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., Department of Psychology, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, 91905 Israel. Email: reut.avinun@mail.huji.ac.il

Abstract

Depression is genetically influenced, but the mechanisms that underlie these influences are unknown. Recently, shared genetic influences were found between depression and educational attainment (EA), which is often used as a proxy for cognitive ability. Although genetic influences are often thought to represent direct biological pathways, they can also reflect indirect pathways, including modifiable environmental mediations (gene-environment-trait correlations). I set out to test whether the genetic correlation between EA and depressive symptoms partly reflects an environmental mediation involving socioeconomic status (SES). Summary statistics from a recent genome-wide association study of EA were used to calculate EA polygenic scores. Two independent samples were used: 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. Results indicated that SES partly mediated the genetic association between EA and depressive symptoms in both samples, such that higher EA polygenic scores predicted higher SES, which in turn predicted lower depressive symptoms. Current findings suggest that some of the genetic correlates of depressive symptoms depend on an environmental mediation and consequently that modifying the environment, specifically through social and economic policies, can affect the genetic influences on depression. Additionally, these results suggest that findings from genetic association studies of depression may be context-contingent and reflect social, cultural, and economic processes in the examined population.

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

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 & Loncar, 2006). The heritability of major depression has been estimated to be about 30-40% (Kendler, Gatz, Gardner, & Pedersen, 2006; Sullivan, Neale, & Kendler, 2000), but the mechanisms that underlie these genetic influences are mostly unknown. 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, López-Noval, & Mira, 2014). A recent study (Wray et al., 2018) further showed that the link between EA and depression is partly due to shared genetic influences. Notably however, our understanding of the genetic link between these phenotypes is lacking.

Recently, it was hypothesized that socioeconomic status (SES) may mediate the link between EA and depression (Avinun, 2020). More generally, it was suggested that the environment may mediate genetic correlations between two phenotypes within the same individual, in a process termed gene-environment-trait correlations (Avinun, 2020). This hypothesis stems from accumulating research showing passive, active, and evocative processes that lead to correlations between genetic variation and environmental measures, such as parenting and stressful life events (Avinun & Knafo, 2014; Kendler & Baker, 2007). These passive, active, and evocative processes, known as gene-environment correlations (Plomin, DeFries, & Loehlin, 1977; Scarr & 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 within the same individual and contribute to the widespread genetic correlations observed between numerous

phenotypes (Bulik-Sullivan et al., 2015), including EA and depression (Wray et al., 2018). In other words, it is possible that genetic influences on two different phenotypes, like EA and depression, are linked, because one genetically influenced phenotype (e.g., EA) affects an environment (e.g., SES) that, in turn, affects another phenotype (e.g., depressive symptoms).

Identifying gene-environment-depression correlations can shed light on environments that play a role in pathways that connect between certain genetic variation and depression. Disrupting such pathways through public policy will modify these indirect genetic influences on depression. Additionally, such environmental mediations can demonstrate the importance of context in the discovery of the genetic correlates of depression, because different contexts can translate into different environmental mediations.

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 & Anaya, 2014), has been shown to be genetically influenced (Hill et al., 2016; Marioni et al., 2014). Put differently, 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 & Deary, 2015) and highly genetically correlated with EA, indicating shared genetic influences (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 & Anaya, 2014; Galobardes, Lynch, & Davey Smith, 2004; Werner, Malaspina, & Rabinowitz, 2007), including depression (Everson, Maty, Lynch, & Kaplan, 2002), and a genetic correlation between SES and depression has also been 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 ~1.1 million European-descent participants, making it one of the most powerful, and consequently prevalently used, GWASs in psychology (for comparison, a recent GWAS of intelligence included 269,867 individuals; Savage et al., 2018). A polygenic score based on the summary statistics from this GWAS explained ~11% of the variance in EA. In the current study, I tested whether SES mediated an association between EA polygenic scores and depressive symptoms. The EA polygenic score was used as a proxy for a cognitive ability polygenic score, as has been previously done to increase statistical power in genetic studies (e.g., Hill et al., 2019; Rietveld et al., 2014), because of their high phenotypic and genetic correlation (Calvin et al., 2012; Marioni et al., 2014).

