

# Supplementary Note for : No evidence for social genetic effects or genetic similarity among friends beyond that due to population stratification.

## Interpretation of Domingue's genetic similarity in terms of kinship differences

In Domingue *et al.* (2014), the authors measure the genetic similarity between two groups of pairs of individuals as the area under the curve defined by the quantiles of the distribution of kinship coefficients under the null (random mating) versus the quantiles of the distribution of kinship coefficients under the alternative (e.g. assortative mating).

To interpret this measure in terms of mean kinship differences, we consider two distributions of kinship coefficients under the null ( $H_0 : \mathcal{N}(0, \sigma^2)$ ) and under the alternative ( $H_1 : \mathcal{N}(-\mu, \sigma^2)$ ). The area of the shaded zone (Fig 1. in Domingue et al; 2014) is therefore defined as

$$I(\mu) = \int_0^1 \Phi_1[\Phi_0^{-1}(u)]du - \frac{1}{2}. \quad (1)$$

where  $\Phi_k$  is the cumulative distribution function of relatedness coefficients under  $H_k$ .

If we posit  $v = \Phi_0^{-1}(u)$ , i.e  $u = \Phi_0(v)$ , then  $du = \phi_0(v)dv$ , with  $\phi_0(\cdot)$  being the probability density function under the null. When  $u = 0$ ,  $v = \Phi_0^{-1}(0) = -\infty$  and when  $u = 1$ ,  $v = \Phi_0^{-1}(1) = +\infty$ .

Equation (1) can then be rewritten as

$$I(\mu) = \int_{-\infty}^{+\infty} \Phi_1(v)\phi_0(v)dv - \frac{1}{2}. \quad (2)$$

Under Gaussian assumptions

$$\Phi_1(v) = \int_{-\infty}^v \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(t+\mu)^2}{2\sigma^2}\right] dt = \int_{-\infty}^{v+\mu} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right] dx = \Phi_0(v + \mu). \quad (3)$$

Therefore

$$I(\mu) = \int_{-\infty}^{+\infty} \Phi_0(v + \mu)\phi_0(v)dv - \frac{1}{2}. \quad (4)$$

The Taylor's series expansion of  $I(\mu)$  near 0 is

$$I(\mu) \approx I(0) + I'(0)(\mu - 0). \quad (5)$$

Noting that  $I(0) = 0$  and that

$$I'(0) = \int_{-\infty}^{+\infty} \phi_0^2(v) dv = \frac{\sqrt{\pi\sigma^2}}{2\pi\sigma^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[-\frac{x^2}{\sigma^2}\right] dx = \frac{1}{\sqrt{4\pi\sigma^2}}, \quad (6)$$

It follows for small values of  $\mu$  that

$$I(\mu) \approx \frac{\mu}{\sqrt{4\pi\sigma^2}}. \quad (7)$$

In Domingue *et al.* (2018), the reported genetic similarity between friends is .031 (Table 1). This implies a difference of kinship coefficients ( $0.5 \times$  the off-diagonal terms of Genetic Relationship Matrix, GRM), between friends compared to random pairs, of  $\sim .031 \times \sqrt{4\pi\sigma^2}$ . The variance of the off-diagonal terms of the GRM can be approximated as  $\sigma^2 \approx 1/M_e = 1/50,000 = 2 \times 10^{-5}$  (Visscher *et al.*; 2014). Therefore a genetic similarity of .031 would correspond to a kinship difference of  $\sim 2.45 \times 10^{-4}$ , i.e. a difference in GRM off-diagonal elements  $\sim 0.0005$ . Numerous subtle factors might lead to such a small difference, e.g. population stratification.

## Expectations of ‘social genetic’ effects reported in Domingue *et al.*

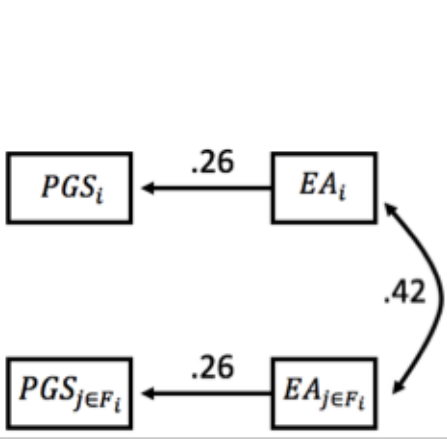
Domingue *et al.* define social-genetic effects as “the influence of one organism’s genotype on a different organism’s phenotype” (p. 704)—in this case, the effects of friends’ genomes on a focal individual’s phenotype. The purpose of this section is to argue that the results reported by Domingue *et al.* in support of social genetic effects among friends are more simply and parsimoniously explained by a model of people befriending individuals whose Educational Attainment (EA) values are similar to one another. Such a process is called homophily or “phenotypic assortment”. Such phenotypic assortment among friends on EA combined with the fact that EA can be predicted from its polygenic score (PGS) necessarily implies correlations between friends’ PGS values as well as correlations between focal individuals’ EA values and their friends’ PGS values. While social-genetic effects cannot be ruled out, we show here that the relationships reported by Domingue *et al.* are expected consequences of simple assortment among friends on EA, and therefore there is no need to resort to more complicated explanations such as social-genetic effects.

