

HUMAN GENOMICS

The nature of nurture: Effects of parental genotypes

Augustine Kong,^{1,2,3*} Gudmar Thorleifsson,¹ Michael L. Frigge,¹ Bjarni J. Vilhjalmsón,^{4,5} Alexander I. Young,^{1,2,6} Thorgeir E. Thorgeirsson,¹ Stefania Benonisdóttir,¹ Asmundur Oddsson,¹ Bjarni V. Halldorsson,¹ Gisli Masson,¹ Daniel F. Gudbjartsson,^{1,3} Agnar Helgason,^{1,7} Gyda Bjornsdóttir,¹ Unnur Thorsteinsdóttir,^{1,8} Kari Stefansson^{1,8*}

Sequence variants in the parental genomes that are not transmitted to a child (the proband) are often ignored in genetic studies. Here we show that nontransmitted alleles can affect a child through their impacts on the parents and other relatives, a phenomenon we call “genetic nurture.” Using results from a meta-analysis of educational attainment, we find that the polygenic score computed for the nontransmitted alleles of 21,637 probands with at least one parent genotyped has an estimated effect on the educational attainment of the proband that is 29.9% ($P = 1.6 \times 10^{-14}$) of that of the transmitted polygenic score. Genetic nurturing effects of this polygenic score extend to other traits. Paternal and maternal polygenic scores have similar effects on educational attainment, but mothers contribute more than fathers to nutrition- and health-related traits.

How the human genome (nature) and the environment (nurture) work together to shape members of our species is a fundamental question, and any insights into this topic would be an important milestone. One challenge encountered by those who aspire to shed light on this matter is the lack of independence between the genome and the environment; thus, models that fail to account for this limitation are incomplete. Here we demonstrate how the genomes of close relatives—parents and siblings—can affect the proband through their contributions to the environment.

In animal studies, it is well established that alleles in a parent that are not transmitted to the offspring can nonetheless influence the offspring’s phenotypes (1, 2). Most examples involve effects manifested at the fetal stage, at which only the nontransmitted maternal alleles are relevant. In humans, the nontransmitted maternal alleles have been used to examine the potential causal relationships between the state of the mother during pregnancy and the outcomes of the child (3, 4). Here, for humans, we consider an alternative causal path where both paternal and maternal nontransmitted alleles can have effects that are

mostly manifested after birth. A sequence variant that affects the phenotype of an individual is also likely to affect the parent from whom it was inherited (Fig. 1A). For some phenotypes, the state of a parent can influence the state of its child. This gives rise to a situation in which a child’s phenotype is influenced not only by the transmitted paternal and maternal alleles (T_P and T_M) (Fig. 1A) but also by the alleles that were not transmitted (NT_P and NT_M). A good example

is educational attainment (EA) (5, 6): The EA of parents provides an environmental effect for children, but one that has a genetic component (7, 8). We call this phenomenon “genetic nurture.” The transmitted and nontransmitted alleles (Fig. 1A) both exert effects on the parents, and thus both induce genetic nurturing effects. The effect of the transmitted allele includes both its direct effect on the proband and its effect manifested through nurturing from blood relatives. Because the amount of trait variance explained is proportional to the square of effect size, genetic nurture could have a larger impact on variance explained through the transmitted alleles (by magnifying the direct effect) than the nontransmitted alleles. However, data on the nontransmitted alleles are needed to separate the genetic nurturing effects from the direct effects of the transmitted alleles. Specifically, $\hat{\theta}_T$ (transmitted) and $\hat{\theta}_{NT}$ (nontransmitted) denote the respective estimated effects of the alleles when the paternal and maternal alleles are grouped together. Denoting the direct effect as δ , we propose to estimate it by $\hat{\delta} = (\hat{\theta}_T - \hat{\theta}_{NT})$. By calculating the difference, genetic nurturing effects and other potential confounding effects induced by population structure and assortative mating (9, 10) (see below) are cancelled out. Even though the implementations are different, this approach is related to the transmission-disequilibrium test (TDT) (11, 12), as both use nontransmitted alleles as controls (13). However, the potential effects of the nontransmitted alleles are ignored in the TDT. Mathematically, genetic nurture is a form of associative (or indirect) genetic effect, as defined by the animal-breeding literature (2). Genetic nurture is not limited to effects manifested through the phenotypes of the parents, as additional contributions (albeit probably substantially smaller

