Supplementary Online Materials

The role of parental genotype in predicting offspring years of education: Evidence for genetic nurture

Emily A. Willoughby, Matt McGue, William G. Iacono,
Aldo Rustichini, James J. Lee

Contents

Here we supplement the main text with additional methodological details, results of secondary analyses, and conceptual discussion.

1 Calculation of the polygenic scores (PGS)

A polygenic score (PGS), often called the polygenic risk score in disease prediction, is calculated from a set of SNPs that are tested in the initial GWAS sample for association with a trait of interest. Effect sizes are estimated for each marker to construct a PGS in a replication sample. The basic definition of the PGS is

$$\hat{S} = \sum_{j=1}^{m} \hat{\beta}_j G_j,$$

where $\hat{S}$ is the individual's polygenic score, $\hat{\beta}_j$ is the weight of SNP j as derived from the GWAS training sample, and G_j is the individual's count of the reference allele at SNP j.

Polygenic scores for years of education (*EduYears*) in the current study were derived from downloadable EA3 summary statistics using the LDpred software package, which uses the correlations between SNPs estimated in an external reference panel to convert the univariate regression coefficients making up the GWAS summary statistics to what are effectively partial regression coefficients [1]. We used the MCTFR white parents as the reference panel. Following the EA3 authors [2], we set the LDpred shrinkage parameter equal to unity—the highest possible value and the one leading to the least shrinkage of the PGS weights. The developers recommend trying a grid of values and choosing the one leading to the best prediction accuracy in the validation sample, but we eschewed this optimization in order to avoid any possibility of "double dipping." Because the MCTFR was part of EA3, data from the target MCTFR sample were removed from the SSGAC GWAS before weights were derived.

2 Analysis details

Data were analyzed and figures were created in `R` [3–6]. Only those individuals with data indicating white European ancestry were included in analyses. Participants with alternate or missing ethnicity data were excluded; a PGS declines in accuracy when used in a population different from the one studied in the training sample. All polygenic scores were standardized prior to data analysis. In cases where β coefficients were compared for different phenotypes, these coefficients were standardized by multiplying the slope by the standard deviation of the predictor over the standard deviation of the outcome variable. In all cases where two regression models were compared and assessed for significant differences, all missing data were removed for each variable present in the models, ensuring that while the compared samples were sometimes smaller as a result, the models were necessarily fitted to the same datasets for these comparisons.

Bootstrap resampling over twin pairs was used to compute standard errors and P values for offspring predictions in order to ensure that the statistical inferences are not affected by the non-independence of twins from the same family. Bootstrapping was computed over 100 iterations.

3 The correlation between midparent and offspring PGS

As a quality-control check on our genotyping and polygenic scoring, we calculated the correlation between the midparent and offspring PGS. The value of this correlation in the population is theoretically predicted to be somewhat greater than $\sqrt{1/2}$; it will be exactly this value in the absence of assortative mating. A derivation of this fact can be found in elementary texts [e.g., 7]. We found a sample correlation of 0.73 (Table 1), in very good accord with the theoretical prediction.

4 An apparent effect of genetic nurture cannot be explained by unreliability of the offspring PGS

Fig. 1 shows that in the absence of an environmentally mediated effect of the parent PGS on the offspring phenotype, the offspring PGS d-separates the parent PGS from the offspring phenotype [8, 9]; consequently, in a regression model predicting the offspring phenotype with the PGS of both parent and offspring, only the offspring PGS should have predictive power. It has been suggested to us that this reasoning may be flawed if the PGS is noisy as a result of sampling error in the GWAS, because in this case the parent PGS might compensate for the unreliability of the offspring PGS. This suggestion seems to be inspired by standard psychometric theory, in which the predictive validity of a composite measuring a single common factor increases with the number of measurements contributing to the composite [10].

This suggestion is in fact incorrect, as we now demonstrate.

4.1 Negative controls with unreliable PGS do not show the behavior of *EduYears*

Perhaps the simplest argument against the unreliability of the offspring PGS as an explanation of our findings is that we did not observe a comparable pattern in our negative-control phenotypes (Fig. 3).

4.2 Genotyping/imputation are highly accurate

To the extent that the offspring's SNP genotypes are called incorrectly as a result of genotyping/imputation error, the use of the parents' SNP genotypes may indeed increase

the prediction R^2. Genotyping/imputation is typically quite accurate, however, and the closeness of our sample correlation between midparent and offspring PGS to its theoretical value (see the previous section) confirms this accuracy in our own case [11]. Consequently, any improvement in the prediction R^2 for this reason should be quite small. We henceforth do not treat this possibility further.

4.3 Simulation of training GWAS, polygenic scoring, and out-of-sample validation does not show incremental predictive value of parent PGS

We conducted simulations to test the soundness of our argument that the weights in the parent PGS are distorted by exactly the same sampling errors and so cannot lead to an improvement in the prediction of the offspring phenotype. In simple simulations based on a genome where all pairs of causal sites are unlinked and in linkage equilibrium, we found conclusively that the addition of the parent PGS does nothing to ameliorate GWAS sampling error (results not shown). To examine the more realistic case of linkage and LD, we conducted more complex simulations based on the MCTFR genetic data.

