to depression through lower SES. Consequently, polygenic scores that will be based on such a GWAS will be weaker predictors of depression in low SES samples. Furthermore, the current results imply that social policies aimed at reducing socioeconomic inequalities and the negative factors that correlate with low SES may weaken the genetic effects on depression.

Low SES may lead to depression by adding stress to one's life. Stress that stems from having to make ends meet and from living in a disadvantaged neighborhood, which is associated with higher crime and fewer resources (Santiago et al., 2011). Low SES has also been associated with poorer access to green spaces (Dai, 2011), and with health damaging behaviors, such as physical inactivity, higher alcohol consumption, and poor nutrition (Nandi et al., 2014; Pampel et al., 2010), which are

thought to affect mental health (e.g., Avinun and Hariri, 2019; Beyer et al., 2014; Boden and Fergusson, 2011). All of these risk factors can be possible targets for policy makers.

The strengths of the current study include the use of two independent samples with markedly different measures and characteristics (e.g., young university students versus older community volunteers) and a GWAS-derived polygenic score, but it is also limited in ways that can be addressed in future studies. First, the findings are limited to populations of European descent and to the Western culture. Second, both samples consisted of volunteers and consequently do not fully represent the general population. However, it may be speculated that the observed associations would strengthen with the inclusion of more individuals from low SES backgrounds, which are usually characterized by higher levels of depression (Lorant et al., 2003). Third, the mediation model should be replicated within longitudinal designs in which measures of SES and depressive symptoms are available at multiple time points.

In conclusion, the current results shed light on the genetic associations that have been observed between EA and depression (Wray et al., 2018), and suggest that a part of this association may be mediated by SES. The mediation by SES is important because it suggests that the genetic influences on depression may be moderated by public policy. In addition, the current findings suggest that the genetic composition of depression is likely to depend on the social context in which it is examined.

References

Adler, N.E., Epel, E.S., Castellazzo, G., Ickovics, J.R., 2000. Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. Health Psychol. 19(6), 586.

Avinun, R., in press. The E is in the G: Gene-Environment-Trait Correlations and Findings from Genome-Wide Association Studies. Perspectives on Psychological Science.

Avinun, R., Hariri, A.R., 2019. A Polygenic Score for Body Mass Index is Associated with Depressive Symptoms via Early Life Stress: Evidence for Gene-Environment Correlation. J. Psychiatr. Res. 118, 9-13.

Avinun, R., Knafo, A., 2014. Parenting as a Reaction Evoked by Children's Genotype A Meta-Analysis of Children-as-Twins Studies. Pers. Soc. Psychol. Rev. 18(1), 87-102.

Beyer, K., Kaltenbach, A., Szabo, A., Bogar, S., Nieto, F., Malecki, K., 2014. Exposure to neighborhood green space and mental health: evidence from the survey of the health of Wisconsin. Int. J. Environ. Res. Public. Health 11(3), 3453-3472.

Boden, J.M., Fergusson, D.M., 2011. Alcohol and depression. Addiction 106(5), 906-914.

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., 2015. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47(11), 1236.

305 Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A.,
306 Vukcevic, D., Delaneau, O., O'Connell, J., 2017. Genome-wide genetic data on~
307 500,000 UK Biobank participants. bioRxiv, 166298.
308 Calixto, O.-J., Anaya, J.-M., 2014. Socioeconomic status. The relationship with health
309 and autoimmune diseases. Autoimmun. Rev. 13(6), 641-654.
310 Crespo, L., López-Noval, B., Mira, P., 2014. Compulsory schooling, education,
311 depression and memory: New evidence from SHARELIFE. Economics of
312 Education Review 43, 36-46.
313 Dai, D., 2011. Racial/ethnic and socioeconomic disparities in urban green space
314 accessibility: Where to intervene? Landscape and Urban Planning 102(4), 234-
315 244.
316 Davis, K.A., Coleman, J.R., Adams, M., Allen, N., Breen, G., Cullen, B., Dickens, C., Fox,
317 E., Graham, N., Holliday, J., 2018. Mental health in UK Biobank: development,
318 implementation and results from an online questionnaire completed by 157
319 366 participants. BJPsych Open 4(3), 83-90.
320 Do, C.B., Tung, J.Y., Dorfman, E., Kiefer, A.K., Drabant, E.M., Francke, U., Mountain,
321 J.L., Goldman, S.M., Tanner, C.M., Langston, J.W., 2011. Web-based genome-
322 wide association study identifies two novel loci and a substantial genetic
323 component for Parkinson's disease. PLoS Genet. 7(6), e1002141.
324 Dudbridge, F., 2013. Power and predictive accuracy of polygenic risk scores. PLoS
325 Genet. 9(3), e1003348.
326 Eriksson, N., Macpherson, J.M., Tung, J.Y., Hon, L.S., Naughton, B., Saxonov, S., Avey,
327 L., Wojcicki, A., Pe'er, I., Mountain, J., 2010. Web-based, participant-driven