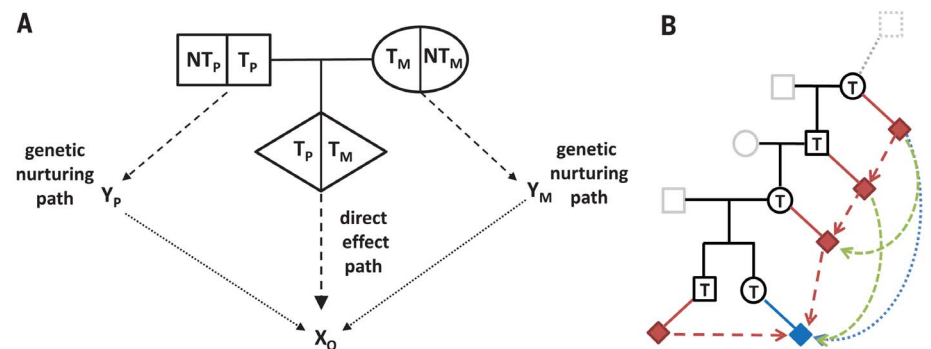


Fig. 1. Direct genetic effect and genetic nurturing effect. (A) Alleles at an autosomal site carried by a parents-offspring trio are labeled with respect to the offspring (proband). T_P and T_M denote, respectively, the alleles transmitted from the father and the mother to the proband, and NT_P and NT_M denote the paternal and maternal alleles that are not transmitted. The transmitted alleles can influence the phenotype of the offspring, X_O , through a direct path. The alleles of the parents, both transmitted and nontransmitted, can influence the parents’ phenotypes, Y_P and Y_M , and through them may have a nurturing effect on X_O . This pathway combines a genetic effect (T_P , NT_P , T_M , and NT_M) on Y_P and Y_M with a nurturing effect (Y_P and Y_M) on X_O . Note that although X_O is often an individual trait of interest, Y would include a much broader set of phenotypes and is not completely known. (B) Red diamonds denote phenotypes of relatives; the blue diamond denotes the phenotype of the proband. Using the maternally transmitted allele as an example (denoted by T), we highlight that, in addition to the parents, the genetic nurturing effect can be manifested through the phenotypes of older ancestors and nonancestors such as siblings.

¹deCODE genetics/Amgen, 101 Reykjavik, Iceland. ²Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7LF, UK. ³School of Engineering and Natural Sciences, University of Iceland, 101 Reykjavik, Iceland. ⁴Bioinformatics Research Centre, Aarhus University, 8000 Aarhus, Denmark. ⁵Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. ⁶Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ⁷Department of Anthropology, University of Iceland, 101 Reykjavik, Iceland. ⁸Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.

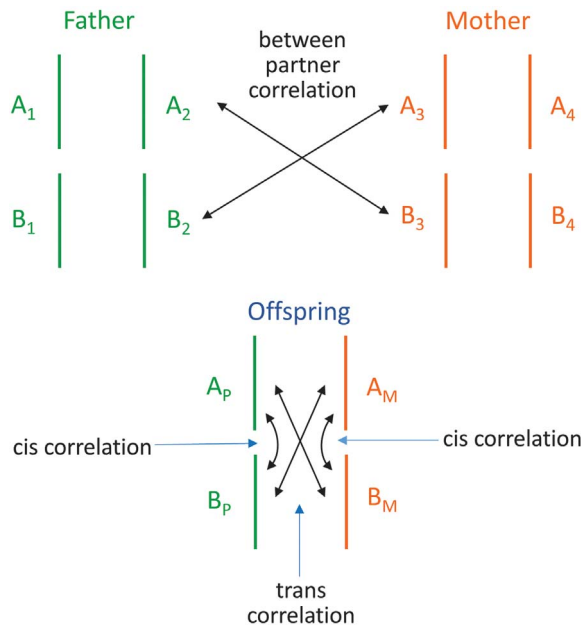
*Corresponding author. Email: augustine.kong@bdi.ox.ac.uk (A.K.); kstefans@decode.is (K.S.)

Table 1. Decomposition of the observed effect of the polygenic score into direct, genetic nurturing, and confounding effects. Traits: educational attainment (EA), age at first child (AGFC), high-density lipoprotein level (HDL), body mass index (BMI), fasting glucose level (FG), height (HT), cigarettes per day for smokers (CPD), and composite health trait (HLTH). Traits are standardized to have a variance of 1. *N*: number of probands with at least one parent genotyped; *N_{NTP}*: number with father genotyped; *N_{NMT}*: number with mother genotyped. $\hat{\theta}_T$ and $\hat{\theta}_{NT}$: estimated effects of the polygenic scores computed for the transmitted and nontransmitted alleles, respectively, when they are analyzed jointly.

$\hat{\delta} = (\hat{\theta}_T - \hat{\theta}_{NT})$: estimated direct effect of the polygenic score. R^2 : estimated variance accounted for by the transmitted polygenic score, which captures both the direct effect and the genetic nurturing effect. R^2_{δ} : estimated variance accounted for by the direct effect alone. These fractions of variance explained are for trait values adjusted for sex, yob (year of birth), and PCs. Corresponding values for unadjusted trait values would be somewhat smaller (13). $\hat{\phi}_{\delta}$, $\hat{\eta}$, and $\hat{\phi}_{\eta}$: estimates, respectively, of the assortative mating–induced confounding effect for the direct effect component, the genetic nurturing effect, and the confounding effect of the genetic nurturing component.