We focus first on the central finding that Domingue *et al.* argue is evidence for social genetic-effects of EA: the EA of focal individuals is significantly related with the PGS of their friends. We derive this expected relationship based only on the EA-PGS relationship and the relationship between EAs of friends. We define the following terms:

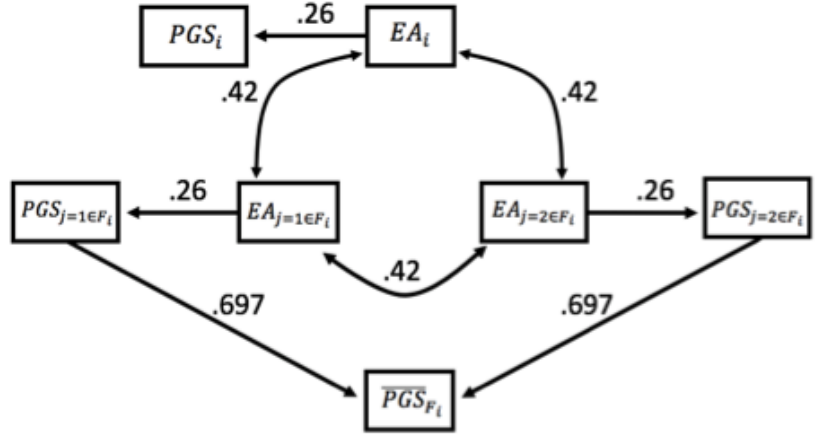
- $EA_i$ : the EA of focal individual  $i$
- $EA_{j \in F_i}$ : the EA of individual  $j$ , a member of individual  $i$ ’s friend group ( $F_i$ )
- $PGS_i$ : the PGS for EA of focal individual  $i$

- $PGS_{j \in F_i}$ : the PGS of EA of individual  $j$ , a member of individual  $i$ 's friend group ( $F_i$ )
- $\overline{PGS}_{F_i}$ : the mean PGS of EA across all friends of individual ( $F_i$ )
- $PGS_{k \in S_i}$ : the PGS of EA of individual  $k$ , a member of individual  $i$ 's schoolmates ( $S_i$ )
- $\overline{PGS}_{S_i}$ : the mean PGS of EA across all schoolmates of individual  $i$  ( $S_i$ )

Using path tracing rules in Figure RS1 below and the correlations presented in Table 2 (of the Domingue et al. manuscript), the expected correlation between  $EA_i$  and  $PGS_{j \in F_i}$  (or equivalently between  $PGS_i$  and  $EA_{j \in F_i}$ ) is  $.26 \times .42 = .109$ , and the expected correlation between  $PGS_i$  and  $PGS_{j \in F_i}$  is  $.26 \times .42 \times .26 = .029$  (all variables are standardized as noted on p. S3 of Domingue et al. Supplementary Materials).



**Figure RS1**



**Figure RS2**

In Figure 2 and Table S6, Domingue *et al.* report the relationship (the slope in this case) between  $EA_i$  and  $\overline{PGS}_{F_i}$ , which is different from the relationship between  $EA_i$  and  $PGS_{j \in F_i}$  derived above and depends on the number of friends included in the average PGS score,  $\overline{PGS}_{F_i}$ . The number of friends an individual had in Add Health varied across individual but had a mean of 2 (Figure S3). For mathematical tractability, we assume the number of friends was constant at 2 across all focal individuals. Furthermore, the slope of the regression  $EA_i \sim \overline{PGS}_{F_i}$  depends crucially on the variances of both variables. The authors state that outcomes and predictors were standardized for this analysis (caption, Figure 2), and our expectations below agree with this. We therefore assume that  $\overline{PGS}_{F_i}$  was standardized after taking the mean. The coefficients to  $\overline{PGS}_{F_i}$  in Figure RS2 below (.697) are those that lead to  $\text{var}(\overline{PGS}_{F_i}) = 1$ , after accounting for the correlation between  $EA_{j=1 \in F_i}$  and  $EA_{j=2 \in F_i}$ . Our model assumes that co-friends of a focal individual are correlated as highly as each friend is to the focal individual, but this assumption has only a minor influence on results: the coefficient from  $PGS_{j \in F_i}$  to  $\overline{PGS}_{F_i}$  would be only slightly different ( $\sqrt{.5} \approx .707$ ) if co-friends' EA values were uncorrelated.

