The nature of nurture: Widespread covariation of early environmental exposures and trait-associated polygenic variation

E Krapohl[1]; L J Hannigan[1]; J-B Pingault[2]; H Patel[1,3]; C Curtis[1,3]; S Newhouse[1,3]; T C Eley[1]; P F O'Reilly[1]; R Plomin[1*]

[1]MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[2]Clinical, Educational and Health Psychology, University College London, London, UK
[3]NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, UK

* Corresponding authors: eva.krapohl@kcl.ac.uk; robert.plomin@kcl.ac.uk

Author Contributions: Conceived, directs, and received funding for the study: RP. Conceived and designed the analyses: EK. Analyzed data, and processed and quality controlled genotype data: EK. Performed/supervised genotyping and manual quality control and calling of genotype data: HP SN CC. Wrote the paper: EK RP. Discussed analysis strategy: EK LH JBP RP. All authors contributed to and critically reviewed the manuscript.

Abstract

Although gene-environment correlation is recognized and investigated by family studies and recently by SNP-heritability studies, the possibility that genetic effects on traits capture environmental risk factors or protective factors has been neglected by polygenic prediction models. We investigated covariation between trait-associated polygenic variation identified by genome-wide association studies (GWAS) and specific environmental exposures, controlling for overall genetic relatedness using a genomic-relatedness-matrix restricted maximum-likelihood model. In a UK-representative sample (N=6,710), we find widespread covariation between offspring trait-associated polygenic variation and parental behavior and characteristics relevant to children's developmental outcomes – independently of population stratification. For instance, offspring genetic risk for schizophrenia was associated with paternal age (R^2=0.002; P=1e-04), and offspring education-associated variation was associated with variance in breastfeeding (R^2=0.021; P=7e-30), maternal smoking during pregnancy (R^2=0.008; P=5e-13), parental smacking (R^2=0.01; P=4e-15), household income (R^2=0.032; P=1e-22), watching television (R^2=0.034; P=5e-47), and maternal education (R^2=0.065; P=3e-96). Education-associated polygenic variation also captured covariation between environmental exposures and children's inattention/hyperactivity, conduct problems, and educational achievement. The finding that genetic variation identified by trait GWAS partially captures environmental risk factors or protective factors has direct implications for risk prediction models and the interpretation of GWAS findings.

Significance Statement

Environmental exposures are among the best predictors of health and educational outcomes. Models that estimate the effect of environmental exposures on developmental outcomes typically ignore genetic factors, or focus on gene-environment interaction (whether individuals' response to environmental exposures depends on their genotype). Here we test gene-environment correlation (whether individuals' exposure to environments depends on their genotype). Using a method that tests specific genetic effects while controlling for background genetic effects, we estimate covariation between children's genetic liability/propensity for core developmental outcomes and a wide range of environmental exposures. Findings suggest that genetic variants associated with traits, such as educational attainment, body-mass index, and schizophrenia, also capture environmental risk and protective factors.

\body

Introduction

Environmental exposures are among the best early predictors of developmental outcomes. For instance, maternal smoking during pregnancy, socioeconomic deprivation, and time spent watching television and playing video games are associated with lower academic achievement (1–9). Harsh parental physical discipline such as hitting has been linked to increased emotional and behavioral problems including aggression in adolescence (10–14). Paternal age is a risk factor for a range of disorders and subclinical phenotypes including low academic achievement (15), with the link to autism spectrum disorders and schizophrenia most robustly replicated (16–21). Breastfeeding and higher parental socio-economic status (education, income, occupation) are protective factors for a range of outcomes including educational achievement (7, 8, 22).

Evidence from many family, twin, and adoption studies converges in showing that individuals' exposure to environments partially depends on their genotype (i.e. genotype-environment correlation). This includes both parenting characteristics and broad socio-economic variables; all are partially heritable (23–28). In the past decade, quantitative genetic research of this type has been extended to explore genetic and environmental contributions to correlations between environmental factors and children's outcomes (29–32). Some new designs such as the children-of-twins designs make it possible to tease apart different types of genotype-environment correlation and identify environmental influences free of genetic confounds (33–37). These designs are limited by the extent to which environmental variables differ between close relatives.

Converging evidence for gene-environment correlation comes more recently from 'single-nucleotide-polymorphism (SNP)-heritability' studies that estimate overall genetic influences from genome-wide DNA differences in unrelated individuals. These studies have shown that variation in individuals' social deprivation, household income, stressful life events, and family socio-economic status partially reflects individual' differences across genome-wide common genetic variants measured on SNP arrays (38–44). There have also been a few reports of extending SNP heritability analysis to estimate genetic correlations between environmental measures and measures of children's developmental outcomes (38–40).

Gene-environment correlation is recognized and investigated by family studies and recently by SNP-heritability studies. However, the possibility that genetic effects on traits capture environmental risk factors or protective factors has been neglected by polygenic prediction models, which use trait-associated genetic variants identified by genome-wide association studies (GWAS) to estimate genetic trait propensities for individual-level trait prediction.