Trait	<i>N</i>	<i>N_{NTP}</i>	<i>N_{NMT}</i>	Transmitted <i>T</i> (<i>T</i> = <i>T_P</i> + <i>T_M</i>)			Nontransmitted <i>NT</i> (<i>NT</i> = <i>NT_P</i> + <i>NT_M</i>)		<i>R</i> ² _δ (%)	$\hat{\delta}/\hat{\theta}_T$	$\hat{\phi}_{\delta}/\hat{\theta}_T$	$\hat{\eta}/\hat{\theta}_T$	$\hat{\phi}_{\eta}/\hat{\theta}_T$
				$\hat{\theta}_T$	<i>P</i>	<i>R</i> ² (%)	$\hat{\theta}_{NT}$	<i>P</i>					
EA	21637	13948	19012	0.223	1.6×10^{-174}	4.98	0.067	1.6×10^{-14}	2.45	0.701	0.046	0.224	0.029
AGFC	54372	35294	47052	0.108	9.7×10^{-110}	1.17	0.039	2.9×10^{-13}	0.48	0.640	0.052	0.264	0.043
HDL	46872	30855	40788	0.065	9.0×10^{-29}	0.42	0.027	6.0×10^{-6}	0.14	0.586	0.046	0.319	0.050
BMI	39078	26433	34533	-0.060	1.0×10^{-22}	0.36	-0.017	0.0077	0.19	0.718	0.055	0.197	0.030
FG	34767	22959	30222	-0.051	7.6×10^{-18}	0.26	-0.018	0.0059	0.11	0.655	0.052	0.252	0.040
HT	39270	26563	34703	0.052	6.6×10^{-14}	0.28	0.030	1.5×10^{-5}	0.05	0.422	0.031	0.476	0.071
CPD	18887	12371	16589	-0.055	1.4×10^{-12}	0.31	-0.030	5.3×10^{-4}	0.06	0.461	0.035	0.439	0.066
HLTH	62328	41996	54546	0.082	2.7×10^{-60}	0.67	0.033	8.9×10^{-11}	0.23	0.592	0.051	0.305	0.052

Fig. 2. Correlation and confounding induced by assortative mating. An example of two loci, A and B, contributing to the phenotype. Through assortative mating, alleles in the father become correlated with alleles in the mother. Consequently, the transmitted paternal alleles (*A_P* and *B_P*) become correlated with the maternally transmitted alleles (*A_M* and *B_M*). This correlation between alleles with different parental origins is referred to as trans correlation, whereas correlation between alleles with the same parental origins (e.g., *A_P* and *B_P*) is referred to as cis correlation. When *A_P*/*A_M* and *B_M*/*B_P* are correlated, association analysis between the phenotype and A alone will also capture part of the effect of B.



ones) may go through grandparents and great-grandparents, for example (Fig. 1B). This study takes advantage of our human data to empirically examine the magnitudes of such effects for traits such as EA.

Estimating direct effects

To maximize the power to detect the effects of the nontransmitted alleles, we used 618,762 single-nucleotide polymorphisms (SNPs) spanning the genome to construct polygenic scores (14). The per-locus allele-specific weightings for the polygenic scores were derived from applying LDpred

(15) to the results of a large genome-wide association study (8) (GWAS) of EA measured in years of education, with Icelandic data removed (13). The first analysis focused on 21,637 Icelandic probands, born between 1940 and 1983 (9139 males, 12,498 females), with EA data and at least one parent genotyped (Table 1). Because we could establish the parent of origin of the transmitted alleles (16), the nontransmitted allele from a genotyped parent was easily determined. *poly_{TP}* and *poly_{TM}* represent the polygenic scores computed from the transmitted paternal and maternal alleles, respectively, and *poly_{NTP}* and *poly_{NMT}* de-

note the corresponding polygenic scores for the nontransmitted alleles. To maximize power, we start by providing the results for *poly_T* = *poly_{TP}* + *poly_{TM}* and *poly_{NT}* = *poly_{NTP}* + *poly_{NMT}*. Here, *poly_{TP}* and *poly_{TM}* are scaled so that *poly_T* has a mean of 0 and a variance of 1, and the trait EA is standardized to have a variance of 1. *poly_{NTP}* and *poly_{NMT}* were similarly calculated, and a 0 was imputed when the parent was not genotyped (13). Associations between EA and the polygenic scores computed from a joint analysis of *poly_T* and *poly_{NT}* that adjusts for sex, year of birth (yob) up to the cubic term, interactions between sex and yob, and 100 principal components (PCs) (13) are presented in Table 1. The estimated effect of *poly_T*, $\hat{\theta}_T$, is 0.223 and significant [*P* = 1.6×10^{-174} , calculated with genomic control adjustment (13, 17)]. Because both *poly_T* and EA are standardized, the estimated fraction of the trait variance explained by *poly_T* is $\hat{\theta}_T^2 = (0.223)^2 = 4.98\%$ (*R*² in Table 1). However, the estimated effect of *poly_{NT}*, $\hat{\theta}_{NT} = 0.067$, is also significant (*P* = 1.6×10^{-14}). Thus, the estimated direct effect of *poly_T*, $\hat{\delta} = (\hat{\theta}_T - \hat{\theta}_{NT}) = 0.157$, explains only $R^2_{\delta} = 0.157^2 = 2.45\%$ of the trait variance, approximately one-half of *R*². Noting that $R^2_{\delta}/R^2 = (\hat{\delta}/\hat{\theta}_T)^2$, the value of $\hat{\delta}/\hat{\theta}_T$ is determined (Table 1). In addition to the polygenic scores, individual results for 120 genome-wide significant SNPs (*P* < 5×10^{-8}) in the Iceland-excluded meta-analysis are provided (table S1). Fifteen of the 120 SNPs (12.5%) have a one-tailed *P* value that is <0.05 for the nontransmitted alleles, which is more than that expected from noise [*P* = 1.5×10^{-3} (13)]. The average estimated effect of the nontransmitted alleles is 34.2% of that of the transmitted alleles. These results are consistent with previous observations that within-family EA effects calculated for dizygotic twins tended to be

