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

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. RA received support from US-National Institutes of Health grant R01AG049789 and from a fellowship from the Jerusalem Brain Community. The Brain Imaging and Analysis Center was supported by the Office of the Director, National Institutes of Health under Award Number S10OD021480. The author declares no competing financial or other interests.

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

Abstract

Lower educational attainment (EA) has been suggested as a possible risk factor for depression, partly because of research demonstrating a genetic correlation between these phenotypes. As EA is closely related to socioeconomic status (SES), which was shown to be genetically influenced and associated with depression, I hypothesized that higher EA polygenic scores would be indirectly associated with lower depressive symptoms through higher SES. Evidence in support of this hypothesis would be consistent with a gene-environment-trait correlation (rGET), a process in which genetic influences on one phenotype also affect a different phenotype through an environment. The identification of rGET stresses the importance of environmental context and can provide support for intervention targets. Summary statistics from a recent genome-wide association study of EA were used for the calculation of the EA polygenic scores. Analyses in a discovery sample of 524 non-Hispanic Caucasian university students from the Duke Neurogenetics Study (278 women, mean age 19.78±1.23 years) revealed a significant mediation (indirect effect = -.12, bootstrapped SE=.06, bootstrapped 95% CI: -.27 to -.02), wherein higher EA polygenic scores predicted higher SES, which in turn predicted lower depressive symptoms. Analyses in an independent replication sample of 5,500 white British volunteers (2,831 women, mean age 62.40±7.42 years) from the UK biobank, confirmed this mediation (indirect effect = -.038, bootstrapped SE=.0075, bootstrapped 95% CI: -.053 to -.02). These results suggest that public policy and interventions that aim to increase 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 & 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, López-Noval, & Mira, 2014). A recent study further showed that the link between EA and depression is partly due to shared genetic influences, and determined that EA is causally related to depression (Wray et al., 2018). It is not clear what mediates this link, but previous research suggests socioeconomic status (SES) as one possibility.

SES, which can be defined as an individual's or group's position within a social hierarchy that is based on variables such as education, occupation, income, and wealth (Calixto & Anaya, 2014), 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). Interestingly, SES has been shown to be genetically influenced (Hill et al., 2016). In other words, individuals with certain genetically influenced traits are more likely to be characterized by a specific SES, a gene-environment correlation (Plomin, DeFries, & Loehlin, 1977; Scarr & McCartney, 1983). Indeed, children's EA polygenic scores have been shown to explain about 2.5% of the variance in family SES (Krapohl & Plomin, 2016).

A mediation in which EA affects depression indirectly through SES will provide evidence for a gene-environment-trait correlation (rGET; Avinun, in press). rGET

occurs when genetic influences on one phenotype also affect a different phenotype through an environment. Identifying such environmentally mediated genetic effects can provide modifiable targets for interventions and also demonstrate the importance of context in the discovery of genetic influences on a certain phenotype. In other words, if high SES partly mediates the association between high EA and low depressive symptoms, then providing a context in which high SES is more attainable, may not only lower the prevalence of depression, but also modify the genetic correlates of depression, as the genetic variants that affect EA will no longer be as strongly associated with depression.

Based on the above, I hypothesized that a polygenic score derived from the most recent genome wide association study (GWAS) of EA (Lee et al., 2018), which included about 1,131,881 European-descent participants, will be associated with the individual's SES, which in turn will be associated with depressive symptoms, consistent with an rGET. I tested this hypothesis in two independent samples: a discovery sample of 524 non-Hispanic Caucasian university students from the Duke Neurogenetics Study and a replication sample of 5,500 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

47 *Participants*

48 Our discovery sample consisted of 524 self-reported non-Hispanic Caucasian
49 participants (278 women, mean age 19.78±1.23 years) from the Duke
50 Neurogenetics Study (DNS) who were not related and for whom there was complete
51 data on genotypes, SES, depressive symptoms, and all covariates. Participants were
52 recruited through posted flyers on the Duke University campus and through a Duke
53 University listserv. All procedures were approved by the Institutional Review Board
54 of the Duke University Medical Center, and participants provided informed consent
55 before study initiation. All participants were free of the following study exclusions:
56 1) medical diagnoses of cancer, stroke, diabetes requiring insulin treatment, chronic
57 kidney or liver disease, or lifetime history of psychotic symptoms; 2) use of
58 psychotropic, glucocorticoid, or hypolipidemic medication; and 3) conditions
59 affecting cerebral blood flow and metabolism (e.g., hypertension).