Here we tested whether genetic variation identified by trait GWAS capture variation in environmental risk factors or protective factors. Specifically, as children's environments and genetic propensities are both 'provided by' their parents, these are expected to correlate because parents pass on genetic variants to their offspring that influence parents' environment-providing behaviors. Therefore, we examine to what extent offspring trait-associated alleles covary with parental traits and behaviors previously reported to be environmental risk or protective factors for important child outcomes. We also tested to what extent offspring genetic trait propensities contribute to the correlation between parenting characteristics and children's developmental outcomes.

First, we conducted a systematic investigation of covariation between children's genetic propensities for specific developmental outcomes and a wide range of environmental exposures, previously shown to be risk or protective factors for these outcomes (SI Appendix, Methods S3). We focus on genetic propensities – that is, individual-specific genomic profiles of trait-associated alleles – for three core developmental outcomes: educational attainment (45), body mass index (BMI) (46), and schizophrenia (47). These traits from three important domains of child development: social-cognitive, mental health, and physical health, each are robust predictors of mortality and life expectancy, with substantial associated societal and personal burden (48–55). They were chosen because of the availability of statistically powerful GWAS summary statistics for these traits (56).

Second, we tested whether the environmental exposures predicted children's developmental outcomes (as would be expected based on previous literature) and to what extent these associations are captured by children's polygenic propensities for education, BMI, and schizophrenia. For this, we examined associations between the environmental exposures and three developmental outcomes assessed at age 16 in our sample: educational achievement, inattention-hyperactivity symptoms, and conduct problems (SI Appendix, Methods S3).

We used a sample of 6,710 unrelated individuals, drawn from the Twins Early Development Study (TEDS), for whom genotype data and a wide range of specific environmental exposure measures and developmental outcomes from birth to adolescence are available. TEDS is a multivariate longitudinal study that recruited over 11 000 twin pairs born in England and Wales in 1994, 1995 and 1996 (57, 58), shown to be representative of the UK population (38, 59).

We created genome-wide polygenic scores for trait-associated genetic variants for each individual in the sample using summary statistics from the independent genome-wide association study (GWAS) of years of education (EDU) (45), BMI (46), and schizophrenia (SCZ) (47). We used a Bayesian approach (60) that estimates posterior mean effect size of each marker by using a point-normal mixture prior on effect sizes and linkage disequilibrium information (*Materials and Methods*).

Because of the salience of possible population stratification when investigating the genetic effect on differences in environmental exposures, we estimated the effect of the polygenic scores while controlling for overall genetic relatedness in the form of a genomic-relatedness-matrix restricted maximum-likelihood model. Specifically, we fit the effects of all SNPs as random effects, while estimating the fixed effects of the polygenic scores (*Materials and Methods*).

Results

To estimate the univariate effect of each polygenic score on the environmental exposures, we fit a series of single-score models, which reveal significant trait-associated polygenic effects across a wide range of environmental exposures. Figure 1a (and SI Appendix, Table S1) shows the estimated variance explained by each polygenic score for each of the environmental measures. Environmental factors varied significantly as a function of trait-associated polygenic variation, independently of population stratification. This provides evidence for trait-associated genotype-environment correlation. However, given the robust evidence for extensive pleiotropy across complex traits (61), we aimed to isolate the effects of each trait-associated polygenic score using a multi-score model. To test the trait-specificity of the polygenic effects on environmental exposures, we jointly modelled the three scores for years of education, BMI, and schizophrenia, allowing us to estimate the effects of each polygenic score while adjusting for the effects of the others. Figure 1b (and SI

Appendix, Table S2) shows that the multi-score models revealed some attenuation of the polygenic score effects compared to the single-score models, suggesting that the effects of the three scores on environmental exposures are non-independent. Specifically, the effects of BMI-associated polygenic variation on several environmental measures (including watching television and parental education) were no longer significant.

Breastfeeding duration was positively associated with offspring education polygenic score, adjusted for BMI and schizophrenia polygenic scores (R^2=0.021, beta=0.144; P=7e-30). Figure 2a displays children's adjusted education polygenic score as a function of whether and for how long they were breastfed. Children who were breastfed had, on average, an education polygenic score approximately one third standard deviation higher (Hedges' *g* = 0.30) than children who were not breastfed (t =-11.55, df= 5664.2, P=1.6e-30).

Maternal smoking during pregnancy was negatively associated with offspring education polygenic score adjusted for BMI and schizophrenia polygenic scores (R^2=0.008, beta=0.090; P=5e-13; Figure 2b). Children exposed to maternal smoking prenatally had, on average, an education polygenic score approximately one quarter standard deviation lower (Hedges' *g* = 0.26) than children whose mothers did not smoke (t =7.93, df=1556.3; P=4e-15).

Other effects of education-associated polygenic variation on environmental exposures included: 3.3% in household income (beta=0.181, P=1e-22), 6.5% in maternal education level (beta=0.255, P=3e-96), 1% in parental smacking (beta= -0.10, P=4e-15), and 3.4% in television watching in the household (beta= -0.184, P=5e-47).

