



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 Yengo¹ · Morgan Sidari² · Karin J. H. Verweij³ · Peter M. Visscher¹ · Matthew C. Keller^{4,5} · Brendan P. Zietsch² 

Received: 21 May 2019 / Accepted: 24 October 2019 / Published online: 11 November 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Using data from 5500 adolescents from the National Longitudinal Study of Adolescent to Adult Health, Domingue et al. (Proc Natl Acad Sci 25:256., 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. We argue that neither claim is justified by the data. Mathematically we show that (1) the genetic similarity reported between friends is far larger than theoretically possible if it was caused by phenotypic assortment as the authors claim; uncontrolled population stratification is a likely reason for the genetic similarity they observed, and (2) significant association between individuals’ EA and their friends’ polygenic scores for EA is a necessary consequence of EA similarity among friends, and does not provide evidence for social-genetic effects. Going forward, we urge caution in the analysis and interpretation of data at the intersection of human genetics and the social sciences.

Keywords Genomic similarity · Social-genetic effects · Confound · Kinship

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 addresses 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; Walsh and Lynch 2018), and rigorous benchmarking against these underpinnings is essential—especially when novel or unexpected results in human behaviour are reported. In this paper, 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 5500 adolescents from the National Longitudinal Study of Adolescents 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 find 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.

Handling Editor: Stacey S Cherny.

✉ Brendan P. Zietsch
zietsch@psy.uq.edu.au

¹ Institute for Molecular Bioscience, The University of Queensland, Brisbane 4072, Australia

² School of Psychology, The University of Queensland, Brisbane 4072, Australia

³ Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

⁴ Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA

⁵ Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA

Genetic Similarity of Friends—Phenotypic Assortment or Uncontrolled Population Stratification?

One might intuit that individuals who pair up according to their similarity on a heritable trait (i.e. phenotypic assortment) should be genetically more similar than random pairs of individuals. Although this intuition is technically correct, the induced genetic similarity is trivially small for polygenic traits, as has been shown before (Robinson et al. 2017) and as we reiterate below. In this context, we refer to genetic similarity as the degree to which pairs of individuals—say, romantic partners or in this case friends—share alleles across the whole genome. More specifically, when single nucleotide polymorphism (SNP) data is available, the genetic similarity between two individuals i and j is classically measured using a genomic relationship coefficient (Yang et al. 2010) A_{ij} defined as

$$A_{ij} = \frac{1}{M} \sum_{k=1}^M \frac{(x_{ik} - 2p_k)(x_{jk} - 2p_k)}{2p_k(1 - p_k)}, \quad (1)$$

where M is the number of SNPs used to estimate G_{ij} , p_k the minor allele frequency of SNP k and x_{ik} and x_{jk} the numbers of minor alleles at SNP k that individual i and j possess respectively.

Robinson et al. (2017) previously showed that the expected genomic relationship coefficient between individuals who assort on a trait equals $rh^2\sigma_{GRM}^2$, where r is the trait correlation between assorted individuals, h^2 is the heritability of the trait and σ_{GRM}^2 is the variance of genomic relationship coefficients in the population. Previous studies (Goddard et al. 2009) have shown for a given population and a given set of common SNPs used to calculate A_{ij} , that σ_{GRM}^2 is a fixed quantity which only depends on the effective size of human populations (N_e). Assuming $N_e \sim 10,000$ (Takahata 1993), Visscher et al. (2014) estimated in individuals of European ancestry that for common SNPs, $\sigma_{GRM}^2 \sim 2 \times 10^{-5}$. Therefore, assuming the SNP heritability of EA to be $h^2 \sim 0.12$ (Lee et al. 2018), and given the correlation of EA between friends ($r = 0.415$; from their Table 2) reported by Domingue et al., we would expect the mean genomic relationship coefficients between friends assorting on EA to be $0.12 \times 0.415 \times 2 \times 10^{-5}$, i.e. $\sim 10^{-6}$. Such a slight increase in genetic similarity would require millions of friend pairs to be reliably detected, as already emphasised in Robinson et al. (2017).