60 The replication sample consisted of 5,500 white British volunteers (2,831
61 women, mean age 62.40±7.42 years), who participated in the UKB's imaging wave,
62 completed an online mental health questionnaire (Davis et al., 2018), and had
63 complete genotype, SES, depressive symptoms and covariate data. The UKB
64 (www.ukbiobank.ac.uk; Sudlow et al., 2015) includes over 500,000 participants,
65 between the ages of 40 and 69 years, who were recruited within the UK between
66 2006 and 2010. The UKB study has been approved by the National Health Service
67 Research Ethics Service (reference: 11/NW/0382), and current analyses were
68 conducted under UKB application 28174.

69

70 *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, Nevo, Knodt, Elliott, & 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, 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 in 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, 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

117 (NGI), a CLIA-certified clinical laboratory and subsidiary of Laboratory Corporation of
118 America. One of two different Illumina arrays with custom content was used to
119 provide genome-wide SNP data, the HumanOmniExpress (N=329) or
120 HumanOmniExpress-24 (N=195; Do et al., 2011; Eriksson et al., 2010; Tung et al.,
121 2011). In the UKB, samples were genotyped using either the UK BiLEVE (N=525) or
122 the UKB axiom (N=4,975) array. Details regarding the UKB's quality control can be
123 found elsewhere (Bycroft et al., 2017).
124
125 *Quality control and polygenic scoring*
126 For genetic data from both the DNS and UKB, PLINK v1.90 (Purcell et al., 2007) was
127 used to apply quality control cutoffs and exclude SNPs or individuals based on the
128 following criteria: missing genotype rate per individual >.10, missing rate per SNP
129 >.10, minor allele frequency <.01, and Hardy-Weinberg equilibrium p<1e-6.
130 Additionally, in the UKB, quality control variables that were provided with the
131 dataset were used to exclude participants based on a sex mismatch (genetic sex
132 different from reported sex), a genetic relationship to another participant, outliers
133 for heterozygosity or missingness (unique Data Identifier 22010-0.0), and UKBiLEVE
134 genotype quality control for samples (unique Data Identifiers 22050-0.0-22052-0.0).
135 Polygenic scores were calculated using PLINK's (Purcell et al., 2007) "--score"
136 command based on published SNP-level summary statistics from the most recent EA
137 GWAS (Lee et al., 2018). Published summary statistics do not include the data from
138 23andMe per the requirements of this company. SNPs from the GWAS of EA were
139 matched with SNPs from the DNS and the UKB. For each SNP the number of the
140 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 25. 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, 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. The mediation was first analyzed 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.

Results

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

The EA polygenic scores were significantly associated with SES (b=.20, SE=.06, p=.0015), 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=.01) and men (b=-.46, SE=.12, p=.0002) 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 positively predicted depressive symptoms (b=-.61, SE=.22, p=.007). Of the covariates, age was significantly associated with depressive symptoms, so that being younger was associated with higher depressive symptoms (b=-.52, SE=.25, p=.041).

EA polygenic scores and depressive symptoms in the DNS

The EA polygenic scores did not significantly predict depressive symptoms (b=-.07, SE=.32, *ns*). 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, Krull, & Lockwood, 2000; Rucker, Preacher, Tormala, & Petty, 2011), which was the main interest of the current study.