To minimize computational burden, we only used chromosome 1. We randomly selected 1,520 SNPs present in the intersection of our MCTFR imputed dataset and the 1000 Genomes Phase 3 European reference panel to be the causal sites affecting the simulated phenotype; this density is consistent with the total number of SNPs affecting EduYears, as estimated in a recent paper [12]. To ensure the clear predictive validity of the PGS, we gave these SNPs a normal distribution of per-standardized-allele effects consistent with a contribution to heritability equaling 0.15. We postmultiplied the correlation matrix of the SNPs (estimated in 1000 Genomes) by the vector of per-standardized-allele effects (a sparse vector with 1,520 nonzero entries) to obtain the vector of per-standardized-allele univariate regression coefficients. In more detail, we assumed the negligibility of any correlation between SNPs in different LD blocks defined by the method of Berisa and Pickrell [13] and concatenated the results of postmultiplying the correlation matrix of an individual block with the effect vector of just the SNPs in that block. To simulate GWAS sampling error, we took a random draw from the multivariate normal distribution with a mean vector of zero and a covariance matrix equal to the LD correlation matrix scaled by the reciprocal of the simulated GWAS sample size. We then added this random draw to the vector of per-standardized-allele univariate regression coefficients to obtain a vector of simulated GWAS summary statistics. Note that this method is similar to the second method employed by O'Connor and Price (Supplementary Note, pp. 22–23) [14]. We examined three different GWAS sample sizes: 30 thousand, 100 thousand, and 300 thousand. For each distinct size of the GWAS training sample, we resampled the identities and effect sizes of the causal SNPs.

To generate simulated phenotypes, we used GCTA [15] to take the dot product of each MCTFR individual's imputed genotypes and the causal effects. For each distinct GWAS sample size, we simulated 200 replicates. Note that the replicates varied the non-genetic

residuals of the individuals but retained the same causal SNPs and their effect sizes.

We combined the simulated summary statistics with an LD reference panel consisting of the MCTFR parents to obtained LDpred PGS weights for all ∼39,000 genotyped SNPs on chromosome 1 in the MCTFR dataset. (In our real data analysis, we used a PGS based on genotyped SNPs because imputed data in the MCTFR dataset are not available for non-whites, whom we plan to study in other projects. Even though the genotyped SNPs are a subset of all common SNPs, the predictive power of a PGS based on genotyped SNPs is still close to maximal because the genotyped SNPs tag nearly all common genetic variation [16, 17].) We set the prior shrinkage parameter to unity—as in our real data analysis, although a different value might have led to a higher prediction R^2. We then used PLINK 1.9 [18] to calculate each individual's PGS on the basis of these weights.

In the validation stage of the simulation, we estimated the coefficients of two regression models. The first model used only the offspring's own PGS as a predictor of the offspring phenotype. The second model used the PGS of the offspring, mother, and father. A significant increment in the prediction R^2 as a result of adding the mother and father PGS would show that the noisiness of the offspring PGS can indeed be ameliorated by the use of the parent PGS as additional predictors.

The results of our simulation are summarized in Supplementary Fig. S1. There are some anomalies, but overall the results demonstrate the implausibility of GWAS sampling error as an explanation of what appears to be genetic nurture.

- As expected, the baseline prediction R^2 increased with each (approximate) tripling of GWAS sample size (4.0, 6.4, and 7.5 percent of the phenotypic variance respectively).

- At the smallest simulated sample size (30,000), the increment in the prediction R^2 (adjusted for number of predictors) from the addition of the parent PGS was statistically significant ($P < 10^{-7}$). We do not have an explanation of this result. Regardless, the unadjusted increment in the prediction R^2 was quantitatively quite small, amounting to an additional 0.3 percent of the phenotypic variance. Recall that in our real data analysis of *EduYears*, the addition of the midparent PGS led to an increment of 1.8 percent.

- At the intermediate sample size (100,000), the increment in the prediction R^2 (adjusted for number of predictors) did not reach statistical significance ($P = 0.75$) and was quantitatively small (0.1 percent).

- At the largest sample size (300,000), the increment in the prediction R^2 (adjusted for number of predictors) did reach nominal statistical significance ($P = 0.04$) but was actually negative; without adjustment for number of predictors, the increment was a positive but minuscule 0.03 percent.

The simulation results thus fail to support the notion that, in the absence of genetic

nurture, the use of the parent PGS as covariates can substantially improve the prediction R^2.

5 An apparent effect of genetic nurture in our data is unlikely to be the result of population stratification

In the first submitted version of this manuscript, we employed the GWAS summary statistics of the GIANT Consortium [19, 20] to construct our height and BMI PGS. In all of our initial analyses, regardless of the phenotype, we set the LDpred shrinkage parameter to 0.3, as we found this value to produce the maximal validation R^2 for *EduYears*. In response to a reviewer's suggestion to refrain from optimization of the shrinkage parameter, we set it to unity for the revised manuscript and thereby found that the incremental predictive power of the parent PGS increased significantly and substantially in the case of height. In the course of investigating this anomaly, we found that the distribution of the height PGS (with the shrinkage parameter equal to unity) was extremely right-skewed with a substantial second mode, quite unlike that of the normally distributed *EduYears* PGS. Because of the central limit theorem, a skewed multimodal distribution is extremely unlikely in the absence of population stratification. That is, our sample probably represents at least two cryptic subpopulations with very different averages of the height PGS.