How does this expected genetic similarity compare to that reported by Domingue et al.? Unfortunately, such a comparison is not straightforward, because the friend-pair genetic similarity reported by Domingue et al. does not represent a standard measure used in the human genetics literature

(Yang et al. 2010); instead it is an alternative measure introduced by the authors in a previous publication (Domingue et al. 2014). To enable a comparison, we derive the mathematical relation between Domingue et al.'s similarity measure, hereafter denoted $I(\mu)$, and the genomic relationship coefficient (Yang et al. 2010).

$I(\mu)$ is defined in Domingue et al. (2014) as an estimator of the area under the curve defined by the quantiles of the distribution of kinship coefficients under the null (random pairing) versus the quantiles of the distribution of kinship coefficients under the alternative (e.g. assortative mating). Given that kinship coefficients equal half of genomic relationship coefficients, we derive below an interpretation of $I(\mu)$ in terms of differences in mean genomic relationship coefficients between two groups of pairs. We consider two distributions of genomic relationship coefficients under the null ($\mathbf{H}_0: N(0, \sigma_{GRM}^2)$) and under the alternative ($\mathbf{H}_1: N(-\mu, \sigma_{GRM}^2)$). Therefore $I(\mu)$ can be expressed as

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

where Φ_k is the cumulative distribution function of genomic relationship 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 genomic relationship coefficients under the null. When $u = 0$, $v = \Phi_0^{-1}(0) = -\infty$ and when $u = 1$, $v = \Phi_0^{-1}(1) = +\infty$. Moreover, we can show, under Gaussian assumptions, that $\Phi_1(v) = \Phi_0(v + \mu)$. Therefore, $I(\mu)$ can be rewritten as

$$I(\mu) = \int_{-\infty}^{+\infty} \Phi_0(v + \mu)\phi_0(v)dv - 1/2. \quad (3)$$

$I(\mu)$ cannot be calculated analytically from Eq. (3). However, we can still derive its Taylor's series expansion for small values of $|\mu|$ as

$$I(\mu) \underset{|\mu| \rightarrow 0}{\approx} I(0) + I'(0)(\mu - 0) = \mu \sqrt{4\pi\sigma_{GRM}^2} \quad (4)$$

In Domingue et al. (2018), the reported genetic similarity between friends is 0.031 (95% CI 0.022–0.036) (data from Table 1 in Domingue et al. 2018). Therefore, Eq. (4) implies that such a level of genetic similarity would correspond to a mean difference in genomic relationship coefficients of $\sim 0.031 \times \sqrt{4\pi\sigma_{GRM}^2}$, i.e. $\sim 4.9 \times 10^{-4}$ (95% CI 3.5×10^{-4} – 5.7×10^{-4}) between friend pairs and random pairs.

Although quite small in absolute terms, this value is still about 500 times larger than the theoretical value of $\sim 10^{-6}$ expected from friends phenotypically assorting on EA. Indeed, even if friends were perfectly correlated on a phenotype that was 100% heritable, we would still only expect a genetic similarity of $\sim 2 \times 10^{-5}$, still an order of magnitude smaller than the lower bound of the smallest of Domingue

et al.’s relatedness estimate (i.e. 0.11 from their Table S3 $\times \sqrt{4\pi\sigma_{GRM}^2} = 1.7 \times 10^{-4}$). Therefore, the claim by Domingue et al. that the genetic similarity between friends could be due to phenotypically assorting on EA is incompatible with theory and inconsistent with known properties of the human genome and trait variation.