Offspring genetic risk for schizophrenia was positively associated with paternal age, even when adjusting for education and BMI-associated polygenic variation (R^2=0.002, beta=0.049; P=1e-04). Figure 2c shows children's adjusted genetic risk for schizophrenia as a function of paternal age. Children whose father was aged over 45 at their birth had, on average, a genetic risk score for schizophrenia over one quarter standard deviation (Hedges' *g* = 0.26) higher than children whose father was under the age of 26 at their birth (t=-3.01, df=411.91; P=3e-03).

Next, we examined the extent to which associations between environmental exposures and developmental outcomes are explained by trait-associated polygenic variation for education, BMI, and schizophrenia (SI Appendix, Fig. S3). We examined associations between environmental exposures and three developmental outcomes: educational achievement, inattention-hyperactivity symptoms, and conduct problems. Of the three polygenic scores, only the education polygenic score captured covariation between environmental exposures and the three developmental outcomes (SI Appendix, Table S3).

On average education-associated polygenic variation explained 15% of the associations between the environmental measures and children's developmental outcomes. For example, the education polygenic score explained 23% (P=1.2e-18) of the beta = 0.19 covariance between child educational achievement and breastfeeding. Education-associated polygenic variation also captured 6% (P=1.9e-05) and 7% (P=4.4e-06) of the associations between parental slapping/smacking and conduct problems and hyperactivity/inattention problems (beta=0.20 for both).

Discussion

We report evidence for covariation between trait-associated polygenic variation and early environmental exposures independently of population stratification. We show that a wide range of parental, neighborhood, and parent-child perinatal characteristics, representing key early life 'environmental' influences, present at birth or early in life, correlate with offspring genetic propensity – specifically, with the allele frequency at loci associated with education, BMI, and schizophrenia. We also demonstrate that covariance between environments and important developmental outcomes are partially captured by education-associated polygenic variation.

The present study combines family and molecular data. In addition to replicating the general finding that individuals' environmental exposures vary as a function of their genotype, the current findings suggest that trait GWAS are detecting genetic variants associated with parental characteristics and their correlation with child outcomes.

Importantly, the association between exposures and outcomes was by no means entirely captured by offspring trait-associated polygenic variation. There are three likely, non-mutually exclusive, explanations for this. First, a substantial proportion of the exposure-outcome associations is likely due to non-genetic factors. Second, polygenic scores intrinsically underestimate the total genetic effects on the exposure-outcome associations because they are limited to the additive effects of common variants on a particular trait that the discovery GWAS was powered to detect. Third, we only measure offspring polygenic variation, but offspring phenotype can be influenced not only by transmitted but also by non-transmitted parental alleles via parental phenotype (i.e. child exposure).

The education-associated polygenic variation showed the strongest and most consistent correlations with environmental exposures. This is consistent with research showing associations between educational attainment and many parental behaviors and characteristics (e.g. 12, 31, 63). Moreover, the multi-polygenic score models showed that the association between BMI-associated polygenic variation and environmental exposures such as television watching and parental education are explained by education-associated genetic variations. This suggests the potential for multi-polygenic models for isolating polygenic effects, provided the underlying discovery GWAS are similarly powered. The finding of an association between paternal age and offspring genetic risk for schizophrenia is consistent with previous evidence for older fathers' elevated risk for conceiving a child who will go on to develop schizophrenia (18, 19, 63). Although the current findings provide evidence for the relevance of gene-environment correlation for polygenic trait prediction methods, they are not informative about the mechanisms involved.

The observed associations could arise from passive or active gene-environment correlation, or via environmentally-mediated genetic effects, all of which are non-mutually exclusive. Figure 3 illustrates these possibilities schematically. Many of the observed associations between offspring genotype and environment-providing parental characteristics are outside of the offspring's influence (e.g. parental age and education level at child birth) and are therefore likely to result from passive gene-environment correlation. That is, parental genetic propensities that were passed down to offspring also influence environment-providing parental behavior (through both path a and b Figure 3). However, some of the investigated parental behaviors could partially be evoked by offspring genetic propensities (through paths c, and d in Figure3; e.g. breastfeeding, watching television). Finally, genetic correlations could arise as a result of environmentally-mediated genetic effects (e.g. if education-associated genetic variation influenced mothers' predisposition to smoke during pregnancy, and prenatal exposure to nicotine had an environmental effect on

offspring attention problems, this could result in offspring education-associated polygenic variation being associated with maternal smoking pregnancy as well as capturing part of its correlation with offspring attention problems).

The design of the current study is unable to distinguish environmentally-mediated genetic effects, passive-, and evocative gene-environment correlations. One way to investigate the contributions of these different mechanisms would be to use samples incorporating parental genotype data. In analyses of such samples, confounding of offspring genotype by parental genotypes could be accounted for. Provided that paternal, maternal, and offspring genotype and phenotype data were available in a single sample, cross-generational effects of genetic and environment could be further disentangled (see Figure 3 for schematic illustration).