Two groups have recently shown that the height GWAS on which we initially relied [19] was subject to a form of population stratification such that alleles apparently associated with height have higher frequencies in Northern Europe [21, 22]. Suppose, for the sake of argument (but not unrealistically), that our sample of Minnesota families contains a subset with unusually high Scandinavian ancestry [23, 24]. If compared to other individuals with the same values of the height PGS, the Scandinavians will be shorter because their height PGS reflects to some extent their ancestry rather their genetic potential for height. But since the parent PGS provides additional information about ancestry, the parents of the Scandinavians will have a higher PGS than the parents of the non-Scandinavians, and thus the parent PGS can provide a correction of the overprediction. That is, the partial regression coefficients of offspring and parent PGS will have opposite sign—specifically, the parent PGS having a negative sign—because a high parent PGS indicates Scandinavian ancestry. This was exactly the borderline-significant pattern observed in our first submission, and the pattern became stronger upon switching the LDpred shrinkage parameter from 0.3 to 1 (perhaps because more aggressive shrinkage somewhat purifies the PGS of population stratification in the original GWAS).

For this reason we decided to use different GWAS as the data sources for our anthropometric PGS. Specifically, we used the UK Biobank summary statistics provided by Loh et al. [25]. Besides offering the advantage of larger sample size, these GWAS used a more homogenous population and applied to their single cohort a linear mixed model to deal with residual confounding. It has been shown that these height summary statistics are

largely free from the population stratification present in the GIANT Consortium summary statistics that we initially employed [19]. And indeed, when the PGS derived from these statistics was used to predict height in our MCTFR sample, we no longer observed a statistically significant partial regression coefficient of parent PGS in the case of height (Fig. 3; $p = .67$). We also did not observe a statistically significant partial regression coefficient of parent PGS in the case of BMI ($p = .55$).

The question then arises as to whether population stratification can explain the significant incremental predictive power of the parent PGS in the case of *EduYears*. We think this is extremely unlikely to be the case. First, the GIANT Consortium's GWAS of height shows a mean chi-square statistic of 2.92 and a rather large LD Score regression intercept of 1.28 [21, Supplementary Table 5]. Despite a larger mean chi-square statistic of HapMap3 SNPs in EA3 (3.81), the LD Score regression intercept in that GWAS is only 1.11 [2]. Furthermore, to construct our PGS, we used a version of the EA3 summary statistics without the contribution of 23andMe, which would tend to reduce whatever stratification is present in the full meta-analysis, because a larger proportion comes from the homogeneous UK Biobank cohort and none at all from the likely heterogeneous 23andMe sample. Second, we have applied some of the analyses in the Sohail and Berg papers to the downloadable EA3 summary statistics and found no evidence of severe population stratification (results not shown). For example, when applying the tSDS analysis to EA3, the results look extremely similar to those in the UKB panel of Sohail et al.'s Fig. 3a and not at all like those in the GIANT panel.

In summary, we discontinued the use of the summary statistics from the GIANT Consortium to construct our height and BMI PGS and switched to the UK Biobank summary statistics provided by Loh et al. The latter summary statistics are based on a larger and more homogenous sample. The resulting PGS show absolutely no evidence of parent genetics providing incremental predictive power.

6 Within-family prediction and incremental effect of parent PGS as complementary strategies

Because the operation of genetic nurture implies both a within-family effect smaller than implied by the population GWAS results and a significant partial regression coefficient of the parent PGS, it might seem that the demonstration of both consequences is redundant. Here we point out why these two items of evidence are not redundant but rather complementary.

A within-family effect smaller than implied by the population GWAS follows from an environmentally mediated effect of parent genotype, but the converse is not true. That is, there are alternative explanations of a smaller within-family effect. One such explanation is assortative mating. In one of many striking passages in *The Genetical Theory of Natural Selection*, R.A. Fisher wrote

> [i]n human stature, for example, the correlation found between married persons is sufficient to ensure that each gene tending to increase the stature must be associated with other genes having a like effect, to an extent sufficient to make the average excess [i.e., univariate regression coefficient] associated with each gene substitution exceed its average effect [i.e., within-family effect] by about a quarter. [26, p. 31]

The possibility that assortative mating explains the *EduYears* discrepancy has been examined extensively [2]. Since assortative mating was not found to be a plausible explanation for the whole of the discrepancy, the EA3 authors suggested the mechanism of genetic nurture[1]. The approach of EA3 does rely on the critical assumption that the phenotypic correlation between mates is the result of a preference for mates with either higher or similar values of *EduYears*, which leads in turn to the phenotypic correlation being larger than the correlation between the true polygenic scores. There are some hints in the literature that this assumption may be incorrect—that, perhaps because *EduYears* is merely a downstream effect of some other trait affecting assortment, spousal genetic similarity actually exceeds similarity with respect to phenotypic *EduYears* [e.g., 29].

The upshot of all this is that the conditional significance of the parent PGS as a predictor of offspring phenotype is by no means a redundant piece of evidence in the prosecution of parent genotype for affecting *EduYears* through an environmental mechanism.