Table 2. Parent-of-origin-specific effects of the polygenic scores. Traits are as defined in Table 1. $\hat{\theta}_{TP}$, $\hat{\theta}_{TM}$, $\hat{\theta}_{NTP}$, and $\hat{\theta}_{NTM}$: estimates of the effect of the polygenic scores $poly_{TP}$, $poly_{TM}$, $poly_{NTP}$ and $poly_{NTM}$, respectively. $\hat{\eta}_P$ and $\hat{\eta}_M$: estimates of the paternal and maternal genetic nurturing effects, respectively.

Trait	Transmitted				Nontransmitted				Estimate of $\eta_M - \eta_P$	P	$\hat{\eta}_M/\hat{\eta}_P$
	$\hat{\theta}_{TP}$	P	$\hat{\theta}_{TM}$	P	$\hat{\theta}_{NTP}$	P	$\hat{\theta}_{NTM}$	P			
EA	0.214	1.0×10^{-89}	0.232	9.9×10^{-103}	0.066	5.2×10^{-7}	0.067	3.6×10^{-9}	0.011	0.31	1.26
AGFC	0.100	1.4×10^{-52}	0.116	1.9×10^{-68}	0.034	2.7×10^{-5}	0.043	1.2×10^{-9}	0.013	0.067	1.59
HDL	0.062	5.0×10^{-16}	0.068	7.8×10^{-19}	0.013	0.13	0.037	2.0×10^{-6}	0.014	0.077	2.05
BMI	-0.059	5.2×10^{-13}	-0.062	1.2×10^{-13}	-0.019	0.055	-0.016	0.062	-0.000	0.98	1.02
FG	-0.043	1.0×10^{-7}	-0.059	4.1×10^{-13}	-0.011	0.27	-0.023	0.0073	-0.014	0.090	3.99
HT	0.035	8.8×10^{-5}	0.070	1.3×10^{-14}	0.027	0.0082	0.033	3.5×10^{-4}	0.023	0.011	2.85
CPD	-0.042	1.3×10^{-4}	-0.069	2.6×10^{-10}	-0.036	0.0071	-0.025	0.028	-0.011	0.33	1.63
HLTH	0.070	1.4×10^{-26}	0.093	1.0×10^{-44}	0.026	7.5×10^{-4}	0.039	8.4×10^{-9}	0.019	0.0048	2.32

smaller than the standard GWAS effect estimates (7, 8).

Assortative mating and estimating the genetic nurturing effect

We designate η to denote the magnitude of the genetic nurturing effect. Even though our analyses have adjustment for 100 PCs, which should have eliminated much of the population stratification-induced confounding, $\hat{\theta}_{NT}$ can still capture effects other than η . When there is assortative mating with respect to the genetic component underlying EA (10), a subtle confounding effect may result. Figure 2 illustrates a simple scenario in which the phenotype is assumed to be influenced by two loci: A and B. If there is assortative mating in the parents' generation, it would lead to correlation of alleles between partners; for instance, the A alleles of the father (A_1 and A_2 in Fig. 2) will be correlated with the B alleles of the mother (B_3 and B_4) and vice versa. Consequently, the paternally transmitted A allele A_P will be positively correlated with the maternally transmitted B allele B_M , and A_M will be correlated with B_P . This correlation between alleles inherited from different parents is referred to as trans correlation, whereas the correlation between alleles inherited from the same parent (e.g., A_P and B_P) is referred to as cis correlation. This assortative mating-induced correlation differs from correlation between markers that are close physically, that is, within the same linkage-disequilibrium block. The latter correlation is mainly driven by the cis component, whereas the assortative mating-induced correlation could be dominated by the trans component. If trait association is calculated for locus A individually, the observed effect will capture both the effect of locus A and part of the effect of locus B. We let ϕ_δ denote this added confounding effect. Similarly, assortative mating would also lead the A alleles to capture some of the nurturing effect of locus B, an effect denoted by ϕ_η . Under our model assumptions (13)

$$\phi_\eta/\eta = 2 \times (\phi_\delta/\delta)$$

The factor of 2 arises because the nontransmitted alleles have the same nurturing effects as the

transmitted ones, and thus the transmitted and nontransmitted A alleles are capturing, through correlation, the nurturing effects of both the transmitted and nontransmitted B alleles. Additionally, we have the decompositions