Nurture has a genetic component; trait-associated alleles in the offspring explain variation in environment-providing parental behaviors, and their covariation with offspring developmental outcomes. This provides evidence that the observed effects from GWAS are not only reflecting direct trait effects. This evidence resonates with the hypothesis that trait GWAS capture variation in risk factors as well as direct genetic effects on the trait (64). Here we showed that polygenic scores derived from trait GWAS predict variation in variables beyond the target trait, including variables often presumed to be environmental in origin such as parenting. This suggests incorporating genetic variants associated with environmental risk or predictive factors into polygenic prediction models might improve trait prediction.

In summary, we show that genetic variation identified by trait GWAS partially captures environmental risk or protective factors, indicating that some of the same genetic variation underlies both traits and environments. In contrast to the conceptual dichotomy often imposed between traits and environments, this finding implies that the pleiotropy widely found in phenome-genome associations also crosses over to the realm of environments and manifests across generations. Findings illustrate the relevance of gene-environment correlation for polygenic prediction models, and that combining family and molecular data might help reveal mechanisms by which genetic variation is translated into phenotypic variation.

Materials and Methods

We used genome-wide SNP and environment-wide phenotype data from 6,710 unrelated individuals drawn from the UK-representative Twins Early Development Study (57, 58),. We processed the 6,710 genotypes using stringent quality control procedures followed by imputation of SNPs to the Haplotype Reference Consortium reference panel (65) (SI Appendix, Methods S1). This included removing one individual from any pair of individuals with an estimate SNP marker relatedness >0.05. After quality control, 7,581,516 genotyped or well-imputed (info >.70) variants remained.

Polygenic scores

For each individual in the sample, we created polygenic scores for years of education, schizophrenia, and BMI. After coordinating overlapping markers between each of the three GWA summary statistics and the target data by excluding markers due to nucleotide inconsistencies or low minor allele frequency (<1%), we retained 5,690,632 for the years of education (45), 5,781,731 for schizophrenia (47), and 1,810,667 for BMI (46). We constructed polygenic scores as the effect-size weighted sums of individuals' trait-associated alleles across all SNPs. We used LDpred (60), which places a prior on the markers' effect sizes and adjusts summary statistics for linkage disequilibrium (LD) between markers. For each trait, we created score using three different priors on the fraction of causal markers, 0.01, 0.1, and 1.0, from which the one yielding the largest R^2 in the single-polygenic score models was then entered

into the multi-polygenic score model. For details on the polygenic score construction see SI Appendix, Methods S2.

To account for population stratification, we adjusted the polygenic predictors by the first 30 principal components generated from genotype data prior to the analysis. We used the top 30 PCs as well as genotyping array and plate to create a N*P matrix Z of eigenvectors across the P selected principal components. We then regressed the genetic polygenic predictor onto the eigenvectors as S = μ + Zβ + e, where μ is the mean and β is a P×1vector of the regression coefficients, and e is the residual error.

Single-score and multi-score genomic-relatedness-matrix restricted* maximum-*likelihood models

When estimating genetic effects on environmental exposures, the possibility of population stratification is especially salient. This is because genetic and common environment effects, even if uncorrelated, may be confounded as close relatives share both genes and their environment to a greater extent than other individuals. We control this type of confounding because, under only population stratification, we would not expect an association between polygenic predictors and environmental measures within the mixed effect model of equations 1 and 2. This is because they account for population stratification by both regressing PCs from the polygenic predictors (see above), and fitting a relationship matrix estimated from the SNP markers (see below).

To estimate the degree to which trait-associated polygenic variation captures variation in environmental measures, we estimated the relationship between the polygenic scores and the environmental measures, while controlling for net genetic relatedness by fitting the effects of all the SNPs as random effects by a mixed linear model.

Single-score model (Eq. 1): $var(y) = \mu + S_i\beta + \mathrm{A}\sigma_g^2 + \mathrm{I}\sigma_e^2$

Multi-score model (Eq. 1): $var(y) = \mu + S_{BMI}\beta + S_{SCZ}\beta + S_{EDU}\beta + \mathrm{A}\sigma_g^2 + \mathrm{I}\sigma_e^2$

y is an n × 1 vector containing the level of environmental exposure, with n being the sample size. β is a vector of fixed effects estimating the effects of the polygenic predictor, independently of overall genetic relatedness g.

In the single-score model (Eq. 1), S_i is a vector containing individuals' polygenic score for one of i ∈ {years of education (EDU) (45), Body Mass Index (BMI) (46), schizophrenia (SCZ) (47)}; adjusted for 30 principal components, genotyping array and plate (see section above). g is an n × 1 vector of the total genetic effects of the individuals, independently of β, with $g \sim N(0, A\sigma^2 g)$, and A is interpreted as the genetic relationship matrix (GRM) between individuals (MAF >0.01; relatedness <0.05 as described above). The genomic relationship of each pair of subjects j and k is calculated as $Ajk = 1N\sum Ni=1(xij-2pi)(xik-2pi)/2pi(1-pi)$ with x_{ij} being the number of copies of the reference allele for the i^{th} SNP of the j^{th} individual and p_i being the frequency of the reference allele (66).