Supplementary Fig. S2 compares the within-family effects of the PGS on our four outcomes to the between-family coefficients (i.e., the coefficient in the regression of the twinship mean outcome on the twinship mean PGS). The between-family coefficients in Supplementary Fig. S2 are larger than the individual-level coefficients in Fig. 2 of the main text, but this is to be expected. If the individual-level coefficient is not equal to the within-family coefficient, then the individual-level coefficient is not necessarily equal to the between-family coefficient either [**selzam2019**]. In any case, the qualitative interpretation of Supplementary Fig. S2 is the same: the between-family coefficient for *EduYears* is larger than its corresponding within-family coefficient, pointing to some combination of assortative mating and genetic nurture.

7 Magnitude of genetic nurture as estimated in different studies

Previous studies of genetic nurture have used estimation procedures slightly different from ours. We now comment on the relationship between these procedures so as to determine the extent to which our estimate of genetic nurture is consistent with those obtained by previous studies.

[1] Interestingly the EA3 authors failed to find a discrepancy in the case of height of the magnitude predicted by Fisher, although later research has found evidence of assortative mating for height by other means [27]. The shortfall may possibly be attributable to natural selection [28].

The first published study of genetic nurture, by Kong et al. [30], used families of genotyped parent-offspring trios. They partitioned each parent's PGS into a portion that was transmitted to the offspring and a portion that was not so transmitted. Note that the offspring's own PGS is simply equal to the transmitted portions of both parents.

In a multiple regression of the offspring's phenotype on both transmitted and nontransmitted PGS, let θ_{T} be the coefficient of the transmitted portion and θ_{NT} the coefficient of the nontransmitted portion. Kong et al. took $\delta := \theta_{\mathrm{T}} - \theta_{\mathrm{NT}}$ to be the causal effect of the offspring PGS on the offspring phenotype, free from inflation ascribable to genetic nurture. To see the justification of this, first let γ denote the effect of the parent PGS operating through genetic nurture. Then

$$\begin{aligned}
&(\text{offspring PGS}_{\mathrm{T}})\delta + (\text{parent PGS}_{\mathrm{T}} + \text{parent PGS}_{\mathrm{NT}})\gamma \\
&= (\text{offspring PGS}_{\mathrm{T}})\delta + (\text{parent PGS}_{\mathrm{T}})\gamma + (\text{parent PGS}_{\mathrm{NT}})\gamma \\
&= \text{PGS}_{\mathrm{T}}(\delta + \gamma) + (\text{parent PGS}_{\mathrm{NT}})\gamma.
\end{aligned}$$

It thus follows that θ_{T} is equal to $\delta + \gamma$ and θ_{NT} to γ alone. The difference $\theta_{\mathrm{T}} - \theta_{\mathrm{NT}}$ is indeed the direct causal effect of the PGS on the phenotype. Kong et al. then took $\delta/\theta_{\mathrm{T}}$ as the percentage of the PGS regression coefficient in the prediction of the phenotype attributable to the direct causal effect. The denominator, θ_{T}, will not be exactly equal to the univariate PGS regression coefficient because of a slight correlation between PGS_{T} and PGS_{NT} induced by assortative mating, but the two quantities will be quite close.

In our own work, we first regressed the offspring phenotype on just the offspring's own PGS and subsequently on both the offspring and midparent PGS. We used the estimates obtained in this way (Table 3) to compute a quantity equivalent to $\delta/\theta_{\mathrm{T}}$ as follows. We placed in the numerator the coefficient of offspring PGS in the regression model predicting the offspring phenotype with both offspring and midparent PGS, as this is also an estimate of the direct causal effect. We placed in the denominator the sum of the numerator and half the coefficient of midparent PGS; it is evident from the first line of the equation above that the coefficient of midparent PGS is equal to 2γ, and thus our denominator is equivalent to Kong et al.'s θ_T.

Supplementary Table S2 shows our estimated ratio as well as the others published to date. Our estimate of 0.613 is the one that assigns the largest role to genetic nurture. Standard errors are not available for any of these estimates; if these were to be calculated, statistically significant heterogeneity would possibly be found. Nevertheless, even after we allow for some heterogeneity, all estimates are consistent with at least 60 percent of the correlation between the PGS and *EduYears* being attributable to the direct causal effect of the offspring's own PGS on their own *EduYears*.

8 Effects of genotypes belonging to ancestors more remote than the parents

The relationship between the midparent PGS and whatever parental traits may be affecting the offspring *EduYears* may not be purely causal but rather itself reflect genetic nurture to some extent. That is, the parents themselves may have received both genotypes and trait-affecting environments from the uncontacted grandparents in our pedigrees. Here, we explore this possibility in greater depth and argue that the majority of the confounding due to genetic nurture is attributable to the parental generation rather than the grandparental and more remote generations. Our discussion below abuses notation in that symbols used previous ly may now have different meanings.