An alternative explanation of such genetic similarity among friends is 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 we do not believe their correction was effective. The standard way to correct for stratification in these types of analyses is to control for principal components from a genomic relationship matrix (Price et al. 2010). Instead, Domingue et al. reported (in their Supplementary Materials) a secondary analysis using the program REAP (Thornton et al. 2012), which they argue is robust to stratification and still reveals a significant genetic similarity among friends (though reduced: from 0.031 (95% CI 0.022–0.036) to 0.02 (95% CI 0.011–0.028); data from Table S3 in Domingue et al. 2018). However, REAP was designed to estimate kinship among related individuals in admixed samples with heterogeneous continental ancestry (Thornton et al. 2012)—not for estimating genetic 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). There is no evidence that REAP effectively corrects for subtle population stratification within a continental ancestry group (e.g. Northern vs. Southern European ancestry). Given the results REAP has yielded in this case are implausible, as described above, it appears to us the most likely and more parsimonious explanation for the observed genomic similarity among friends is residual within-continental population stratification. We cannot rule out other explanations for the high genetic similarity observed, but, importantly, we can rule out the explanation—friends’ assortment on EA scores—provided by Domingue et al.

‘Social-Genetic Effects’ or Simply Like-Befriending-Like?

Domingue et al. report a significant association between focal individuals’ educational attainment (EA) and their friends’ polygenic score for EA (PGS_{EA}), controlling for focal individuals’ PGS_{EA} . They argue that this is evidence for social-genetic effects, whereby individuals’ educational attainment (EA) is influenced by their friends’ genes via direct effects of one’s (heritable) EA on friends’ EA. Here, we demonstrate that, given the predictive ability of PGS_{EA} ,

Domingue et al.’s findings are necessary consequences of the well-documented observation that friends tend to have similar EA values. Thus, the genetic findings they report are irrelevant to understanding why such similarity occurs, and provide no evidence for social-genetic effects over simpler alternatives, such as social homophily (the general tendency to associate and bond with similar others; Smirnov and Thurner 2017; Tuma and Hallinan 1979).

We derive the results reported by Domingue et al. to be evidence for social-genetic effects based only on the EA-PGS correlation and the correlation 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)
- PGS_{F_i} : the mean PGS of EA across all friends of individual i (F_i)

Applying path tracing rules to Fig. 1a (in which correlations are taken from Table 2 of Domingue et al.), the expected correlation between EA_i and $PGS_{j \in F_i}$ (or equivalently

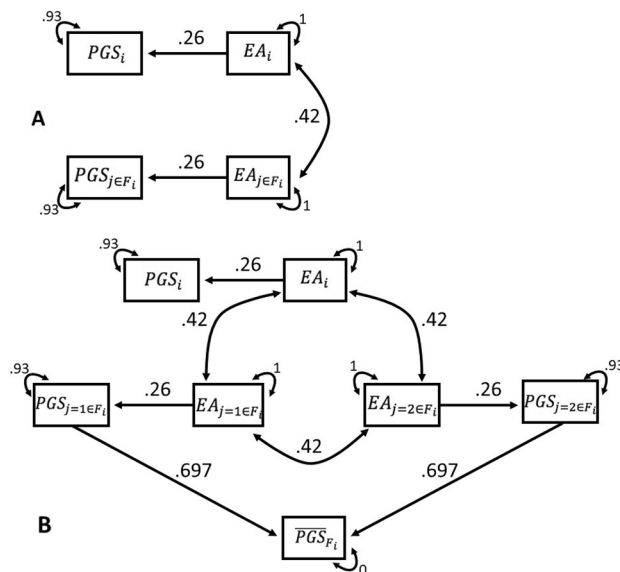


Fig. 1 Path models of the results Domingue et al. use to argue for social-genetic effects. **a** can be used to derive the expected relationship between the polygenic risk scores (PGS) of education between friends given the phenotypic correlation of education between friends reported by Domingue et al. **b** can be used to derive the expected slope of educational attainment (EA_i) of a focal individual regressed on the average educational PGS of their friends ($PGS_{j \in F_i}$). As shown in the diagrams, such associations are necessary consequences of phenotypic assortment

between PGS_i and $EA_{j \in F_i}$) is $0.26 \times 0.42 = 0.109$, and the expected correlation between PGS_i and $PGS_{j \in F_i}$ is $0.26 \times 0.42 \times 0.26 = 0.029$ (all variables are standardized as noted on p. S3 of Domingue et al. Supplementary Materials). Thus, PGS’s are expected to correlate between mates or friends whenever there is phenotypic assortment.