328 studies yield novel genetic associations for common traits. PLoS Genet. 6(6),
329 e1000993.
330 Everson, S.A., Maty, S.C., Lynch, J.W., Kaplan, G.A., 2002. Epidemiologic evidence for
331 the relation between socioeconomic status and depression, obesity, and
332 diabetes. J. Psychosom. Res. 53(4), 891-895.
333 Ferrari, A., Somerville, A., Baxter, A., Norman, R., Patten, S., Vos, T., Whiteford, H.,
334 2013. Global variation in the prevalence and incidence of major depressive
335 disorder: a systematic review of the epidemiological literature. Psychol. Med.
336 43(3), 471-481.
337 Galobardes, B., Lynch, J.W., Davey Smith, G., 2004. Childhood socioeconomic
338 circumstances and cause-specific mortality in adulthood: systematic review
339 and interpretation. Epidemiol. Rev. 26(1), 7-21.
340 Hayes, A.F., 2009. Beyond Baron and Kenny: Statistical mediation analysis in the new
341 millennium. Communication monographs 76(4), 408-420.
342 Hayes, A.F., 2017. Introduction to mediation, moderation, and conditional process
343 analysis: A regression-based approach. Guilford Publications.
344 Hayes, A.F., Scharkow, M., 2013. The relative trustworthiness of inferential tests of
345 the indirect effect in statistical mediation analysis: Does method really matter?
346 Psychol. Sci. 24(10), 1918-1927.
347 Hill, W.D., Hagenaars, S.P., Marioni, R.E., Harris, S.E., Liewald, D.C., Davies, G., Okbay,
348 A., McIntosh, A.M., Gale, C.R., Deary, I.J., 2016. Molecular genetic
349 contributions to social deprivation and household income in UK Biobank. Curr.
350 Biol. 26(22), 3083-3089.

351 Kendler, K.S., Baker, J.H., 2007. Genetic influences on measures of the environment:
352 a systematic review. Psychol. Med. 37(5), 615-626.
353 Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief
354 depression severity measure. J. Gen. Intern. Med. 16(9), 606-613.
355 Lee, J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.,
356 Bowers, P., Sidorenko, J., Karlsson, L.R., 2018. Gene discovery and polygenic
357 prediction from a genome-wide association study of educational attainment in
358 1.1 million individuals. Nat. Genet. 50(8), 1112–1121.
359 Lorant, V., Deliège, D., Eaton, W., Robert, A., Philippot, P., Ansseau, M., 2003.
360 Socioeconomic inequalities in depression: a meta-analysis. Am. J. Epidemiol.
361 157(2), 98-112.
362 MacKinnon, D.P., Krull, J.L., Lockwood, C.M., 2000. Equivalence of the mediation,
363 confounding and suppression effect. Prev. Sci. 1(4), 173-181.
364 MacKinnon, D.P., Lockwood, C.M., Williams, J., 2004. Confidence Limits for the
365 Indirect Effect: Distribution of the Product and Resampling Methods.
366 Multivariate Behavioral Research 39(1), 99.
367 Marioni, R.E., Davies, G., Hayward, C., Liewald, D., Kerr, S.M., Campbell, A., Luciano,
368 M., Smith, B.H., Padmanabhan, S., Hocking, L.J., 2014. Molecular genetic
369 contributions to socioeconomic status and intelligence. Intelligence 44, 26-32.
370 Mathers, C.D., Loncar, D., 2006. Projections of global mortality and burden of disease
371 from 2002 to 2030. PLoS Med. 3(11), e442.
372 Nandi, A., Glymour, M.M., Subramanian, S., 2014. Association among socioeconomic
373 status, health behaviors, and all-cause mortality in the United States.
374 Epidemiology 25(2), 170-177.