The direct causal effect of offspring genotype on offspring trait can be represented by

$$\text{offspring PGS} \xrightarrow{\alpha} \text{offspring trait},$$

whereas parental confounding can be represented by

$$\text{offspring PGS} \xleftarrow{1} \text{parent PGS} \xleftarrow{\alpha} \text{parent trait} \xrightarrow{\delta} \text{offspring trait}.$$

If genetic nurture perpetuates itself across generations, then we have grandparental confounding that can be represented by

$$\text{offspring PGS} \xleftarrow{1} \text{parent PGS} \xleftarrow{1} \text{grandparent PGS} \xrightarrow{\alpha}$$
$$\text{grandparent trait} \xrightarrow{\delta} \text{parent trait} \xrightarrow{\delta} \text{offspring trait}.$$

In this path-diagrammatic representation, we have assumed invariance of the causal background across generations. That is, we have assumed that the effect of genotype on phenotype (α) and the direct effect of parent on offspring trait (δ) is the same from one generation to the next. We have also assumed that there is no direct effect of grandparent trait on offspring trait. Some recent evidence favors this assumption [31].

Although we are assuming that the parent trait that affects the offspring trait is the same as the offspring trait itself (i.e., *EduYears*), our argument still follows if the parent and offspring trait show a strong genetic correlation and comparable heritabilities. These conditions in fact hold if the parent trait is indeed SES, a composite including *EduYears* and income [32]. Furthermore, including parent *EduYears* by itself as a covariate does eliminate the statistical significance of parent PGS as a predictor (Supplementary Table S1).

It is well known that the slope in the regression of offspring PGS on midparent PGS is equal to unity [e.g., 7]. Calling the parental generation the first generation, the grandparental generation the second, and so on, we now prove that the slope in the regression of the average PGS of one's 2^k ancestors in generation k on the average PGS of one's 2^{k+1}

ancestors in generation $k+1$ is always unity.

$$\begin{aligned}
\text{PGS}_k &:= \frac{1}{2^k}\sum_{i=1}^{2^k}\text{PGS of ancestor } i \text{ in generation } k \\
&= \frac{1}{2^k}\sum_{i=1}^{2^k}\frac{1}{2}\left(\text{PGS of ancestor } i\text{'s mother} + \text{PGS of ancestor } i\text{'s father}\right) \\
&= \frac{1}{2^{k+1}}\sum_{j=1}^{2^{k+1}}\text{PGS of ancestor } j \text{ in generation } k+1 \\
&:= \text{PGS}_{k+1},
\end{aligned}$$

where we ignore orthogonal zero-mean terms representing Mendelian segregation. Since we have been able to go from the top to bottom line without the introduction of any terms or factors, it follows that regression of PGS_k on PGS_{k+1} indeed has a slope of unity.

An analogous argument shows that if the direct effect of the midparent phenotype on offspring phenotype is δ, then the effect of the average ancestor's phenotype in generation $k+1$ on that in k is also δ.

By the path-tracing rules, the direct effect of offspring PGS on the offspring trait makes the contribution α to the total coefficient of the offspring PGS, the confounding in the parental generation makes the contribution $\alpha\delta$, and the confounding in the grandparental generation makes the contribution $\alpha\delta^2$. In general, the contribution of confounding in the kth previous generation is $\alpha\delta^k$. If the gene-environment correlations extend indefinitely into the past, then the total coefficient of the offspring PGS is the infinite geometric series

$$\sum_{k=0}^{\infty}\alpha\delta^k = \frac{\alpha}{1-\delta}.$$

Suppose that genetic nurture leads to an inflation of the PGS direct effect α by about a third (Supplementary Table S2). Setting the geometric series equal to $4\alpha/3$ and solving for δ, we get the solution $1/4$. Thus, $(1/4)/(1/3) = 3/4$ of the inflation attributable to genetic nurture is the result of gene-environment correlation in the parental generation and only $1/4$ to the grandparental and more remote generations. This conclusion is not particularly sensitive to the postulated size of the total inflation due to genetic nurture. If we set the geometric series equal to $5\alpha/3$, consistent with a nearly 70-percent inflation, then 60 percent of the total inflation is attributable to the parental generation.

A remaining issue is whether a direct effect of parent trait on offspring trait implied by our calculation above ($\sim$1/4) is plausible. One adoption study with plausibly random assignment of adoptees to parents has found that the partial regression coefficient of rearing mother's *EduYears* in the prediction of adopted offspring's *EduYears* is 0.097 with a

standard error of 0.027; the covariates in this model included the logarithm of parents' income, although the latter was not significant [33]. The comparable coefficient in biological families was found to be about three times larger. If we assume that 0.097 is an unbiased estimate of the direct causal effect, that there is no assortative mating, and that the effect of rearing father's *EduYears* is equal to that of the mother's, then the estimated effect of midparent *EduYears* on offspring *EduYears* is 0.19 with a confidence interval extending beyond 1/4. These assumptions are not realistic, but nevertheless it is evident that an effect of roughly 1/4 is not at all unreasonable.

References

[1] Vilhjlmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrm S, Ripke S, *et al.* Modeling linkage disquilibrium increases the accuracy of polygenic risk scores. *American Journal of Human Genetics* 2015; **97**: 576–592. DOI: 10.1016/j.ajhg.2015.09.001.

[2] Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* 2018; **50**: 1112–1121. DOI: 10.1038/s41588-018-0147-3.