In their Fig. 2 and Table S6, Domingue et al. report the association (the slope in this case) between EA_i and PGS_{F_i} . This differs from the association 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 the study sample varied across individual but had a mean of 2 (Domingue et al. Figure S3). For mathematical tractability, we assume the number of friends was constant at 2 across all focal individuals. Furthermore, the slope of $EA_i \sim \overline{PGS}_{F_i}$ depends on the variances of both variables. The authors state that outcomes and predictors were standardized for this analysis (Domingue et al. Figure 2 caption), and our expectations below agree with this. We therefore assume that PGS_{F_i} was standardized after taking the mean. The coefficients to \overline{PGS}_{F_i} in Fig. 1b (0.697 and 0.697) generate $var(\overline{PGS}_{F_i}) = 1$, after accounting for the correlation between $EA_{j=1 \in F_i}$ and $EA_{j=2 \in F_i}$. Our path 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} , $E[\hat{\beta}_{EA_i \sim \overline{PGS}_{F_i}}]$, can be expressed as $E\left[\hat{\beta}_{EA_i \sim \overline{PGS}_{F_i}}\right] = E\left[\frac{cov(EA_i, \overline{PGS}_{F_i})}{var(\overline{PGS}_{F_i})}\right] = E\left[cov(EA_i, \overline{PGS}_{F_i})\right]$ given that all variables are standardized. Using path tracing rules, $E\left[cov(EA_i, \overline{PGS}_{F_i})\right] = 2 \times [0.697 \times 0.26 \times 0.42] = 0.152$, which agrees closely with the reported $\hat{\beta}_{EA_i \sim} = 0.175 \pm 0.03$ (Table S6, column 4).

Domingue et al. then control for PGS_i and find that this partial slope is only slightly reduced ($\hat{\beta}_{EA_i \sim \overline{PGS}_{F_i} | PGS_i} = 0.154 \pm 0.03$) 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,

$$E[\hat{\beta}_{EA_i \sim \overline{PGS}_{F_i} | PGS_i}] = E\left[\frac{cov(EA_i, \overline{PGS}_{F_i} | PGS_i)}{var(\overline{PGS}_{F_i} | PGS_i)}\right] = E\left[\frac{r_{EA_i, \overline{PGS}_{F_i}} - (r_{EA_i, PGS_i})(r_{PGS_i, \overline{PGS}_{F_i}})}{1 - r_{PGS_i, \overline{PGS}_{F_i}}^2}\right] \tag{5}$$

Using path tracing rules and Fig. 1b, the correlation between PGS_i and \overline{PGS}_{F_i} is $2 \times [0.697 \times 0.26 \times 0.42 \times 0.26] = 0.0396 \cong 0.04$. Therefore, this expected partial slope is

$$E[\hat{\beta}_{EA_i \sim \overline{PGS}_{F_i} | PGS_i}] = \frac{0.152 - 0.26 \times 0.04}{1 - 0.04^2} = 0.142 \tag{6}$$

which is, again, not significantly different from the partial slope (0.154 ± 0.03) reported in the manuscript (Domingue et al. Fig. 2 and Column 2 of Table S7).

Thus, the results interpreted by Domingue et al. as evidence for social-genetic effects are expected under a simple model of individuals befriending others of similar educational attainment values. It is well established that friends have similar educational performance (Smirnov and Thurner 2017; Tuma and Hallinan 1979); more broadly, the general tendency to associate and bond with similar others is one of the most pervasive observations in the social sciences (McPherson et al. 2001). Therefore, there is no need to invoke social-genetic effects to explain Domingue et al.’s findings. Furthermore, other research suggests such effects are unlikely in any substantive sense. Using a large, longitudinal sample of high school and university students (N=6000), Smirnov and Thurner (2017) showed that friend similarity in academic performance is due to initial choice of similar friends, not change in individuals’ academic performance towards that of their friends. This lack of an effect on friends’ academic performance is inconsistent with ‘social-genetic effects’ as envisaged by Domingue et al.

