

## No evidence for social genetic effects or genetic similarity among friends beyond that due to population stratification: a reappraisal of Domingue et al (2018)

Loic Yengo<sup>a</sup>, Morgan Sidari<sup>b</sup>, Karin J. H. Verweij<sup>c</sup>, Peter M. Visscher<sup>a,d</sup>, Matthew C. Keller<sup>e,f</sup>, Brendan P. Zietsch<sup>b,\*</sup>

<sup>a</sup>Institute for Molecular Bioscience, The University of Queensland, Brisbane 4072, Australia; <sup>b</sup>School of Psychology, The University of Queensland, Brisbane 4072, Australia; <sup>c</sup>Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; <sup>d</sup>Queensland Brain Institute, The University of Queensland, Brisbane 4072, Australia; <sup>e</sup>Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA; <sup>f</sup>Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA.

\*To whom correspondence may be addressed.

Email: [zietsch@psy.uq.edu.au](mailto:zietsch@psy.uq.edu.au)

### Abstract

**Using data from 5,500 adolescents from the National Longitudinal Study of Adolescent to Adult Health, Domingue et al. (2018) claimed to show that friends are genetically more similar to one another than randomly selected peers, beyond the confounding effects of population stratification by ancestry. The authors also claimed to show ‘social-genetic’ effects, whereby individuals’ educational attainment (EA) is influenced by their friends’ genes. Neither claim is justified by the data. Mathematically we show that 1) although similarity at causal variants is expected under assortment, the genome-wide relationship between friends (and similarly between mates) is extremely small (an effect that could be explained by subtle population stratification) and 2) significant association between individuals’ EA and their friends’ polygenic score for EA is expected under homophily with no socio-genetic effects.**

The availability of large samples of individuals with genome-wide genetic data in combination with behavioural phenotypes and social outcomes has led to a resurgence in research that address questions at the interface of genetics and the social sciences. Some of that research is hypothesis driven while much of it is data-driven and hypothesis generating. The genetics and statistical analysis of human traits has a solid underpinning theory in quantitative and population genetics (Lynch and Walsh 1998<sup>1</sup>; Walsh and Lynch 2018<sup>2</sup>), and rigorous benchmarking against these underpinnings is essential, especially when novel or unexpected results in human behaviour are reported. In this short note, we highlight one example (and list others) where novel results and claims are not justified by the data presented and instead have alternative and more parsimonious explanations. Using data from 5,500 adolescents from the National Longitudinal Study of Adolescent to Adult Health, Domingue et al.<sup>3</sup> claimed to show that friends are genetically more similar to one another than randomly selected peers, beyond the confounding effects of population stratification by ancestry. The authors also report evidence of ‘social-genetic’ effects, whereby individuals’ educational attainment (EA) is influenced by their friends’ genes. Here we argue that neither claim is justified by the data.

First, we show that individuals who match on phenotype – whether as romantic partners or, in this case, friends – do not exhibit a substantially increased genome-wide similarity. Indeed, Robinson et al.<sup>4</sup> previously showed that the expected genomic relatedness between individuals phenotypically matching on a given trait equals  $rh^2/M_e$ , where  $r$  is the phenotypic correlation between matched

individuals,  $h^2$  the heritability of the trait driving the assortment and  $M_e$  the effective number of independent markers in the genome, estimated in European descents  $\sim 50,000$ .<sup>5</sup> Therefore, assuming the SNP heritability of EA to be  $h^2 \sim 0.12$  (Ref.<sup>6</sup>) and given the correlation of EA between friends ( $r=0.415$ ; from their Table 2) reported by Domingue et al., we would expect a genomic relatedness between friends of  $0.12 \times 0.415 / 50,000$ , i.e.  $\sim 10^{-6}$ . Unfortunately, the friend-pair genomic similarity of 0.031 (95% confidence interval (CI): 0.022 – 0.036) reported by Domingue et al.<sup>3</sup> does not represent an actual estimate of genomic relationship as classically used in the human genetics literature<sup>7</sup>, but is instead an alternative measure introduced by the authors in a previous publication<sup>8</sup>.