$$E[\hat{\theta}_T] = \delta + \phi_\delta + \eta + \phi_\eta$$

and

$$E[\hat{\theta}_{NT}] = \phi_\delta + \eta + \phi_\eta$$

where $E[\]$ denotes expectation. Because both the transmitted and nontransmitted A alleles capture the confounding effects, $\hat{\delta} = (\hat{\theta}_T - \hat{\theta}_{NT})$ remains an appropriate estimate of the direct effect δ . Locus A and locus B in Fig. 2 can be generalized to represent two nonoverlapping sets of loci. For our study, we think of locus A as the EA polygenic score, whereas locus B represents the genetic component of EA that is statistically orthogonal to locus A (under a scenario of no assortative mating). The mathematical relationships highlighted above continue to hold for the polygenic scores, either exactly or approximately. By using a method for estimating heritability that also incorporates data on the nontransmitted alleles (18), we estimate the full genetic component of EA to have a direct effect that explains 17.0% of the variance of EA. In other words, $poly_T$ is estimated to be $2.45/17.0 = 14.4\%$ of the full genetic component, whereas the remaining 85.6% corresponds to the B components. From this estimate, we extrapolate the correlations observed between the paternal polygenic scores ($poly_{TP}$ and $poly_{NTP}$) and the maternal polygenic scores ($poly_{TM}$ and $poly_{NTM}$) to estimate the correlations between them and the unobserved B components (13). From the latter, ϕ_δ/δ and ϕ_η/η are estimated as 0.065 and 0.130, respectively. For this calculation, we avoided making the assumption that assortative mating between parents was manifested only through correlation of their EAs, which would have led to lower estimates for the ϕ values (13). From these estimates and the above equations, $\hat{\phi}_\delta$, $\hat{\eta}$, and $\hat{\phi}_\eta$ were computed and presented in Table 1 as fractions of $\hat{\theta}_T$. For EA, $\hat{\eta}$ accounts for ~75% of the value of $\hat{\theta}_{NT}$ and $\hat{\eta}$ is

31.9% of $\hat{\delta}$. Finally, we note that assortative mating occurring before the parents' generation could lead to additional confounding. However, this effect appears to be negligible in our study, as after adjustment for 100 PCs, the within-parent correlation of the transmitted and nontransmitted polygenic scores is actually negative (but $P > 0.05$) (13).

Direct and nurturing effects on other traits

The EA polygenic score is associated with other quantitative traits in our database. Among them, those with the strongest statistical significance (Table 1) are age at first child (AGFC) (19), high-density lipoprotein level (HDL) (20), body mass index (BMI) (21), fasting glucose level (FG) (22), height (HT) (23), and cigarettes smoked per day by smokers (CPD) (24). The effects of the transmitted and nontransmitted EA polygenic scores on these phenotypes were estimated as before for the EA phenotype (Table 1). Although the fraction of variance explained by $poly_T$ (R^2) is smaller than that for EA, the effect of $poly_{NT}$ is statistically significant. Moreover, except for BMI, the ratio $\hat{\eta}/\hat{\delta}$ is higher for these traits than for EA and exceeds 1 for HT.

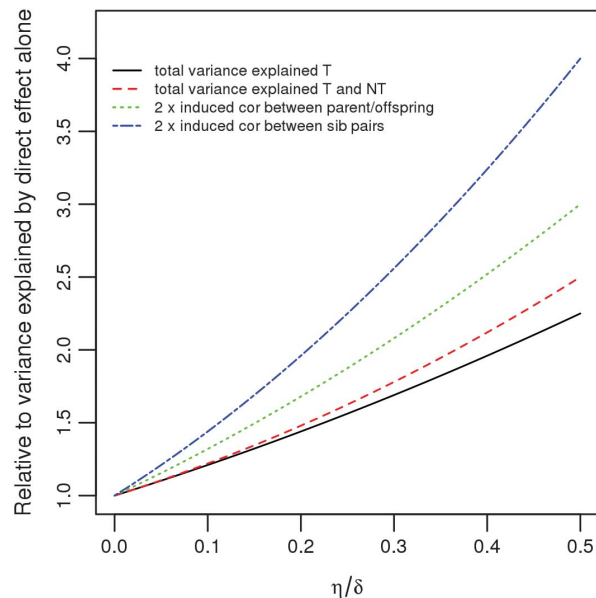
Parent of origin

Table 2 provides the estimated effects of $poly_{TP}$, $poly_{TM}$, $poly_{NTP}$, and $poly_{NTM}$ separately (13). For EA, $\hat{\theta}_{NTP}$, the estimated effect of $poly_{NTP}$, is significant ($P = 5.2 \times 10^{-7}$), and its value is nearly identical to that of $\hat{\theta}_{NTM}$ (the higher P value for $poly_{NTP}$ is due to fewer fathers genotyped than mothers). This indicates that the effect observed for $poly_{NT}$ is not driven by epigenetic effects such as imprinting or genetic interactions between fetus and mother in the womb and does capture a genetic nurturing effect [also see tables S2 and S3, which have results for polygenic scores calculated without SNPs in imprinted regions (25)]. However, even with both parents contributing to genetic nurture, the magnitude of the effect can differ between fathers and mothers. We designate η_P and η_M to denote the paternal and maternal genetic nurturing effects, respectively. Because the transmitted alleles also contribute to the

Fig. 3. Variance explained and induced correlations between phenotypes of parent-offspring and siblings.