Conclusions

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. Several other high-profile papers (Christakis and Fowler 2014; Connolly et al. 2019; Domingue et al. 2014) on the genetic similarity of social or romantic mates have forwarded exciting but unfounded interpretations of results that probably have more parsimonious explanations, such as population stratification (e.g. see commentaries by Abdellaoui et al. 2014; Chen 2014; Wray and Yengo 2019).

Acknowledgements This research was supported by the Australian Research Council (DP160102400; FT160100298; FL180100072), 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.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by the any of the authors.

References

- Abdellaoui A, Verweij KJH, Zietsch BP (2014) No evidence for genetic assortative mating beyond that due to population stratification. *Proc Natl Acad Sci* 111:E4137. <https://doi.org/10.1073/pnas.1410781111>
- Chen G-B (2014) Where is the friend's home? *Front Genet*. <https://doi.org/10.3389/fgene.2014.00400>
- Christakis NA, Fowler JH (2014) Friendship and natural selection. *Proc Natl Acad Sci USA* 111:10796–10801. <https://doi.org/10.1073/pnas.1400825111>
- Connolly S, Anney R, Gallagher L, Heron EA (2019) Evidence of assortative mating in autism spectrum disorder. *Biol Psychiatry*. 25:256. <https://doi.org/10.1016/j.biopsych.2019.04.014>
- Domingue BW, Belsky DW, Fletcher JM, Conley D, Boardman JD, Harris KM (2018) The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health. *Proc Natl Acad Sci* 25:256. <https://doi.org/10.1073/pnas.1711803115>
- Domingue BW, Fletcher J, Conley D, Boardman JD (2014) Genetic and educational assortative mating among US adults. *Proc Natl Acad Sci* 111:7996–8000. <https://doi.org/10.1073/pnas.1321426111>
- Goddard ME, Wray NR, Verbyla K, Visscher PM (2009) Estimating effects and making predictions from genome-wide marker data. *Stat Sci* 24:517–529
- Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Linnér RK, Fontana MA, Cesarini D (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 50:1112–1121. <https://doi.org/10.1038/s41588-018-0147-3>
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA
- McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27:415–444
- Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* 11:459–463. <https://doi.org/10.1038/nrg2813>
- Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, Peyrot WJ, Abdellaoui A, Zietsch BP, Nolte IM, van Vliet-Ostapchouk JV, Visscher PM (2017) Genetic evidence of assortative mating in humans. *Nat Hum Behav* 1:0016. <https://doi.org/10.1038/s41562-016-0016>
- Smirnov I, Thurner S (2017) Formation of homophily in academic performance: students change their friends rather than performance. *PLoS ONE* 12:e0183473. <https://doi.org/10.1371/journal.pone.0183473>
- Takahata N (1993) Allelic genealogy and human evolution. *Mol Biol Evol* 10:2–22
- Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N (2012) Estimating kinship in admixed populations. *Am J Hum Genet*. 91:122–138. <https://doi.org/10.1016/j.ajhg.2012.05.024>
- Tuma NB, Hallinan MT (1979) The effects of sex, race, and achievement on schoolchildren's friendships. *Soc Forces* 57:1265–1285
- Visscher PM, Hemani G, Vinkhuyzen AA, Chen GB, Lee SH, Wray NR, Goddard ME, Yang J (2014) Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet* 10:e1004269. <https://doi.org/10.1371/journal.pgen.1004269>
- Walsh, B., & Lynch, M. (2018). *Evolution and selection of quantitative traits*: Oxford University Press, Oxford
- Wray NR, Yengo L (2019) Assortative mating in autism spectrum disorder: toward an evidence base from DNA data, but not there yet. *Biol Psychiatry* 86:250–252. <https://doi.org/10.1016/j.biopsych.2019.06.007>
- Yang JA, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42:565–U131. <https://doi.org/10.1038/ng.608>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.