Citation for published version (APA):
The nature of nurture: Widespread covariation of early environmental exposures and trait-associated polygenic variation

E Krapohl¹; L J Hannigan¹; J-B Pingault²; H Patel¹,³; C Curtis¹,³; S Newhouse¹,³; T C Eley¹; P F O'Reilly¹; R Plomin¹*

¹MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
²Clinical, Educational and Health Psychology, University College London, London, UK
³NIHR 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.


\textbf{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
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; $P=4e-15$).

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, $P=1.6e-30$).

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; $P=4e-15$).

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 higher (Hedges’ $g = 0.26$) 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 R² 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×1 vector 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): \( \text{var}(y) = \mu + S_i \beta + A \sigma_g^2 + 1 \sigma_e^2 \)

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

\( y \) is an \( n \times 1 \) vector containing the level of environmental exposure, with \( n \) being the
sample size. \( \beta \) 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 \( i \in \{ \text{years of education (EDU)} \ (45), \text{Body Mass Index (BMI)} \ (46), \text{schizophrenia (SCZ)} \ (47) \} \); adjusted for 30 principal components, genotyping array
and plate (see section above). \( g \) is an \( n \times 1 \) vector of the total genetic effects of the
individuals, independently of \( \beta \), with \( g \sim N(0, A \sigma_g^2 g) \), 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 \( \text{A}_{jk} = 1 \sum N_i = 1(x_i - 2p_i)(x_k - 2p_i)/2p_i(1 - p_i) \) with \( x_i \) 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 \textit{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 \( \text{Cov}_{\text{total}} = (a \cdot d) + (b \cdot e) + (c \cdot f) + g \) was decomposed into the effect of the education score: \( \text{Cov}_{\text{EDU}} = (a \cdot d) \), that of the BMI score: \( \text{Cov}_{\text{BMI}} = (b \cdot e) \), that of the schizophrenia score \( \text{Cov}_{\text{SZ}} = (c \cdot f) \), and residual covariance \( g \). We used maximum likelihood estimation with robust (Huber-White) standard errors. The analyses were conducted using the \textit{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 to \( 1 - (1 - \alpha)^{1/k} \), 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 \( 1 - (1 - 0.05)^{1/357} = 1.44 \times 10^{-4} \).

**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**


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. 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).

**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; $P=7\times10^{-30}$). Children who were breastfed had, on average, an education polygenic score approximately one third standard deviation higher ($\text{Hedges' } g=0.30$) than children who were not breastfed ($t=-11.55$, df=5664.2, $P=1.6\times10^{-30}$).

**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; $P=5\times10^{-13}$). Children exposed to maternal smoking prenatally had, on average, an education polygenic score approximately one quarter standard deviation lower ($\text{Hedges' } g=0.26$) than children whose mothers did not smoke ($t=7.93$, df=1556.3; $P=4\times10^{-15}$).

**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; $P=1\times10^{-04}$). Children whose father was aged over 45 at their birth had, on average, a genetic risk score for schizophrenia over one quarter standard deviation ($\text{Hedges' } g=0.26$) higher than children whose father was under the age of 26 at their birth ($t=-3.01$, df=411.91; $P=3\times10^{-03}$).

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