In **Supplementary Note 1**, we derive the mathematical relation between Domingue et al.'s similarity measure and the classical genomic relationship. We show that a similarity of 0.031 on the authors' scale is akin to a genomic relationship between friends of  $\sim 5 \times 10^{-4}$ . Although quite small, this value is still over two orders of magnitude larger than the theoretical value of  $\sim 10^{-6}$  expected under phenotypic matching. Given that phenotypic matching cannot explain this large discrepancy, the most likely cause of such genome-wide similarity among friends would be population stratification, whereby individuals are more likely to befriend others living in close geographical vicinity and thus of likely similar ancestry. Domingue et al. acknowledged the possibility of confounding due to population stratification, but in our view did not control for it in a standard way. Rather, they reported (in their Supplementary Materials) a secondary analysis using the program REAP, which they argue is robust to stratification and still reveals a significant (though reduced: from 0.031 to 0.02 (95%CI: 0.011 – 0.028); data from Table S3 in ref.<sup>3</sup>) genomic similarity among friends. However, REAP was designed to estimate kinship among related individuals in admixed samples with heterogeneous continental ancestry<sup>9</sup> – not for estimating genomic similarity among unrelated individuals with homogenous continental ancestry (e.g. Domingue et al.'s sample of European ancestry; T. Thornton, personal communication, April 8, 2018). Given these limitations, it appears to us the most likely and more parsimonious explanation for the observed genomic similarity among friends is within-continental population stratification (e.g. Northern vs. Southern European ancestry).

Second, Domingue et al.'s results do not provide evidence for 'social-genetic' effects as claimed. The purported evidence is a significant association between focal individuals' EA and their friends' polygenic score for EA ( $PGS_{EA}$ ), controlling for focal individuals'  $PGS_{EA}$ . We show mathematically in **Supplementary Note 2** that simple homophily – the well-established tendency for individuals to befriend others with similar educational performance<sup>10,11</sup> – can explain the significant association that Domingue et al. used as evidence for a complex socio-genetic process. While our analyses do not rule out the possibility of 'social-genetic effects', other research using longitudinal sample of high school and university students ( $N=6,000$ ), showed that friend similarity in academic performance was due to initial choice of similar friends, and not to a change in individuals' academic performance towards that of their friends<sup>10</sup>. This finding is inconsistent with 'social-genetic effects' as envisaged by Domingue et al.

In summary, the advent of large samples of genotyped individuals with known social relationships has provided unprecedented opportunities for research at the intersection of human genetics and social sciences. However, analysis and interpretation of these data require great care, and several high-profile papers<sup>8,12</sup> on the genetic similarity of social or romantic mates have forwarded exciting but perhaps misleading interpretations of results that probably have more parsimonious explanations, such as population stratification (e.g. see commentaries refs.<sup>13,14</sup> on papers refs.<sup>8,12</sup>).

## Acknowledgements

This research was supported by the Australian Research Council (DP160102400; FT160100298), the Australian National Health and Medical Research Council (1113400 and 1078037) and the National Institute of Health (NIH grants R01AH042568 and R01MH100141). K.J.H.V. is supported by the Foundation Volksbond Rotterdam.

## References

1. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits*. (Sinauer, 1998).
2. Walsh, B. & Lynch, M. *Evolution and Selection of Quantitative Traits*. (Oxford University Press, 2018).
3. Domingue, B. W. *et al.* The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health. *PNAS* **115**, 702–707 (2018).
4. Robinson, M. R. *et al.* Genetic evidence of assortative mating in humans. **1**, 0016 (2017).
5. Visscher, P. M. *et al.* Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. *PLoS Genetics* **10**, e1004269 (2014).
6. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
7. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565–569 (2010).
8. Domingue, B. W., Fletcher, J., Conley, D. & Boardman, J. D. Genetic and educational assortative mating among US adults. *PNAS* **111**, 7996–8000 (2014).
9. Thornton, T. *et al.* Estimating Kinship in Admixed Populations. *The American Journal of Human Genetics* **91**, 122–138 (2012).
10. Smirnov, I. & Thurner, S. Formation of homophily in academic performance: Students change their friends rather than performance. *PLoS One* **12**, (2017).
11. Tuma, N. B. & Hallinan, M. T. The Effects of Sex, Race, and Achievement on Schoolchildren's Friendships. *Social Forces* **57**, 1265–1285 (1979).
12. Christakis, N. A. & Fowler, J. H. Friendship and natural selection. *PNAS* **111**, 10796–10801 (2014).
13. Abdellaoui, A., Verweij, K. J. H. & Zietsch, B. P. No evidence for genetic assortative mating beyond that due to population stratification. *PNAS* **111**, E4137–E4137 (2014).
14. Chen, G.-B. Where is the friend's home? *Front Genet* **5**, (2014).

## Supplementary Notes

### Note 1: Interpretation of Domingue et al.'s genetic similarity in terms of kinship differences

In Domingue et al.<sup>8</sup>, the authors introduced a new metric to quantify the similarity within pairs of a certain class (e.g. friends or mates) relative to the similarity of random pairs. That new statistic, hereafter denoted  $I(\mu)$ , is an estimator of 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). We derive below an interpretation of that statistic in terms of differences in mean kinship between the two groups of pairs.

We consider two distributions of kinship coefficients: under the null ( $\mathbf{H}_0 : N(0, \sigma^2)$ ) and under the alternative ( $\mathbf{H}_1 : N(-\mu, \sigma^2)$ ). The area of the shaded zone (Fig 1. in Domingue et al.<sup>8</sup>), hereafter denoted  $I(\mu)$ , is therefore defined as

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

where  $\Phi_k$  is the cumulative distribution function of relatedness coefficients under  $\mathbf{H}_k$ .