Two independent samples were used for the analyses: a sample of 522 non-Hispanic Caucasian university students from the Duke Neurogenetics Study and a sample of 5,243 adult white British volunteers from the UK Biobank (UKB). The two samples complement each other in several ways: 1) the measures used for the assessment of SES and depressive symptoms differed in the two samples as will be detailed below, and therefore finding a significant mediation in both samples would suggest that the result is robust to different operationalizations of these two constructs; 2) the two samples represented different age groups, young adulthood and older adulthood, and therefore finding the hypothesized mediation in both samples could show that it is not specific to a particular age range; 3) in the DNS all participants were students at the same university, and therefore finding variation in their EA polygenic scores would further support the link between EA polygenic

scores and cognitive and non-cognitive abilities, as has been previously shown (e.g., Belsky et al., 2016); and 4) in the UKB it was possible to test a longitudinal mediation model, which can provide further support for causal inference. Lastly, as the EA GWAS included data from the UKB and this may bias the results, in the analyses of the UKB data EA polygenic scores 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 Duke Neurogenetics Study (DNS) sample consisted of 522 self-reported non-Hispanic Caucasian participants (277 women, mean age 19.78±1.24 years) 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 UKB 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 was approved by the National Health Service Research Ethics Service (reference: 11/NW/0382), and current analyses were conducted under UKB application 28174 (because the application also included a request for neuroimaging data, the sample used in this study is limited to individuals who participated in the imaging wave).

Ancestry

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 v1.9 (Purcell et al., 2007). Before running the multidimensional scaling (MDS) 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 MDS components computed for the non-Hispanic Caucasian subgroup, as determined by both self-reports and the MDS 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 MDS components was based on an examination of a scree plot of eigenvalues, which became very similar after the second MDS component (additional information and plots are available at https://www.haririlab.com/methods/genetics.html).

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 principal 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 & Hariri, 2019; Whalley et al., 2016). Further details on the computation of the principal 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, Epel, Castellazzo, & Ickovics, 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 assess 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 assess depressive symptoms in the past 2 weeks (Kroenke, Spitzer, & Williams, 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 (i.e., N ≈ 766,345). For the UKB analyses, summary scores from a GWAS that did not include the UKB as a discovery sample were used (N ≈ 324,162). 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 ancestry (two genetic components 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, Lockwood, & Williams,

2004). However, bias-corrected bootstrapping also has its faults (Hayes & 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.

As a post-hoc analysis, 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 testing a mediation/indirect effect (Hayes, 2009; MacKinnon, Krull, & Lockwood, 2000; Rucker, Preacher, Tormala, & Petty, 2011), which was the main aim 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). The indirect effect remained significant when 10 MDS components of genetic ancestry and genotyping platform were included as covariates (indirect effect=-.12, bootstrapped SE=.06, bootstrapped 95% CI: -.27 to -.02).

Indirect Effects in the UBK

For the UKB, EA polygenic scores that were based on a GWAS that did not include the UKB as a discovery sample, were used. additionally, genotyping platform was included as a covariate in the analyses. 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=.10, SE=.01, p<.0001, R^2=0.008; *b path*: b=-.43, SE=.04, p<.0001, R^2=0.017). The indirect path was also significant (Figure 1b; indirect effect=-.04, bootstrapped SE=.008, bootstrapped 95% CI: -.06 to -.03).

To test the robustness of the finding, a post-hoc analysis that excluded participants who, at the first assessment, reported on ever seeing a professional for nerves or depression (leaving 3,447 participants), was conducted. This was done in an attempt to increase the likelihood of only including the individuals who became depressed between the first assessment, in which income was assessed, and the assessment a few years later, in which depressive symptoms were assessed. This analysis 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 current results suggest that the negative genetic association previously found between EA and depressive symptoms is partly mediated by SES, an environment that can be modified through social and economic policies. The 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 first assessment, was also significant, further supporting a causal temporal mediation by predicting change in depressive symptoms.