In the multi-score model (Eq. 2), the effects of the three polygenic predictors are being estimated jointly, thereby allowing to the effect of each polygenic predictor independently of each other and of overall genetic relatedness g.

The genetic relatedness matrix accounts for population stratification in the environmental exposure, because it is equivalent to fitting all the principal components within the model. Equations 1 and 2 were estimated using the restricted maximum likelihood (REML) approach implemented in the *reml* function in GCTA v1.26.0 (56).

Decomposition of covariance between environmental exposures and developmental outcomes

We fit structural equation models to decompose the covariance between environmental exposures and developmental outcomes into effects of the three polygenic scores and residual covariance (SI Appendix, Fig. 3). The total covariance estimated as $Cov_{total} = (a * d) + (b * e) + (c * f) + g$ was decomposed into the effect of the education score: $Cov_{EDU} = (a * d)$, that of the BMI score: $Cov_{BMI} = (b * e)$, that of the schizophrenia score $Cov_{SCZ} = (c * f)$, and residual covariance g. We used maximum likelihood estimation with robust (Huber-White) standard errors. The analyses were conducted using the *lavaan* package in R (68).

Multiple testing correction

P-values obtained for each statistic were corrected for multiple testing using the Šidák correction (69). The Šidák adjusted alpha level is equal to1-(1-alpha)^(1/k), where k is the number of tests. The total number of tests was: 357, with 153 (3 scores * 3 priors * 17 exposures) tests for the single-polygenic score models, 51 (3 scores * 17 exposures) tests for the multi-polygenic score model, and 153 (3 scores * 17 exposures * 3 outcomes) test for the decomposition of covariance models. The multiple comparison adjustments were applied to alpha = 0.05. Hence, the corrected 'experimentwise' alpha level was 1-(1-0.05)^(1/357) = 1.44e-04.

Environmental exposures and child outcome measures

For a detailed description of all measures see the SI Appendix, Methods S3.

Acknowledgements: We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ grant agreement n° 602768 and ERC grant agreement n° 295366. RP is supported by a Medical Research Council Professorship award (G19/2). EK is supported by the MRC/IoPPN Excellence Award. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. This study presents independent research supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, NIHR, Department of Health or King's College London. We gratefully acknowledge capital equipment funding from the Maudsley Charity (Grant Ref. 980) and Guy's and St Thomas's Charity (Grant Ref. STR130505).

References

1. Danner FW (2008) A National Longitudinal Study of the Association Between Hours of TV Viewing and the Trajectory of BMI Growth Among US Children. *J Pediatr Psychol* 33(10):1100–1107.

2. Jago R, Baranowski T, Baranowski JC, Thompson D, Greaves KA (2005) BMI from 3–6 y of age is predicted by TV viewing and physical activity, not diet. *Int J Obes* 29(6):557–564.

3. Anderson CA, et al. (2010) Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychol Bull* 136(2):151–173.

4. Gentile DA, Lynch PJ, Linder JR, Walsh DA (2004) The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance. *J Adolesc* 27(1):5–22.

437 5. Räsänen P, et al. (1999) Maternal Smoking During Pregnancy and Risk of
438 Criminal Behavior Among Adult Male Offspring in the Northern Finland 1966
439 Birth Cohort. *Am J Psychiatry* 156(6):857–862.

440 6. Huizink AC, Mulder EJH (2006) Maternal smoking, drinking or cannabis use during
441 pregnancy and neurobehavioral and cognitive functioning in human offspring.
442 *Neurosci Biobehav Rev* 30(1):24–41.

443 7. White KR (1982) The relation between socioeconomic status and academic
444 achievement. *Psychol Bull* 91(3):461–481.

445 8. Sirin SR (2005) Socioeconomic Status and Academic Achievement: A Meta-
446 Analytic Review of Research. *Rev Educ Res* 75(3):417–453.

447 9. Caspi A, et al. (2016) Childhood forecasting of a small segment of the population
448 with large economic burden. *Nat Hum Behav* 1:5.

449 10. Taylor CA, Manganello JA, Lee SJ, Rice JC (2010) Mothers' Spanking of 3-Year-
450 Old Children and Subsequent Risk of Children's Aggressive Behavior.
451 *Pediatrics* 125(5):e1057–e1065.

452 11. Bender HL, et al. (2007) Use of harsh physical discipline and developmental
453 outcomes in adolescence. *Dev Psychopathol* 19(1):227–242.

454 12. Afifi TO, Mota NP, Dasiewicz P, MacMillan HL, Sareen J (2012) Physical
455 Punishment and Mental Disorders: Results From a Nationally Representative
456 US Sample. *Pediatrics* 130(2):184–192.