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 of relatedness under the null. When  $u = 0$ ,  $v = \Phi_0^{-1}(0) = -\infty$  and when  $u = 1$ ,  $v = \Phi_0^{-1}(1) = +\infty$ . Therefore,  $I(\mu)$  can be rewritten as

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

Under Gaussian assumptions, we can show that  $\Phi_1(v) = \Phi_0(v + \mu)$ . Therefore,

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

$I(\mu)$  cannot be calculated analytically. However, for small values of  $\mu$  (near 0) we can derive its Taylor's series as

$$I(\mu) \underset{\mu \rightarrow 0}{\approx} I(0) + I'(0)(\mu - 0) = \frac{\mu}{\sqrt{4\pi\sigma^2}}.$$

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), in friend pairs 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.<sup>5</sup>). Therefore a genetic similarity of .031 would correspond to a kinship difference of  $\sim 2 \times 10^{-4}$ , i.e. a difference in GRM off-diagonal elements  $\sim 0.0005$ . Such a small difference is more likely created by population stratification.

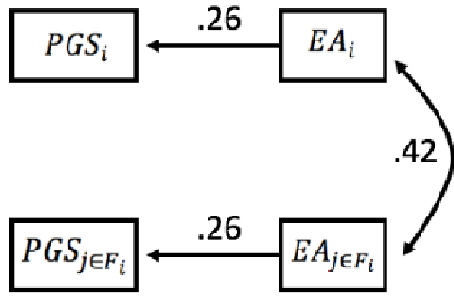
## Note 2: Mathematical expectations of ‘social genetic’ effects reported in Domingue et al.

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 individuals befriending others with similar potential for Educational Attainment (EA) values. 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.

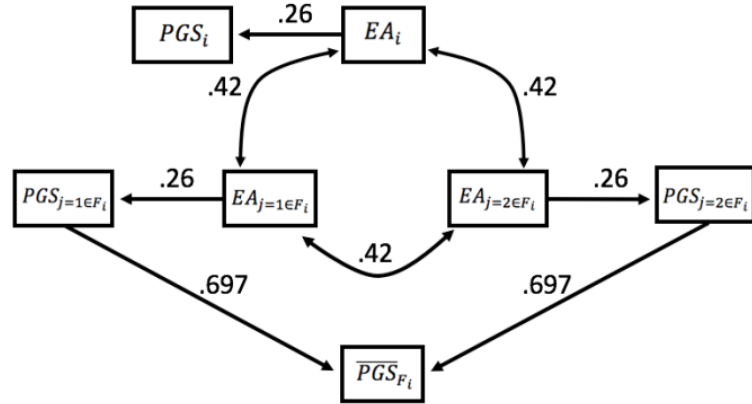
We focus first on the central finding that Domingue et al. argue is evidence for social genetic-effects of EA: the EA of focal individuals is significantly related with the PGS of their friends. We derive this expected relationship based only on the EA-PGS relationship and the relationship between EAs 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  $i$  ( $F_i$ )
- $PGS_{k \in S_i}$ : 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 their Figure 2 and Table S6, Domingue et al. report the relationship (the slope in this case) between  $\overline{PGS}_{F_i}$  and  $PGS_i$ , which is different from the relationship between  $\overline{PGS}_{F_i}$  and  $PGS_{j \in F_i}$  derived above and depends on the number of friends included in the average  $PGS$  score,  $n$ . 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  $\overline{PGS}_{F_i} \sim PGS_i$  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  $PGS_i$ , 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  $\overline{PGS}_{F_i}$  to  $PGS_{j \in F_i}$  would be only slightly different ( $\frac{.26}{.42}$ ) if co-friends' EA values were uncorrelated.

The expected slope of  $PGS_i$  regressed on  $\overline{PGS}_{F_i}$ ,  $\beta$ , can be expressed as

$$\beta = \frac{.26}{.42}$$
 given that all variables are standardized. Using path tracing rules,  $\beta = \frac{.26}{.42}$ , which agrees closely with the reported (Table S6, column 4).

Domingue et al. then control for  $\overline{PGS}_{F_i}$  and find that this partial slope is only slightly reduced ( $\frac{.26}{.42}$ ) 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  $\overline{PGS}_{F_i}$ , is expected to change the slope by only a small amount. In particular, given that all variables are standardized,

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

$$E[\hat{\beta}_{EA_i \sim \overline{PGS}_{F_i} | PGS_i}] = \frac{.152 - .26 \times .04}{1 - .04^2} = .142$$

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 taking 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 of similar educational attainment values. 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 possible, they are not necessary to explain the presented results.