The found mediation supports the gene-environment-trait correlations hypothesis (rGET; Avinun, 2020), 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. More specifically, the current results suggest that a GWAS of depression that relies mostly on a sample that is homogeneous in terms of SES, will be less likely to capture the genetic influences that contribute to depression via SES. Consequently, polygenic scores that will be based on such a GWAS may be weaker predictors of depression. Put differently, because the environment can act as a mediator of some of the genetic influences on depression, if the environment differs between GWASs, the captured genetic influences will differ. Importantly, the current results imply that social policies aimed at reducing socioeconomic inequalities may weaken the genetic effects on depression by decreasing the association between EA and depression. Lastly, this rGET means that statistically controlling for SES may help to remove environmentally mediated effects from the depression polygenic score and lead us a step closer to identifying the direct genetic influences on depression.

Low SES may be a risk factor for depression by leading to an increase in 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, Wadsworth, & Stump, 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, Glymour, & Subramanian, 2014; Pampel, Krueger, & Denney, 2010), which are thought to affect mental health (e.g., Avinun & Hariri, 2019; Beyer et al., 2014; Boden & Fergusson, 2011). All of these mediators 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 they are partly mediated by SES. The mediation by SES is important because it suggests that the genetic influences on depression may be moderated by public policy and that the genetic composition of depression depends on the social context in which it is examined. Additionally, the current results suggest that controlling for SES, or other environmental factors that can act as mediators of genetic influences, may help to isolate the direct genetic influences on depression.

Acknowledgements

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.

Funding

The DNS was supported by Duke University as well as US-National Institutes of Health grant R01DA033369. The author received support from a from a Lady Davis fellowship.

Conflict of Interest

The author declares no competing financial or other interests.

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 Psychology, 19*(6), 586.

Avinun, R. (2020). The E is in the G: Gene-Environment-Trait Correlations and Findings from Genome-Wide Association Studies. *Perspectives on Psychological Science, 15*(1), 81-89. doi: 10.1177/1745691619867107

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. *Journal of Psychiatric Research, 118*, 9-13. doi: 10.1016/j.jpsychires.2019.08.008

Avinun, R., & Knafo, A. (2014). Parenting as a Reaction Evoked by Children's Genotype A Meta-Analysis of Children-as-Twins Studies. *Personality and Social Psychology Review, 18*(1), 87-102.

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., . . . Williams, B. S. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science, 27*(7), 957-972.

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. *International Journal of Environmental Research and 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., . . . Robinson, E. B. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics, 47*(11), 1236.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., . . . O'Connell, J. (2017). Genome-wide genetic data on~ 500,000 UK Biobank participants. *bioRxiv*, 166298.

Calixto, O.-J., & Anaya, J.-M. (2014). Socioeconomic status. The relationship with health and autoimmune diseases. *Autoimmunity Reviews, 13*(6), 641-654.

Calvin, C. M., Deary, I. J., Webbink, D., Smith, P., Fernandes, C., Lee, S. H., . . . Visscher, P. M. (2012). Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behavior Genetics, 42*(5), 699-710.

Crespo, L., López-Noval, B., & Mira, P. (2014). Compulsory schooling, education, depression and memory: New evidence from SHARELIFE. *Economics of Education Review, 43*, 36-46.

Dai, D. (2011). Racial/ethnic and socioeconomic disparities in urban green space accessibility: Where to intervene? *Landscape and Urban Planning, 102*(4), 234-244.

Davis, K. A., Coleman, J. R., Adams, M., Allen, N., Breen, G., Cullen, B., . . . Holliday, J. (2018). Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *BJPsych Open, 4*(3), 83-90.

Do, C. B., Tung, J. Y., Dorfman, E., Kiefer, A. K., Drabant, E. M., Francke, U., . . . Langston, J. W. (2011). Web-based genome-wide association study identifies

two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genetics, 7*(6), e1002141.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genetics, 9*(3), e1003348.

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., . . . Mountain, J. (2010). Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genetics, 6*(6), e1000993.

Everson, S. A., Maty, S. C., Lynch, J. W., & Kaplan, G. A. (2002). Epidemiologic evidence for the relation between socioeconomic status and depression, obesity, and diabetes. *Journal of Psychosomatic Research, 53*(4), 891-895.