457 13. Knox M (2010) On Hitting Children: A Review of Corporal Punishment in the
458 United States. *J Pediatr Health Care* 24(2):103–107.

459 14. Gershoff ET (2002) Corporal punishment by parents and associated child
460 behaviors and experiences: A meta-analytic and theoretical review. *Psychol Bull*
461 128(4):539–579.

462 15. D'Onofrio BM, et al. (2014) Paternal age at childbearing and offspring psychiatric
463 and academic morbidity. *JAMA Psychiatry* 71(4):432–8.

464 16. Reichenberg A, et al. (2006) Advancing paternal age and autism. *Arch Gen*
465 *Psychiatry* 63(9):1026–32.

466 17. Sandin S, et al. (2015) Autism risk associated with parental age and with
467 increasing difference in age between the parents. *Mol Psychiatry* (April):1–8.

468 18. Malaspina D (2001) Paternal factors and schizophrenia risk: de novo mutations
469 and imprinting. *Schizophr Bull* 27(3):379–93.

470 19. Byrne M, Agerbo E, Ewald H, Eaton W, Mortensen PB (2003) Parental Age and
471 Risk of Schizophrenia. *Arch Gen Psychiatry* 60:673–678.

472 20. de Kluiver H, Buizer-Voskamp JE, Dolan CV, Boomsma DI (2016) Paternal age
473 and psychiatric disorders: A review. *Am J Med Genet B Neuropsychiatr*
474 *Genet*:n/a-n/a.

475 21. Janecka M, et al. (2017) Paternal Age Alters Social Development in Offspring. *J*
476 *Am Acad Child Adolesc Psychiatry* 0(0). doi:10.1016/j.jaac.2017.02.006.

22. Victora CG, et al. (2015) Association between breastfeeding and intelligence, educational attainment, and income at 30 years of age: a prospective birth cohort study from Brazil. *Lancet Glob Health* 3(4):e199–e205.

23. Plomin R, Bergeman CS (1991) The nature of nurture: Genetic influence on "environmental" measures. *Behav Brain Sci* 14(3):373–386.

24. Kendler KS, Baker JH (2007) Genetic influences on measures of the environment: a systematic review. *Psychol Med* 37(5):615–626.

25. Avinun R, Knafo A (2013) Parenting as a Reaction Evoked by Children's Genotype A Meta-Analysis of Children-as-Twins Studies. *Personal Soc Psychol Rev*:1088868313498308.

26. Klahr AM, Burt SA (2014) Elucidating the etiology of individual differences in parenting: A meta-analysis of behavioral genetic research. *Psychol Bull* 140(2):544–586.

27. Vinkhuyzen A a. E, Van Der Sluis S, De Geus EJC, Boomsma DI, Posthuma D (2010) Genetic influences on "environmental" factors. *Genes Brain Behav* 9(3):276–287.

28. Butcher LM, Plomin R (2008) The Nature of Nurture: A Genomewide Association Scan for Family Chaos. *Behav Genet* 38(4):361–371.

29. Larsson H, Sariaslan A, Långström N, D'Onofrio B, Lichtenstein P (2014) Family income in early childhood and subsequent attention deficit/hyperactivity disorder: a quasi-experimental study. *J Child Psychol Psychiatry* 55(5):428–435.

30. Colen CG, Ramey DM (2014) Is breast truly best? Estimating the effects of breastfeeding on long-term child health and wellbeing in the United States using sibling comparisons. *Soc Sci Med* 109:55–65.

31. D'Onofrio BM, et al. (2010) Familial Confounding of the Association Between Maternal Smoking During Pregnancy and Offspring Criminality: A Population-Based Study in Sweden. *Arch Gen Psychiatry* 67(5):529–538.

32. D'Onofrio BM, et al. (2007) Causal Inferences Regarding Prenatal Alcohol Exposure and Childhood Externalizing Problems. *Arch Gen Psychiatry* 64(11):1296–1304.

33. Lynch SK, et al. (2006) A Genetically Informed Study of the Association Between Harsh Punishment and Offspring Behavioral Problems. *J Fam Psychol JFP J Div Fam Psychol Am Psychol Assoc Div 43* 20(2):190–198.

34. Harden KP, et al. (2007) A Behavior Genetic Investigation of Adolescent Motherhood and Offspring Mental Health Problems. *J Abnorm Psychol* 116(4):667–683.

35. Narusyte J, et al. (2008) Testing different types of genotype-environment correlation: An extended children-of-twins model. *Dev Psychol* 44(6):1591–1603.

517 36. Knopik VS, et al. (2006) Maternal alcohol use disorder and offspring ADHD:
518 disentangling genetic and environmental effects using a children-of-twins
519 design. *Psychol Med* 36(10):1461–1471.

520 37. Silberg JL, Maes H, Eaves LJ (2010) Genetic and environmental influences on
521 the transmission of parental depression to children's depression and conduct
522 disturbance: an extended Children of Twins study. *J Child Psychol Psychiatry*
523 51(6):734–744.