Results are displayed as a function of the ratio η/δ . The y axis is the relative amplification; that is, various measures relative to what can be accounted for by the direct effect alone, the latter proportional to δ^2 . The total variance explained by the transmitted alleles is proportional to $(\delta + \eta)^2$ [the plotted curve is hence $(\delta + \eta)^2 / \delta^2$], whereas the total variance explained by the transmitted alleles plus the nontransmitted alleles is proportional to $(\delta + \eta)^2 + \eta^2$. Formulas for the induced parent-offspring and sibling correlations are derived (13).

η , magnitude of genetic nurturing effect; δ , direct effect; T, transmitted; NT, nontransmitted.



nurturing effect, we use a weighted average of $(\hat{\theta}_{TM} - \theta_{TP})$ and $(\hat{\theta}_{NTM} - \hat{\theta}_{NTP})$, with weights proportional to the inverse square of the standard error (13), to estimate $(\eta_M - \eta_P)$ (Table 2). Combining this estimate with $\hat{\eta}$ from Table 1, considered as an estimate of a weighted average of η_P and η_M with weights proportional to the numbers of fathers and mothers genotyped, we calculated individual estimates of η_P and η_M (13), denoted by $\hat{\eta}_P$ and $\hat{\eta}_M$, and the ratio $\hat{\eta}_M/\hat{\eta}_P$ (Table 2). For EA, $(\eta_M - \eta_P)$ is estimated to be 0.011, but it is not significantly different from zero ($P = 0.31$)—that is, the ratio $\hat{\eta}_M/\hat{\eta}_P = 1.26$ is not significantly different from 1. For all of the other six traits, $\hat{\eta}_M/\hat{\eta}_P > 1$ but was significant only for HT ($\hat{\eta}_M/\hat{\eta}_P = 2.85$, $P = 1.1 \times 10^{-2}$). HDL and FG have P values between 0.05 and 0.10. To increase power, for individuals for whom we had data for one or more of the five health- and nutrition-related traits (HDL, BMI, FG, HT, and CPD), a composite health trait (HLTH) was constructed by taking the sum of the standardized values of the available traits (positive signs for HDL and HT; negative signs for BMI, FG, and CPD) and dividing it by the square root of the number of trait values summed. It was then standardized to have a variance of 1. For HLTH, $\hat{\theta}_{NT}$ has a larger value than that for the individual health- and nutrition-related traits and is highly significant ($P = 8.9 \times 10^{-11}$) (Table 1). Both $\hat{\theta}_{NTP}$ and $\hat{\theta}_{NTM}$ are significant, but $\hat{\eta}_M/\hat{\eta}_P = 2.32$ with a P value of 4.8×10^{-3} (Table 2). This supports the notion that mothers have a stronger nurturing effect than fathers on the health of the child.

Variance explained and effects of siblings

The existence of genetic nurture complicates the estimation and interpretation of heritability (18).

For example, maternal effects have been shown to affect heritability estimates from animal-breeding data (26). Though distinct from the direct effect of inherited genetic variants, we demonstrate here how genetic nurture can be measured and taken into consideration. If there are two uncorrelated variants of the same frequency, one having a direct effect δ only and the other having a nurturing effect η only, then the variance explained is proportional to $\delta^2 + \eta^2$. By comparison, if one variant has both effects, then the variance explained is proportional to $(\delta + \eta)^2 = \delta^2 + 2\delta\eta + \eta^2$ (Fig. 3), with the extra $2\delta\eta$ term. Moreover, $(\delta + \eta)^2$ captures the effect of the transmitted allele(s) only; the phenotypic variance accounted for by the transmitted and nontransmitted alleles together is proportional to $(\delta + \eta)^2 + \eta^2$ (Fig. 3). With EA, $\hat{\eta}/\hat{\delta} = 0.319$, $(\hat{\delta} + \hat{\eta})^2/\hat{\delta}^2 = 1.74$, and $[(\hat{\delta} + \hat{\eta})^2 + \hat{\eta}^2]/\hat{\delta}^2 = 1.84$. Assuming that the direct effect alone explains 17.0% of the variance, the variance explained by the transmitted alleles with the nurturing effects included increases to $17.0\% \times 1.74 = 29.6\%$. Additionally including the nontransmitted alleles would increase the variance explained to $17.0\% \times 1.84 = 31.3\%$. The genetic nurturing effect not only magnifies the variance explained, it also induces an even larger amplification of the phenotypic correlations of parents and offspring and of siblings (Fig. 3) (13). Also worth noting is that the $2\delta\eta$ term highlighted above does not exist for adopted children, as then both alleles of a parent would be nontransmitted.

Genetic nurture could go through a sibling (Fig. 1B) if, as proposed (27), the phenotypes of the proband are influenced by the phenotypes or behavior of a sibling. On the basis of the genealogy, for each EA proband who has at least one sibling, the sibling most likely to have the biggest effect on the proband was identified as follows. If the proband has older siblings, the older sibling

with a job closest to the proband was selected (monozygotic twins were excluded, but we count a dizygotic twin of the proband as an older sibling). If the proband is the eldest child, a younger sibling with the closest job was chosen. There are 7798 probands whose chosen sibling is genotyped and whose parents are both genotyped. A polygenic score, denoted by $poly_{TS}$, was computed using the alleles transmitted from the parents to the sibling. The EA of the proband was then regressed on $poly_T$, $poly_{NT}$, and $poly_{TS}$ jointly. The effect of $poly_{TS}$ is significant ($P = 0.015$) and is estimated to be 24.1% (95% confidence interval: 4.7 to 43.6%) of the direct effect. The uncertainty is large because $poly_{TS}$ is strongly correlated with $poly_T$ and $poly_{NT}$. One compensation is that, having adjusted for both $poly_T$ and $poly_{NT}$, the estimated effect of $poly_{TS}$ is free of confounding from assortative mating.