Ferrari, A., Somerville, A., Baxter, A., Norman, R., Patten, S., Vos, T., & Whiteford, H. (2013). Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. *Psychological medicine, 43*(3), 471-481.

Galobardes, B., Lynch, J. W., & Davey Smith, G. (2004). Childhood socioeconomic circumstances and cause-specific mortality in adulthood: systematic review and interpretation. *Epidemiologic Reviews, 26*(1), 7-21.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication monographs, 76*(4), 408-420.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*: Guilford Publications.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, 24*(10), 1918-1927.

Hill, W., Marioni, R., Maghzian, O., Ritchie, S., Hagenaars, S., McIntosh, A., . . . Deary, I. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry, 24*(2), 169-181.

Hill, W. D., Hagenaars, S. P., Marioni, R. E., Harris, S. E., Liewald, D. C., Davies, G., . . . Deary, I. J. (2016). Molecular genetic contributions to social deprivation and household income in UK Biobank. *Current Biology, 26*(22), 3083-3089.

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychological Medicine, 37*(5), 615-626.

Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish national twin study of lifetime major depression. *American Journal of Psychiatry, 163*(1), 109-114.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613.

Lee, J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., . . . Karlsson, L. R. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics, 50*(8), 1112–1121.

Lorant, V., Deliège, D., Eaton, W., Robert, A., Philippot, P., & Ansseau, M. (2003). Socioeconomic inequalities in depression: a meta-analysis. *American Journal of Epidemiology, 157*(2), 98-112.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1*(4), 173-181.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods. *Multivariate Behavioral Research, 39*(1), 99.

Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., . . . Hocking, L. J. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence, 44*, 26-32.

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLOS Medicine, 3*(11), e442.

Nandi, A., Glymour, M. M., & Subramanian, S. (2014). Association among socioeconomic status, health behaviors, and all-cause mortality in the United States. *Epidemiology, 25*(2), 170-177.

Pampel, F. C., Krueger, P. M., & Denney, J. T. (2010). Socioeconomic disparities in health behaviors. *Annual review of sociology, 36*, 349-370.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular Psychiatry, 20*(1), 98.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin, 84*(2), 309-322.

Price, A. L., Weale, M. E., Patterson, N., Myers, S. R., Need, A. C., Shianna, K. V., . . . Taylor, K. D. (2008). Long-range LD can confound genome scans in admixed populations. *The American Journal of Human Genetics, 83*(1), 132-135.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement, 1*(3), 385-401.

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., . . . Lee, J. J. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences, 111*(38), 13790-13794.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5*(6), 359-371.

Santiago, C. D., Wadsworth, M. E., & Stump, J. (2011). Socioeconomic status, neighborhood disadvantage, and poverty-related stress: Prospective effects on psychological syndromes among diverse low-income families. *Journal of Economic Psychology, 32*(2), 218-230.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., . . . Coleman, J. R. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics, 50*(7), 912-919.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype environment effects. *Child Development, 54*(2), 424-435.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*(5), 401-426.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine, 12*(3), e1001779.

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry, 157*(10), 1552-1562.

Tung, J. Y., Do, C. B., Hinds, D. A., Kiefer, A. K., Macpherson, J. M., Chowdry, A. B., . . . Wojcicki, A. (2011). Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS ONE, 6*(8), e23473.

Ware, E. B., Schmitz, L. L., Faul, J. D., Gard, A., Mitchell, C., Smith, J. A., . . . Kardia, S. L. (2017). Heterogeneity in polygenic scores for common human traits. *bioRxiv*, 106062.