Indirect Effects in the DNS

187 The indirect path (*a*b*), EA polygenic scores to depressive symptoms via SES was
188 significant as indicated by the bias corrected bootstrapped 95% CI not including zero
189 (Figure 1a; indirect effect=-.12, bootstrapped SE=.06, 95% CI: -.27 to -.02).
190
191 *Indirect Effects in the UBK*
192 The *a path*, from the EA polygenic scores to SES, and the *b path*, from SES to
193 depressive symptoms while controlling for EA polygenic scores, were significant (*a*
194 *path*: b=.21, SE=.02, p<.0001; *b path*: b=-.18, SE=.03, p<.0001). The indirect path also
195 replicated (Figure 1b; indirect effect=-.038, bootstrapped SE=.0075, bootstrapped
196 95% CI: -.053 to -.02), supporting a mediation in which EA polygenic scores predict
197 depressive symptoms indirectly through SES. Similar results were obtained with the
198 EA polygenic scores that were based on a GWAS that did not include the UKB as a
199 discovery sample (*a path*: b=.13, SE=.02, p<.0001; *b path*: b=-.19, SE=.03, p<.0001;
200 indirect effect=-.024, bootstrapped SE=.005, bootstrapped 95% CI: -.036 to -.014).
201
202 **Discussion**
203 Here, I show that EA polygenic scores are associated with depressive symptoms
204 partly through SES, such that individuals with higher EA polygenic scores, are more
205 likely to be of higher SES, and in turn less likely to experience depressive symptoms.
206 This indirect effect was found in two independent samples with different
207 characteristics and measures, and it represents a gene-environment-trait correlation
208 (rGET; Avinun, in press), a process in which an environment mediates the genetic
209 effects of one phenotype on another phenotype. 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.

Despite the prevalence of depression and its associated individual and societal burden, it is poorly understood. The current results provide causal support for the links between EA, which is viewed as a proxy for cognitive ability, SES, and depressive symptoms. 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, Wadsworth, & Stump, 2011). Lower 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 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. The observed associations may change in at risk populations. Third, the mediation model should be replicated within longitudinal designs in which measures of SES and depressive symptoms are available across developmental stages (i.e., from childhood into adulthood).

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. These links are especially important because they can be moderated by public policy and interventions, so that, for example, improving access to green spaces, providing better access to higher education, and relieving stress by reducing crime in disadvantaged neighborhoods, could weaken the associations between EA, SES, and depression. Consequently, the current findings also suggest that the genetic composition of depression is likely to depend on the social conditions in which it was 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 Psychology, 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. *bioRxiv*. doi: 10.1101/536938

Avinun, R., Nevo, A., Knodt, A. R., Elliott, M. L., & Hariri, A. R. (2019). A genome-wide association study-derived polygenic score for interleukin-1β is associated with hippocampal volume in two samples. *Human Brain Mapping*, 1-8. doi: 10.1002/hbm.24639

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.

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.

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. 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.

Krapohl, E., & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry, 21*(3), 437.

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.

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.

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., 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.

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.

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

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.