[3] Wickham H, Franois R, Henry L, Mller K. *dplyr: A Grammar of Data Manipulation*. R package version 0.7.8. 2018.

[4] Ahlmann-Eltze C. *ggsignif: Significance Brackets for 'ggplot2'*. R package version 0.4.0. 2017.

[5] Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer, 2016.

[6] Wilke CO. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 0.9.4. 2019.

[7] Gillespie J. *Population Genetics: A Concise Guide*. 2nd ed. Baltimore, MD: John Hopkins University Press, 2004.

[8] Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. New York, NY: Cambridge University Press, 2009.

[9] Lee JJ. Correlation and causation in the study of personality (with discussion). *European Journal of Personality* 2012; **26**: 372–412. DOI: 10.1002/per.1863.

[10] McDonald RP. *Test Theory: A Unified Treatment*. Mahwah, NJ: Erlbaum, 1999.

[11] Miller MB, Basu S, Cunningham J, Eskin E, Malone SM, Oetting WS, *et al.* The Minnesota Center for Twin and Family Research genome-wide association study. *Twin Research and Human Genetics* 2012; **15**: 767–774. DOI: 10.1017/thg.2012.62.

[12] Zhang Y, Qi G, Park JH, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nature Genetics* 2018; **50**: 1318–1326. DOI: 10.1038/s41588-018-0193-x.

[13] Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 2016; **32**: 283–285. DOI: 10.1093/bioinformatics/btv546.

[14] O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nature Genetics* 2018; **50**: 1728–1734. DOI: 10.1038/s41588-018-0255-0.

[15] Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics* 2011; **88**: 76–82. DOI: 10.1016/j.ajhg.2010.11.011.

[16] Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 2010; **42**: 565–569. DOI: 10.1038/ng.608.

[17] Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genetics* 2012; **8**: e1002637. DOI: 10.1371/journal.pgen.1002637.

[18] Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**: 7. DOI: 10.1186/s13742-015-0047-8.

[19] Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 2014; **46**: 1173–1186. DOI: 10.1038/ng.3097.

[20] Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**: 197–206. DOI: 10.1038/nature14177.

[21] Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, *et al.* Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* 2019; **8**: e39702. DOI: 10.7554/eLife.39702.

[22] Berg JJ, Harpak A, Sinnott-Armstrong N, Joergensen AM, Mostafavi H, Field Y, *et al.* Reduced signal for polygenic adaptation of height in UK Biobank. *eLife* 2019; **8**: e39725. DOI: 10.7554/eLife.39725.

[23] Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics* 2015; **96**: 37–53. DOI: 10.1016/j.ajhg.2014.11.010.

[24] Han E, Carbonetto P, Curtis RE, Wang Y, Granka JM, Byrnes J, *et al.* Clustering of 770,000 genomes reveals post-colonial population structure of North America. *Nature Communications* 2017; **8**: 14238. DOI: 10.1038/ncomms14238.

[25] Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nature Genetics* 2018; **50**: 906–908. DOI: 10.1038/s41588-018-0144-6.

[26] Fisher RA. *The Genetical Theory of Natural Selection*. Oxford, UK: Oxford University Press, 1930.

[27] Yengo L, Robinson MR, Keller MC, Kemper KE, Yang Y, Trzaskowski M, *et al.* Imprint of assortative mating on the human genome. *Nature Human Behaviour* 2018; **2**: 948–954. DOI: 10.1038/s41562-018-0476-3.

[28] Bulmer MG. The effect of selection on genetic variability. *American Naturalist* 1971; **105**: 201–211.

[29] Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper DJ, Miller MB, *et al.* Genetic evidence of assortative mating in humans. *Nature Human Behaviour* 2017; **1**: 0016. DOI: 10.1038/s41562-016-0016.

[30] Kong A, Thorleifsson G, Frigge ML, Vilhjlmsson BJ, Young AI, Thorgeirsson TE, *et al.* The nature of nurture: Effects of parental genotypes. *Science* 2018; **359**: 424–428. DOI: 10.1126/science.aan6877.

[31] Liu H. Social and genetic pathways in multigenerational transmission of educational attainment. *American Sociological Review* 2018; **83**: 278–304. DOI: 10.1177/0003122418759651.

[32] Hill WD, Hagenaars SP, Marioni RE, Harris SE, Liewald DCM, Davies G, *et al.* Molecular genetic contributions to social deprivation and household income in UK Biobank. *Current Biology* 2016; **26**: 3083–3089. DOI: 10.1016/j.cub.2016.09.035.

[33] Sacerdote B. How large are the effects from changes in family environment? A study of Korean American adoptees. *Quarterly Journal of Economics* 2007; **122**: 119–157. DOI: 10.1162/qjec.122.1.119.

[34] Bates TC, Maher BS, Medland SE, McAloney K, Wright MJ, Hansell NK, *et al.* The nature of nurture: Using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. *Twin Research and Human Genetics* 2018; **21**: 73–83. DOI: 10.1017/thg.2018.11.

[35] Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A, *et al.* Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences USA* 2018; **115**: E7275–E7284. DOI: 10.1073/pnas.1801238115.