524 38. Krapohl E, Plomin R (2016) Genetic link between family socioeconomic status
525 and children's educational achievement estimated from genome-wide SNPs.
526 *Mol Psychiatry* 21(3):437–443.

527 39. Davies NM, Hemani G, Timpson NJ, Windmeijer F, Davey Smith G (2015) The
528 role of common genetic variation in educational attainment and income:
529 evidence from the National Child Development Study. *Sci Rep* 5.
530 doi:10.1038/srep16509.

531 40. Trzaskowski M, et al. (2014) Genetic influence on family socioeconomic status
532 and children's intelligence. *Intelligence* 42:83–88.

533 41. Benjamin DJ, et al. (2012) The genetic architecture of economic and political
534 preferences. *Proc Natl Acad Sci* 109(21):8026–8031.

535 42. Marioni RE, et al. (2014) Molecular genetic contributions to socioeconomic status
536 and intelligence. *Intelligence* 44:26–32.

537 43. Power RA, et al. (2013) Estimating the heritability of reporting stressful life events
538 captured by common genetic variants. *Psychol Med* 43(9):1965–1971.

539 44. Hill WD, et al. (2016) Molecular Genetic Contributions to Social Deprivation and
540 Household Income in UK Biobank. *Curr Biol* 0(0).
541 doi:10.1016/j.cub.2016.09.035.

542 45. Okbay A, et al. (2016) Genome-wide association study identifies 74 loci
543 associated with educational attainment. *Nature* 533(7604):539–542.

544 46. Locke AE, et al. (2015) Genetic studies of body mass index yield new insights for
545 obesity biology. *Nature* 518(7538):197–206.

546 47. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014)
547 Biological insights from 108 schizophrenia-associated genetic loci. *Nature*
548 511(7510):421–427.

549 48. Tiihonen J, et al. (2009) 11-year follow-up of mortality in patients with
550 schizophrenia: a population-based cohort study (FIN11 study). *The Lancet*
551 374(9690):620–627.

552 49. Wang H, et al. (2016) Global, regional, and national life expectancy, all-cause
553 mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a
554 systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*
555 388(10053):1459–1544.

556 50. Brown S (1997) Excess mortality of schizophrenia. A meta-analysis. *Br J*
557 *Psychiatry* 171(6):502–508.

558 51. Berrington de Gonzalez A, et al. (2010) Body-mass index and mortality among
559 1.46 million white adults. *N Engl J Med* 363(23):2211–2219.

560 52. Collaboration PS (2009) Body-mass index and cause-specific mortality in
561 900 000 adults: collaborative analyses of 57 prospective studies. *The Lancet*
562 373(9669):1083–1096.

563 53. OECD (2013) *Education at a Glance 2013* (Organisation for Economic Co-
564 operation and Development, Paris) Available at: http://www.oecd-
565 ilibrary.org/content/book/eag_highlights-2013-en [Accessed December 10,
566 2013].

567 54. Morris JN, Blane DB, White IR (1996) Levels of mortality, education, and social
568 conditions in the 107 local education authority areas of England. *J Epidemiol*
569 *Community Health* 50(1):15–17.

570 55. Huisman M, et al. (2005) Educational inequalities in cause-specific mortality in
571 middle-aged and older men and women in eight western European populations.
572 *The Lancet* 365(9458):493–500.

573 56. Zheng J, et al. (2017) LD Hub: a centralized database and web interface to
574 perform LD score regression that maximizes the potential of summary level
575 GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*
576 33(2):272–279.

577 57. Oliver BR, Plomin R (2007) Twins' Early Development Study (TEDS): A
578 multivariate, longitudinal genetic investigation of language, cognition and
579 behavior problems from childhood through adolescence. *Twin Res Hum Genet*.

580 58. Haworth CMA, Davis OSP, Plomin R (2013) Twins Early Development Study
581 (TEDS): A genetically sensitive investigation of cognitive and behavioral
582 development from childhood to young adulthood. *Twin Res Hum Genet* 16:117–
583 125.

584 59. Kovas Y, Haworth CMA, Dale PS, Plomin R (2007) The genetic and
585 environmental origins of learning abilities and disabilities in the early school
586 years. *Monogr Soc Res Child Dev* 72(3):vii, 1-144.

587 60. Vilhjálmsson BJ, et al. (2015) Modeling linkage disequilibrium increases accuracy
588 of polygenic risk scores. *Am J Hum Genet* 97(4):576–592.

589 61. Visscher PM, Yang J (2016) A plethora of pleiotropy across complex traits. *Nat*
590 *Genet* 48(7):707–708.

591 62. Johnson W, et al. (2011) Does Education Confer a Culture of Healthy Behavior?
592 Smoking and Drinking Patterns in Danish Twins. *Am J Epidemiol* 173(1):55–63.

593 63. Janecka M, Mill J, Basson MA (2017) Advanced paternal age effects in
594 neurodevelopmental disorders — review of potential underlying mechanisms.
595 *Transl Psychiatry* (e1019).