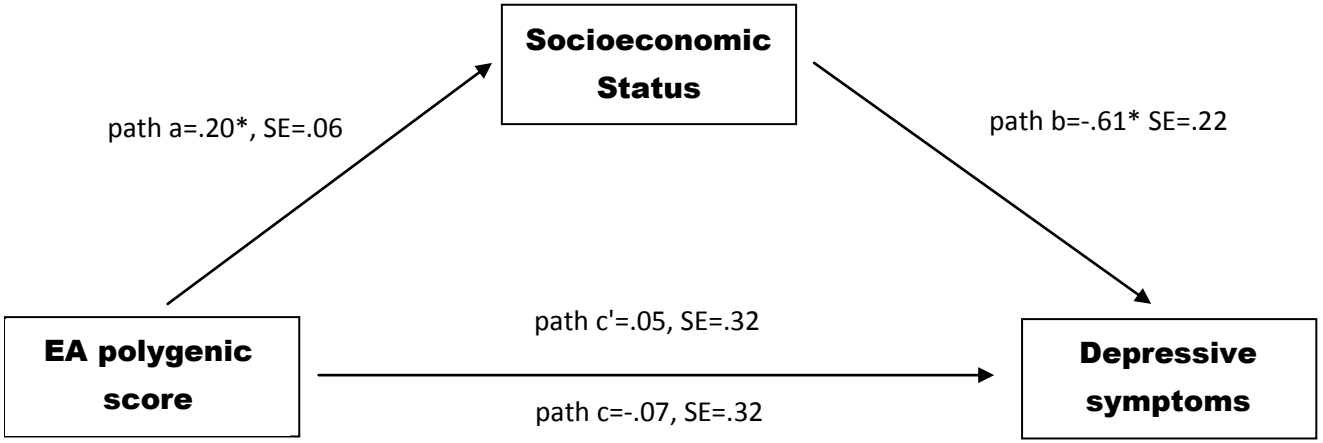
346 Tung, J. Y., Do, C. B., Hinds, D. A., Kiefer, A. K., Macpherson, J. M., Chowdry, A. B., . . .
347 Wojcicki, A. (2011). Efficient replication of over 180 genetic associations with self-
348 reported medical data. *PLoS ONE, 6*(8), e23473.
349 Ware, E. B., Schmitz, L. L., Faul, J. D., Gard, A., Mitchell, C., Smith, J. A., . . . Kardia, S. L.
350 (2017). Heterogeneity in polygenic scores for common human traits. *bioRxiv*,
351 106062.
352 Werner, S., Malaspina, D., & Rabinowitz, J. (2007). Socioeconomic status at birth is
353 associated with risk of schizophrenia: population-based multilevel study.
354 *Schizophrenia Bulletin, 33*(6), 1373-1378.
355 Whalley, H. C., Adams, M. J., Hall, L., Clarke, T.-K., Fernandez-Pujals, A. M., Gibson, J., . . .
356 Davies, G. (2016). Dissection of major depressive disorder using polygenic risk scores
357 for schizophrenia in two independent cohorts. *Translational Psychiatry, 6*(11), e938.
358 Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., . . . the
359 Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium.
360 (2018). Genome-wide association analyses identify 44 risk variants and refine the
361 genetic architecture of major depression. *Nature Genetics, 50*(5), 668-681. doi:
362 10.1038/s41588-018-0090-3

363

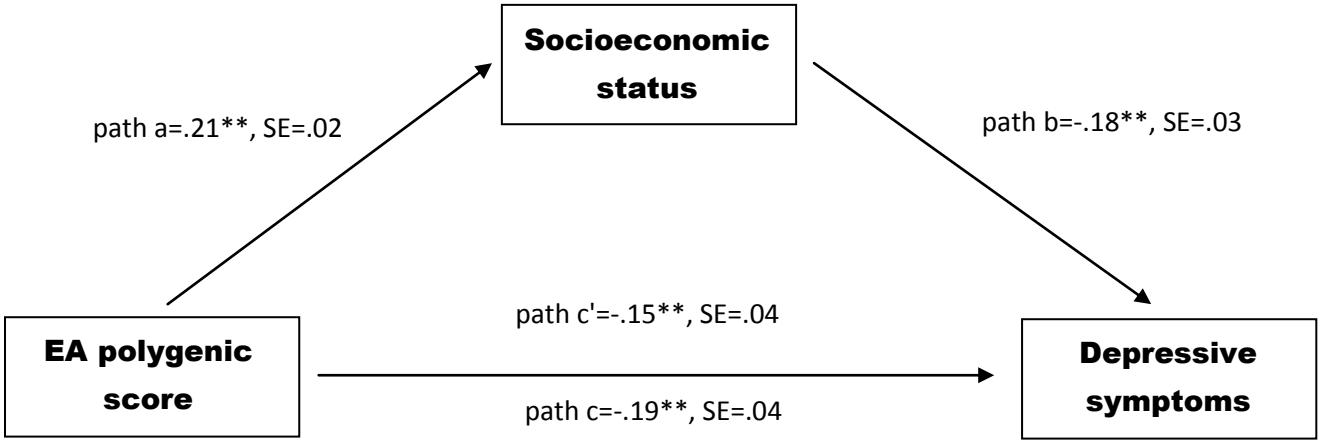
364

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<.001. 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.