375 Pampel, F.C., Krueger, P.M., Denney, J.T., 2010. Socioeconomic disparities in health
376 behaviors. Annual review of sociology 36, 349-370.
377 Plomin, R., Deary, I.J., 2015. Genetics and intelligence differences: five special
378 findings. Mol. Psychiatry 20(1), 98.
379 Plomin, R., DeFries, J.C., Loehlin, J.C., 1977. Genotype-environment interaction and
380 correlation in the analysis of human behavior. Psychol. Bull. 84(2), 309-322.
381 Price, A.L., Weale, M.E., Patterson, N., Myers, S.R., Need, A.C., Shianna, K.V., Ge, D.,
382 Rotter, J.I., Torres, E., Taylor, K.D., 2008. Long-range LD can confound genome
383 scans in admixed populations. The American Journal of Human Genetics 83(1),
384 132-135.
385 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller,
386 J., Sklar, P., De Bakker, P.I., Daly, M.J., 2007. PLINK: a tool set for whole-
387 genome association and population-based linkage analyses. The American
388 Journal of Human Genetics 81(3), 559-575.
389 Radloff, L.S., 1977. The CES-D scale: A self-report depression scale for research in the
390 general population. Applied psychological measurement 1(3), 385-401.
391 Rucker, D.D., Preacher, K.J., Tormala, Z.L., Petty, R.E., 2011. Mediation analysis in
392 social psychology: Current practices and new recommendations. Social and
393 Personality Psychology Compass 5(6), 359-371.
394 Santiago, C.D., Wadsworth, M.E., Stump, J., 2011. Socioeconomic status,
395 neighborhood disadvantage, and poverty-related stress: Prospective effects on
396 psychological syndromes among diverse low-income families. Journal of
397 Economic Psychology 32(2), 218-230.

398 Scarr, S., McCartney, K., 1983. How people make their own environments: A theory
399 of genotype environment effects. Child Dev. 54(2), 424-435.
400 Strenze, T., 2007. Intelligence and socioeconomic success: A meta-analytic review of
401 longitudinal research. Intelligence 35(5), 401-426.
402 Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott,
403 P., Green, J., Landray, M., 2015. UK biobank: an open access resource for
404 identifying the causes of a wide range of complex diseases of middle and old
405 age. PLoS Med. 12(3), e1001779.
406 Tung, J.Y., Do, C.B., Hinds, D.A., Kiefer, A.K., Macpherson, J.M., Chowdry, A.B.,
407 Francke, U., Naughton, B.T., Mountain, J.L., Wojcicki, A., 2011. Efficient
408 replication of over 180 genetic associations with self-reported medical data.
409 PLoS ONE 6(8), e23473.
410 Ware, E.B., Schmitz, L.L., Faul, J.D., Gard, A., Mitchell, C., Smith, J.A., Zhao, W., Weir,
411 D., Kardia, S.L., 2017. Heterogeneity in polygenic scores for common human
412 traits. bioRxiv, 106062.
413 Werner, S., Malaspina, D., Rabinowitz, J., 2007. Socioeconomic status at birth is
414 associated with risk of schizophrenia: population-based multilevel study.
415 Schizophr. Bull. 33(6), 1373-1378.
416 Whalley, H.C., Adams, M.J., Hall, L., Clarke, T.-K., Fernandez-Pujals, A.M., Gibson, J.,
417 Wigmore, E., Hafferty, J., Hagenaars, S.P., Davies, G., 2016. Dissection of major
418 depressive disorder using polygenic risk scores for schizophrenia in two
419 independent cohorts. Translational Psychiatry 6(11), e938.
420 Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A.,
421 Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., Bacanu, S.-A., Bækvad-