Heritability is defined as the fraction of phenotypic variance explained by direct effects alone. The presence of parental genetic nurture introduces bias to estimates of heritability from GREML (genomic relatedness-based restricted maximum likelihood)-type methods (28), such as those embodied in the software package GCTA (29), that use correlations due to transmitted alleles without distinction between direct genetic effects and genetic nurturing effects (18). By contrast, heritability estimates from comparing monozygotic versus dizygotic twins (30) are unaffected as the effects of parental genetic nurture are cancelled out. However, when genetic nurturing effects that go through the phenotypes of a sibling or twin are present, both twin-based heritability estimates (31) and estimates from GREML-type methods are affected.

The nature of genetic nurture and other polygenic scores

To further use the EA trait data, we performed analyses that treated the nontransmitted polygenic score of a genotyped parent as missing if the EA of that parent was unknown. For these data, (unadjusted) estimates of θ_{NT} were calculated as before (table S4). Also given are estimates of θ_{NT} adjusted for the EAs of the parents, obtained by adding the latter to the explanatory variables in the regressions. For EA, AGFC, HT, and HLTH, the adjusted estimate remains significant ($P < 0.005$), and the ratio of the adjusted versus unadjusted estimate is, respectively, 47.6, 63.0, 80.3, and 68.6%. This indicates that the EA of the parent is an important part of the parental phenotypes (Y in Fig. 1A) through which genetic nurture operates, but it is far from all of it. The EA polygenic score is likely associated with intelligence, conscientiousness, and future planning. Parents with a high score enhance the nurturing of their offspring through many behaviors, not exclusively through their own EA.

To contrast the results presented for the polygenic score constructed from a GWAS of EA (EA polygenic score), we examined polygenic scores constructed from GWASs of HT (32) (HT polygenic score) and BMI (33) (BMI polygenic score).

(Results corresponding to Table 1 are in tables S5 and S6.) Noting that the HT and BMI polygenic scores are, respectively, positively ($r = 0.087$) and negatively correlated ($r = -0.146$) with the EA polygenic score, we computed HT and BMI polygenic scores adjusted for the EA polygenic score by regressing the former on the latter and calculating the residuals (tables S7 and S8). Whereas the unadjusted nontransmitted polygenic score has a few significant associations (tables S5 and S6), with adjustment (tables S7 and S8) the only significant effect of the nontransmitted polygenic score is between the HT trait and the nontransmitted HT polygenic score. Furthermore, most of this observed effect is estimated to be due to confounding from assortative mating.

Discussion

Through the study of the nontransmitted alleles, we demonstrated that genetic nurturing effects exist and can have an impact on variance explained. These results also reveal that the observed effects from GWAS do not necessarily reflect direct effects alone. They can be amplified by genetic nurturing effects and, to a lesser extent, assortative mating-induced confounding. Owing to power considerations, we mostly studied variants as an aggregate. However, given the complexity of the EA trait (6) and our observed effects of the EA polygenic score on other traits, for individual variants, the ratio of the genetic nurturing effect versus the direct effect must have variations both between and within traits. Thus, we should aim to gather enough data to perform GWAS with the nontransmitted alleles. This would add insight into the pathway(s) through which the effect of an individual variant is manifested, as well as enable a better understanding of some pleiotropic effects (34).

Although genes have been shown to affect the environment (24, 35, 36), the contribution of a genetic effect manifested through nurturing has mostly been ignored in GWAS. Results here highlight the importance of family data.

Our focus has been on genetic nurture in one direction, but the effects are likely to be bidirectional. For a parent-offspring pair, the magnitude of the effect in the direction of parent to offspring is likely to dominate the effect in the opposite direction. However, with siblings and twins, the effects would be reciprocal.

Our analyses implicitly assume that direct genetic effects and genetic nurturing effects are additive, but interactive effects could certainly exist and further complicate the interpretation of observed effects. Moreover, alleles other than those in the parents can also have an effect; for example, the genetic makeup of the population of the probands could also be an important environmental contributor to their phenotypes.