The expected slope of  $EA_i$  regressed on  $\overline{PGS}_{F_i}$ , hereafter denoted  $\mathbb{E} \left[ \widehat{\beta}_{EA_i \sim \overline{PGS}_{F_i}} \right]$ , can be expressed as

$$\mathbb{E} \left[ \widehat{\beta}_{EA_i \sim \overline{PGS}_{F_i}} \right] = \mathbb{E} \left[ \frac{\text{cov}(EA_i, \overline{PGS}_{F_i})}{\text{var}(\overline{PGS}_{F_i})} \right] = \mathbb{E} [\text{cov}(EA_i, \overline{PGS}_{F_i})], \quad (8)$$

given that all variables are standardized.

Using path tracing rules,  $\mathbb{E} [\text{cov}(EA_i, \overline{PGS}_{F_i})] = 2 \times [.697 \times .26 \times .42] = .152$ , which agrees closely with the reported  $\widehat{\beta}_{EA_i \sim \overline{PGS}_{F_i}} = .175 \pm .03$  (Table S6, column 4).

Domingue *et al.* then control for  $PGS_i$  and find that this partial slope is only slightly reduced ( $\widehat{\beta}_{EA_i \sim \overline{PGS}_{F_i} | PGS_i} = .154 \pm .03$ ) and still significant. They interpret this partial slope as evidence "... that the genetics of individuals in a person's social environment influence that person's phenotype," (p. 705). However, controlling for a variable that is only weakly associated with the outcome and predictor variables, such as  $PGS_i$ , is expected to change the slope by only a small amount. In particular, given that all variables are standardized,

$$\mathbb{E} \left[ \widehat{\beta}_{EA_i \sim \overline{PGS}_{F_i} | PGS_i} \right] = \mathbb{E} \left[ \frac{\text{cov}(EA_i, \overline{PGS}_{F_i} | PGS_i)}{\text{var}(\overline{PGS}_{F_i} | PGS_i)} \right] = \mathbb{E} \left[ \frac{r_{EA_i, \overline{PGS}_{F_i}} - (r_{EA_i, PGS_i})(r_{PGS_i, \overline{PGS}_{F_i}})}{1 - r_{PGS_i, \overline{PGS}_{F_i}}^2} \right]. \quad (9)$$

Using path tracing rules and Figure RS2, the correlation between  $PGS_i$  and  $\overline{PGS}_{F_i}$  is  $2 \times [.697 \times .26 \times .42 \times .26] = .0396 \approx .04$ . Thus, under the assumption only of phenotypic assortment among friends' EA values, this expected partial slope is

$$\mathbb{E} \left[ \widehat{\beta}_{EA_i \sim \overline{PGS}_{F_i} | PGS_i} \right] = \frac{.152 - .26 \times .04}{1 - .04^2} = .142. \quad (10)$$

which is, again, close to the partial slope ( $.154 \pm .03$ ) reported in the manuscript. Similarly, the other relationships reported in Domingue *et al.* do not differ from what is expected under phenotypic assortment of EA. For example, the partial slope  $PGS_i \sim \overline{PGS}_{F_i}$  is expected to be .052 vs.  $.06 \pm .03$  reported, and the partial slope  $EA_i \sim \overline{PGS}_{S_i}$  is expected to be .197 vs.  $.22 \pm .03$  reported. It should be noted that the results reported in Figure 1 and Table S4 appear to be regression coefficients where the predictors,  $\overline{PGS}_{F_i}$  or  $\overline{PGS}_{S_i}$ , were not standardized after subtracting the mean. This is seen most clearly in the slope associated with  $\overline{PGS}_{S_i}$  in Table S4 (.62), which is far too high to be a slope between the standardized values of these two variables. Our expectations assumed that  $\overline{PGS}_{F_i}$  and  $\overline{PGS}_{S_i}$  were unstandardized in these analyses.

In summary, the results interpreted by Domingue *et al.* as evidence for social-genetic effects are expected under a simple model of people befriending individuals whose educational values are similar to one another. The fact that individuals' EA scores are correlated with their friends' PGS scores is a necessary consequence of such assortment on EA. While more complicated models explaining these relationships are certainly possible, they are not necessary to explain the presented results.