596 64. Gage SH, Smith GD, Ware JJ, Flint J, Munafò MR (2016) G = E: What GWAS
597 Can Tell Us about the Environment. *PLOS Genet* 12(2):e1005765.

598 65. McCarthy S, et al. (2015) A reference panel of 64,976 haplotypes for genotype
599 imputation. *bioRxiv*:35170.

600 66. Yang J, et al. (2010) Common SNPs explain a large proportion of the heritability
601 for human height. *Nat Genet* 42(7):565–569.

602 67. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-
603 wide Complex Trait Analysis. *Am J Hum Genet* 88(1):76–82.

604 68. Rosseel Y (2012) lavaan: An R Package for Structural Equation Modeling. *J Stat*
605 *Softw* 48(2):1–36.

606 69. Sidak Z (1971) On Probabilities of Rectangles in Multivariate Student
607 Distributions: Their Dependence on Correlations. *Ann Math Stat* 42(1):169–175.

608 70. Richmond RC, et al. (2017) Using Genetic Variation to Explore the Causal Effect
609 of Maternal Pregnancy Adiposity on Future Offspring Adiposity: A Mendelian
610 Randomisation Study. *PLOS Med* 14(1):e1002221.

611 71. Zhang G, et al. (2015) Assessing the Causal Relationship of Maternal Height on
612 Birth Size and Gestational Age at Birth: A Mendelian Randomization Analysis.
613 *PLOS Med* 12(8):e1001865.

614 72. Eaves LJ, Pourcain BS, Smith GD, York TP, Evans DM (2014) Resolving the
615 Effects of Maternal and Offspring Genotype on Dyadic Outcomes in Genome
616 Wide Complex Trait Analysis (“M-GCTA”). *Behav Genet*:1–11.

617

Figure Legends

Figure 1

a *Single-polygenic score models: Associations between polygenic scores and environmental exposures*
Single-predictor effects of polygenic scores for years of education, BMI, and schizophrenia on the environmental exposures.
b *Multi-polygenic score models: Joint estimation of effects of polygenic scores on environmental exposures*
Effects of polygenic scores for years of education, BMI, and schizophrenia on the environmental exposures while adjusting for other predictors, respectively.
Color gradients represent effect sizes as standardized coefficients, i.e. standard deviations change in the environmental exposure, per standard deviation increase in the polygenic predictor, while adjusting for the other polygenic predictors in the model, respectively (see SI Appendix, Tables S1-3 for full statistics). Single asterisk indicates uncorrected P < 0.05, double asterisks indicate multiple testing corrected P < 0.05 (see Materials & Methods).

Figure 2

a *Offspring adjusted education polygenic score (standardized) by level of breastfeeding:* Education polygenic score was adjusted for schizophrenia and BMI polygenic scores. Positive association (R^2=0.021, beta=0.144; P=7e-30). Children who were breastfed had, on average, an education polygenic score approximately one third standard deviation higher (Hedges' *g* = 0.30) than children who were not breastfed (t =-11.55, df= 5664.2, P=1.6e-30).
b *Offspring adjusted education polygenic score (standardized) by level of maternal smoking during pregnancy:* Education polygenic score was adjusted for schizophrenia and BMI polygenic scores. Negative association (R^2=0.008, beta=0.090; P=5e-13). Children exposed to maternal smoking prenatally had, on average, an education polygenic score approximately one quarter standard deviation lower (Hedges' *g* = 0.26) than children whose mothers did not smoke (t =7.93, df=1556.3; P=4e-15).
c *Offspring adjusted schizophrenia polygenic score (standardized) by paternal age at birth of offspring:* Genetic risk for schizophrenia was adjusted for education and BMI polygenic scores. Positive association (R^2=0.002, beta=0.049; P=1e-04). Children whose father was aged over 45 at their birth had, on average, a genetic risk score for schizophrenia over one quarter standard deviation (Hedges' *g* = 0.26) higher than children whose father was under the age of 26 at their birth (t=-3.01, df=411.91; P=3e-03).
Horizontal lines and bars represent means and 95% confidence intervals. Violin shapes represent probability density of the data.

Figure 3

Schematic illustration of cross-generational effects within family triad
Because of the lack of parental genotype data, the present study was unable to distinguish passive and evocative gene-environment correlation.
Passive gene-environment correlation: $a_{m,p}*b_{m,p}$.
Evocative gene-environment correlation: $c_{m,p}*b_{m,p}$
Offspring phenotype can be influenced by both the transmitted paternal and maternal alleles (red arrows), and by non-transmitted alleles via parental phenotype (green arrows). Provided that paternal, maternal, and offspring genotype and phenotype data were available in a single sample, the effect of parental trait-associated alleles on offspring phenotype independently of genetic sharing between parents and offspring (green arrows) could be estimated (70–72). A testable assumption for

672 investigating these mechanisms is there is no correlation between parental
673 genotypes and between each parent's haplotypes (i.e. assortative mating) (yellow
674 arrows).