Table S1: The persistence of parent PGS as a significant predictor in regression models predicting offspring phenotypes using different parent characteristics as covariates

Model	Offspring outcome								
	Years of education			Height			BMI		
	β (SE)	P value	ΔR^2	β (SE)	P value	ΔR^2	β (SE)	P value	ΔR^2
Parent PGS alone									
Offspring PGS	0.16 (0.04)	$< .001$	—	0.57 (0.03)	$< .001$	—	0.35 (0.03)	$< .001$	—
Parent PGS	0.20 (0.04)	$< .001$	.018	0.01 (0.03)	.705	~ 0	0.02 (0.03)	.552	~ 0
Parent PGS + IQ									
Offspring PGS	0.17 (0.04)	$< .001$	—	0.57 (0.03)	$< .001$	—	0.35 (0.03)	$< .001$	—
Parent PGS	0.11 (0.05)	.015	.005	0.02 (0.03)	.624	~ 0	0.00 (0.03)	.952	~ 0
Parent PGS + EduYears									
Offspring PGS	0.19 (0.04)	$< .001$	—	0.57 (0.03)	$< .001$	—	0.34 (0.03)	$< .001$	—
Parent PGS	0.04 (0.04)	.243	.005	0.01 (0.03)	.624	~ 0	0.00 (0.03)	.998	$\sim$0
Parent PGS + SES									
Offspring PGS	0.19 (0.04)	$< .001$	—	0.56 (0.03)	$< .001$	—	0.33 (0.03)	$< .001$	—
Parent PGS	0.04 (0.04)	.266	.001	0.02 (0.03)	.536	~ 0	0.01 (0.03)	.711	~ 0
Parent PGS + EduYears + SES									
Offspring PGS	0.19 (0.04)	$< .001$	—	0.57 (0.03)	$< .001$	—	0.34 (0.04)	$< .001$	—
Parent PGS	0.03 (0.04)	.490	$\sim$0	0.01 (0.03)	.695	~ 0	0.00 (0.03)	.983	$\sim$0
Parent PGS + IQ + EduYears									
Offspring PGS	0.19 (0.04)	$< .001$	—	0.57 (0.03)	$< .001$	—	0.35 (0.03)	$< .001$	—
Parent PGS	0.03 (0.04)	.536	$\sim$0	0.01 (0.03)	.685	~ 0	0.00 (0.04)	.916	$\sim$0
Parent PGS + IQ + EduYears + SES									
Offspring PGS	0.20 (0.04)	$< .001$	—	0.57 (0.03)	$< .001$	—	0.35 (0.03)	$< .001$	—
Parent PGS	0.02 (0.04)	.747	$\sim$0	0.01 (0.03)	.683	~ 0	–0.02 (0.04)	.668	$\sim$0

Note: The parent attributes used as covariates in a given set of regression models are given in italics. The two rows beneath each model specification give the partial regression coefficients and so forth of the named predictor. Regression coefficients are standardized.

Table S2: Estimated magnitudes of genetic nurture from all published studies

Study	N	$\delta/\theta_{\mathrm{T}}$
Kong et al. (2018) [30]	21,637	0.701
Bates et al. (2018) [34]	2,335	0.629
Belsky et al. (2018) [35]	804	NA
Liu (2018)—Framingham [31]	3,149	0.803
present study	2,517	0.613

Note: N, the number of families with a genotyped child and at least one genotyped parent; $\delta/\theta_{\mathrm{T}}$, the estimated percentage of the offspring PGS's correlation with offspring *EduYears* that is attributable to the direct causal effect of offspring PGS on offspring *EduYears*. The study by Belsky et al. reports an effect in the same direction as the others (i.e., $\hat{\delta}/\hat{\theta}_{\mathrm{T}} < 1$), but genetic data in its sample was available from mothers only.