REFERENCES AND NOTES

- B. Griffing, *Aust. J. Biol. Sci.* **20**, 127–139 (1967).
- M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, 1998).
- G. Zhang *et al.*, *PLOS Med.* **12**, e1001865 (2015).
- R. C. Richmond *et al.*, *PLOS Med.* **14**, e1002221 (2017).
- N. E. Hill, D. F. Tyson, *Dev. Psychol.* **45**, 740–763 (2009).
- E. Krapohl *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15273–15278 (2014).
- C. A. Rietveld *et al.*, *Science* **340**, 1467–1471 (2013).
- A. Okbay *et al.*, *Nature* **533**, 539–542 (2016).
- D. S. Falconer, T. F. C. Mackay, *Introduction to Quantitative Genetics* (Pearson, 1996).
- M. R. Robinson *et al.*, *Nat. Hum. Behav.* **1**, 0016 (2017).
- R. S. Spielman, R. E. McGinnis, W. J. Ewens, *Am. J. Hum. Genet.* **52**, 506–516 (1993).
- D. B. Allison, *Am. J. Hum. Genet.* **60**, 676–690 (1997).
- Materials and methods are available as supplementary materials.
- N. R. Wray, M. E. Goddard, P. M. Visscher, *Genome Res.* **17**, 1520–1528 (2007).
- B. J. Vilhjálmsson *et al.*, *Am. J. Hum. Genet.* **97**, 576–592 (2015).
- A. Kong *et al.*, *Nature* **462**, 868–874 (2009).
- B. K. Bulik-Sullivan *et al.*, *Nat. Genet.* **47**, 291–295 (2015).
- A. Young *et al.*, Estimating heritability without environmental bias. bioRxiv 218883 [Preprint] 14 November 2017.
- N. Barban *et al.*, *Nat. Genet.* **48**, 1462–1472 (2016).
- A. Helgadóttir *et al.*, *Nat. Genet.* **48**, 634–639 (2016).
- G. Thorleifsson *et al.*, *Nat. Genet.* **41**, 18–24 (2009).
- J. Flannick *et al.*, *Nat. Genet.* **46**, 357–363 (2014).
- D. F. Gudbjartsson *et al.*, *Nat. Genet.* **40**, 609–615 (2008).
- T. E. Thorgeirsson *et al.*, *Nature* **452**, 638–642 (2008).
- S. Benonisdóttir *et al.*, *Nat. Commun.* **7**, 13490 (2016).
- K. Meyer, *Livest. Prod. Sci.* **31**, 179–204 (1992).
- C. Nicoletti, B. Rabe, “Sibling spillover effects in school achievement” [Discussion Paper no. 8615, Institute for the Study of Labor (IZA), 2014].
- J. Yang *et al.*, *Nat. Genet.* **42**, 565–569 (2010).
- J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- T. J. Polderman *et al.*, *Nat. Genet.* **47**, 702–709 (2015).
- M. J. Rietveld, D. Posthuma, C. V. Dolan, D. I. Boomsma, *Behav. Genet.* **33**, 247–255 (2003).
- A. R. Wood *et al.*, *Nat. Genet.* **46**, 1173–1186 (2014).
- A. E. Locke *et al.*, *Nature* **518**, 197–206 (2015).
- B. Bulik-Sullivan *et al.*, *Nat. Genet.* **47**, 1236–1241 (2015).
- R. Plomin, *Behav. Genet.* **44**, 629–638 (2014).
- R. Dawkins, *The Extended Phenotype* (Oxford Univ. Press, 1982).

ACKNOWLEDGMENTS

We thank A. Okbay for providing the EA meta-analysis results with Icelandic data removed; J. Hirshhorn for pointing out that the nontransmitted alleles could be capturing some confounding effects due to assortative mating; and the GIANT consortium, A. Wood, and A. Locke for assisting us to obtain meta-analysis results for HT and BMI with Icelandic data removed. B.V.H. is an associate professor in the School of Science and Engineering at Reykjavik University. A summary of the data used in this manuscript is in table S9. Icelandic law allows for unimpeded sharing of summary-level data. However, the law does not allow the sharing of individual-level data on genotypes and phenotypes outside of Iceland.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6374/424/suppl/DC1
Materials and Methods
Tables S1 to S9
References (37–43)

22 May 2017; accepted 13 December 2017
10.1126/science.aan6877

The nature of nurture: Effects of parental genotypes

Augustine Kong, Gudmar Thorleifsson, Michael L. Frigge, Bjarni J. Vilhjalmsón, Alexander I. Young, Thorgeir E. Thorgeirsson, Stefania Benonisdóttir, Asmundur Oddsson, Bjarni V. Halldorsson, Gisli Masson, Daniel F. Gudbjartsson, Agnar Helgason, Gyda Bjornsdóttir, Unnur Thorsteinsdóttir and Kari Stefansson

Science **359** (6374), 424-428.
DOI: 10.1126/science.aan6877

Genetic variants provide a nurturing environment

Genetic variants in parents may affect the fitness of their offspring, even if the child does not carry the allele. This indirect effect is referred to as "genetic nurture." Kong *et al.* used data from genome-wide association studies of educational attainment to construct polygenic scores for parents that only considered the nontransmitted alleles (see the Perspective by Koellinger and Harden). The findings suggest that genetic nurture is ultimately due to genetic variation in the population and is mediated by the environment that parents create for their children.

Science, this issue p. 424; see also p. 386

ARTICLE TOOLS	http://science.sciencemag.org/content/359/6374/424
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2018/01/24/359.6374.424.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/359/6374/386.full
REFERENCES	This article cites 38 articles, 6 of which you can access for free http://science.sciencemag.org/content/359/6374/424#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the [Terms of Service](#)