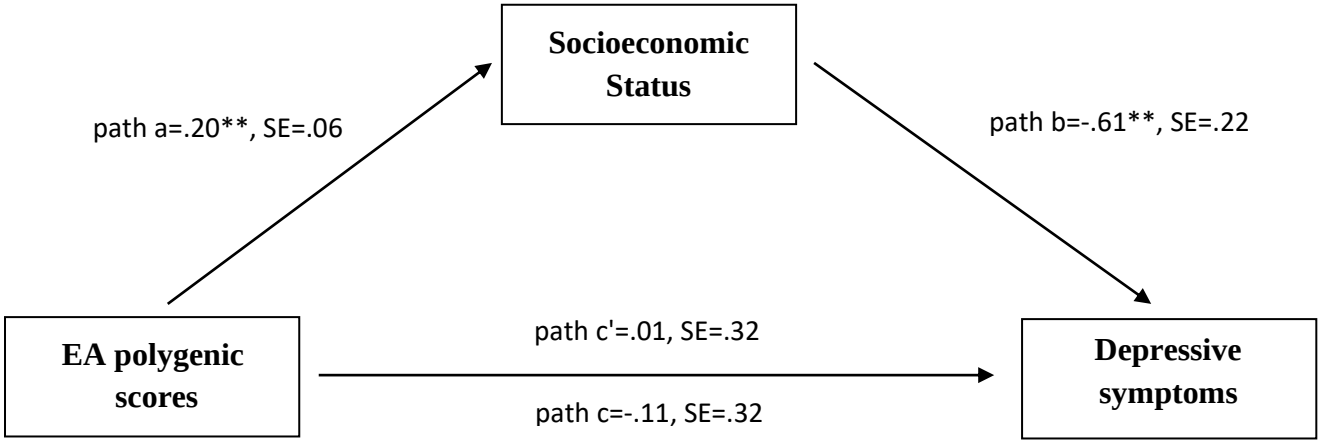
Werner, S., Malaspina, D., & Rabinowitz, J. (2007). Socioeconomic status at birth is associated with risk of schizophrenia: population-based multilevel study. *Schizophrenia Bulletin, 33*(6), 1373-1378.

Whalley, H. C., Adams, M. J., Hall, L., Clarke, T.-K., Fernandez-Pujals, A. M., Gibson, J., . . . Davies, G. (2016). Dissection of major depressive disorder using polygenic risk scores for schizophrenia in two independent cohorts. *Translational Psychiatry, 6*(11), e938.

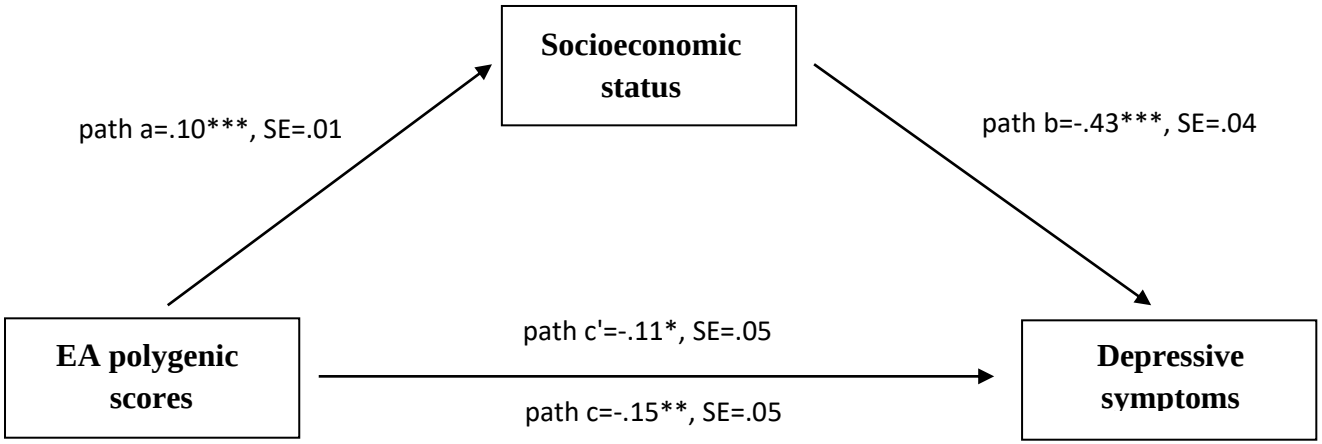
Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., . . . the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics, 50*(5), 668-681. doi: 10.1038/s41588-018-0090-3

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

1a. Duke Neurogenetics Study



1b. UK Biobank



Note. *p<.05, **p<.01, ***p<.0001. EA - educational attainment. 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.