422 Hansen, M., Beekman, A.F.T., Bigdeli, T.B., Binder, E.B., Blackwood, D.R.H.,
423 Bryois, J., Buttenschøn, H.N., Bybjerg-Grauholm, J., Cai, N., Castelao, E.,
424 Christensen, J.H., Clarke, T.-K., Coleman, J.I.R., Colodro-Conde, L., Couvy-
425 Duchesne, B., Craddock, N., Crawford, G.E., Crowley, C.A., Dashti, H.S., Davies,
426 G., Deary, I.J., Degenhardt, F., Derks, E.M., Direk, N., Dolan, C.V., Dunn, E.C.,
427 Eley, T.C., Eriksson, N., Escott-Price, V., Kiadeh, F.H.F., Finucane, H.K., Forstner,
428 A.J., Frank, J., Gaspar, H.A., Gill, M., Giusti-Rodríguez, P., Goes, F.S., Gordon,
429 S.D., Grove, J., Hall, L.S., Hannon, E., Hansen, C.S., Hansen, T.F., Herms, S.,
430 Hickie, I.B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.-J., Hougaard,
431 D.M., Hu, M., Hyde, C.L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles,
432 J.A., Kohane, I.S., Kraft, J., Kretzschmar, W.W., Krogh, J., Kutalik, Z., Lane, J.M.,
433 Li, Y., Li, Y., Lind, P.A., Liu, X., Lu, L., MacIntyre, D.J., MacKinnon, D.F., Maier,
434 R.M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland,
435 S.E., Mehta, D., Middeldorp, C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mill,
436 J., Mondimore, F.M., Montgomery, G.W., Mostafavi, S., Mullins, N., Nauck, M.,
437 Ng, B., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J.,
438 Painter, J.N., Pedersen, C.B., Pedersen, M.G., Peterson, R.E., Pettersson, E.,
439 Peyrot, W.J., Pistis, G., Posthuma, D., Purcell, S.M., Quiroz, J.A., Qvist, P., Rice,
440 J.P., Riley, B.P., Rivera, M., Saeed Mirza, S., Saxena, R., Schoevers, R., Schulte,
441 E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.B.C., Smit, J.H.,
442 Smith, D.J., Stefansson, H., Steinberg, S., Stockmeier, C.A., Streit, F.,
443 Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W.,
444 Thomson, P.A., Thorgeirsson, T.E., Tian, C., Traylor, M., Treutlein, J.,
445 Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Van der Auwera, S., van