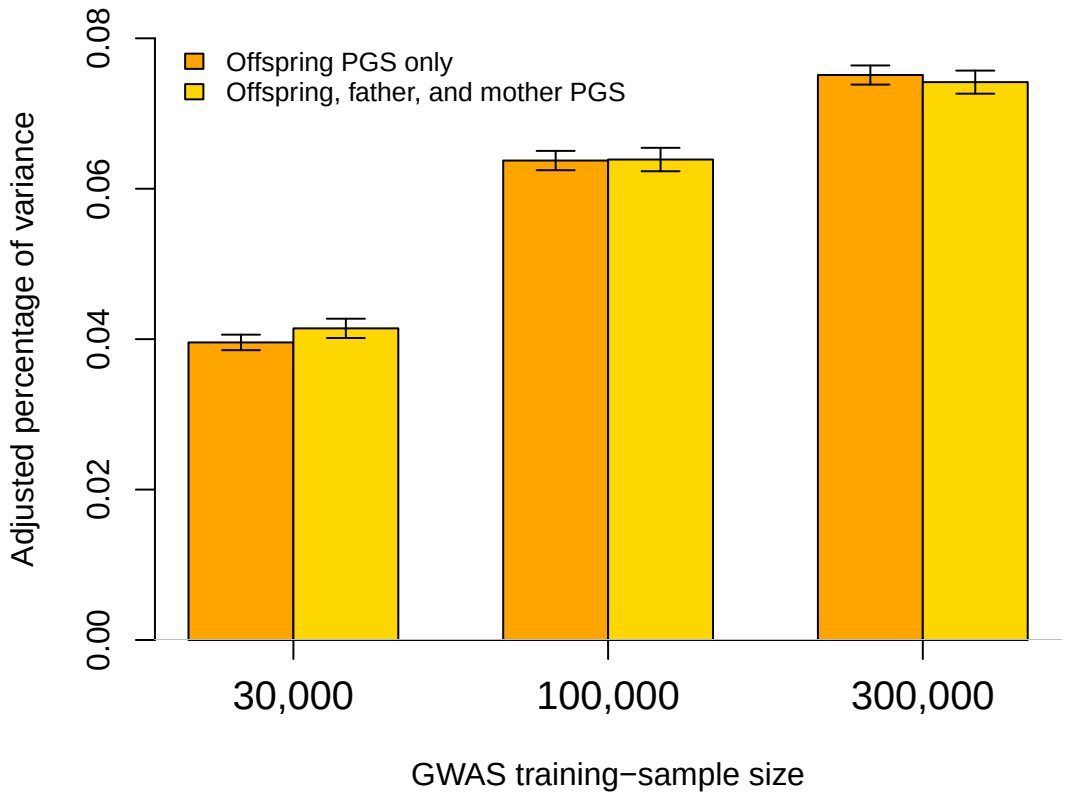

Figure S1: Each bar height corresponds to the average over 200 simulation replicates of the prediction R^2 adjusted for the number of predictors (one vs. three). The error bars enclose 95% confidence intervals. The lack of a conspicuous difference between bar heights within the same sample size indicates that the parent PGS does not provide any substantial increment to the R^2 in the absence of the "genetic nurture" effect, even if the offspring PGS is noisy as a result of sampling error in the training GWAS.

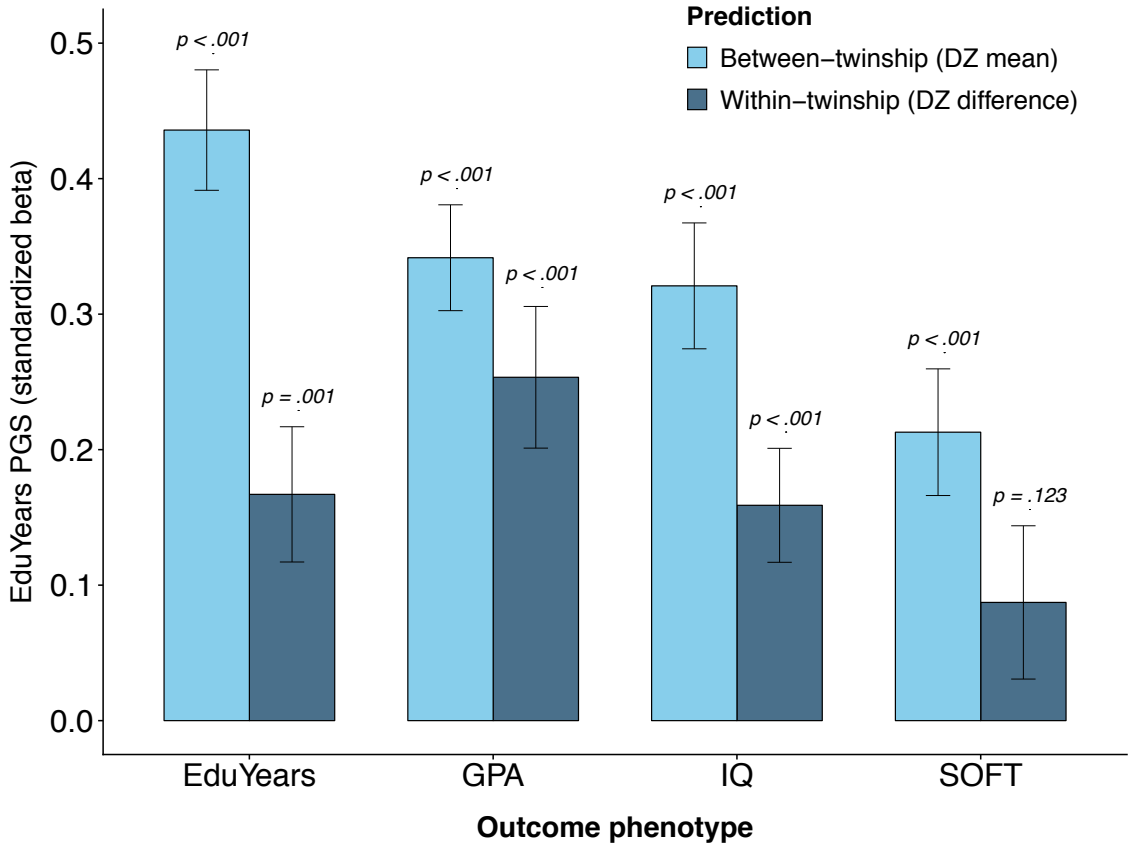

Figure S2: Comparison of β coefficients of offspring PGS on outcomes between dizygotic twinship (mean) and within dyzgotic twinship (difference) for years of education, high school GPA, IQ score, and soft skills (N pairs = 415). Error bars represent ± 1 standard error. All values are standardized.