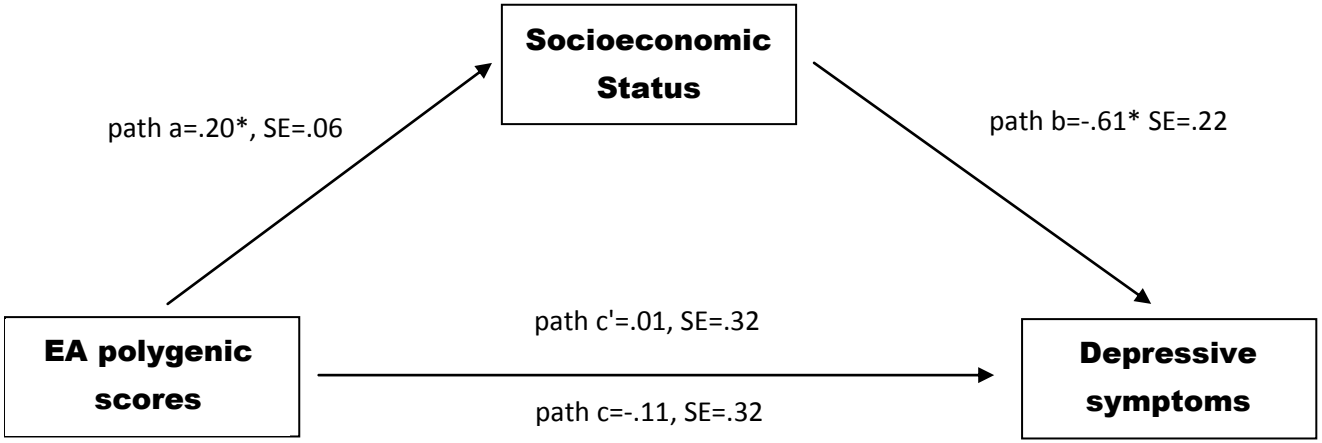
446 Hemert, A.M., Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer,
447 S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F.,
448 Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., de
449 Geus, E.C.J., DePaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J.,
450 Hamilton, S.P., Hayward, C., Heath, A.C., Hinds, D.A., Kendler, K.S., Kloiber, S.,
451 Lewis, G., Li, Q.S., Lucae, S., Madden, P.F.A., Magnusson, P.K., Martin, N.G.,
452 McIntosh, A.M., Metspalu, A., Mors, O., Mortensen, P.B., Müller-Myhsok, B.,
453 Nordentoft, M., Nöthen, M.M., O'Donovan, M.C., Paciga, S.A., Pedersen, N.L.,
454 Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel,
455 M., Schaefer, C., Schulze, T.G., Smoller, J.W., Stefansson, K., Tiemeier, H., Uher,
456 R., Völzke, H., Weissman, M.M., Werge, T., Winslow, A.R., Lewis, C.M.,
457 Levinson, D.F., Breen, G., Børglum, A.D., Sullivan, P.F., eQtlgen, 23andMe, the
458 Major Depressive Disorder Working Group of the Psychiatric Genomics
459 Consortium, 2018. Genome-wide association analyses identify 44 risk variants
460 and refine the genetic architecture of major depression. Nat. Genet. 50(5),
461 668-681.

462

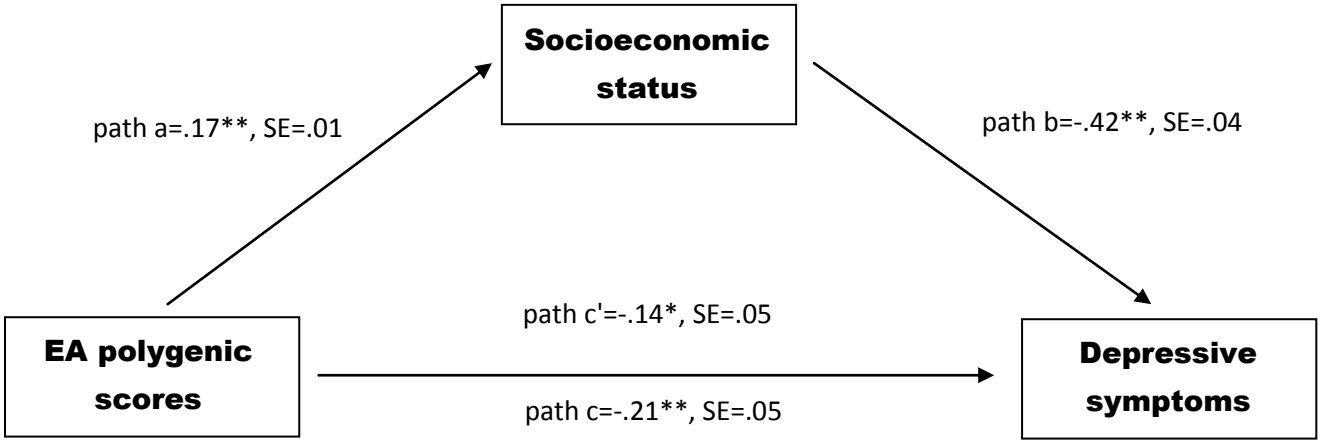
463

Figure 1. Mediation model linking genetic influences on EA to depressive symptoms, via socioeconomic status

1a. Duke Neurogenetics Study: Discovery sample



1b. UK Biobank: Replication sample



Note. *p<.01, **p<.0001. c- the total effect of the EA polygenic scores on depressive symptoms; c'-the effect of EA polygenic scores on depressive symptoms, while controlling for SES.