

No evidence for genetic assortative mating beyond that due to population stratification

Domingue et al. (1) use genome-wide SNPs to show in non-Hispanic US whites that spouses are genetically more similar than random pairs of individuals. We argue that, although this reported result is descriptively true, the spousal genetic similarity can be explained by assortment on shared ancestry (i.e., population stratification) and thus does not reflect genetic assortative mating as interpreted by Domingue et al. This greatly affects the implications of the findings for understanding assortative mating in humans.

Genetic population stratification is a major driver of genetic spouse similarities as a consequence of ethnic or social homogamy and/or geographic proximity. Social homogamy occurs when individuals from similar social, ethnic, or demographic backgrounds are more likely to mate. Residential proximity is a strong predictor of who marries whom (2), and European ancestry within the United States is not independent from geographic location or social group (3). People of European ancestry show systematic genetic differences depending on where in Europe their ancestors originate from, as cataloged in the 1000 Genomes and HapMap projects (www.1000genomes.org).

Excluding non-Caucasian individuals is not sufficient to account for population stratification (especially when only based on self-reports). The standard method to deal with

(subtle) population stratification is to control for the major dimensions of genetic variation [principle components (PCs)] (4). When Domingue et al. use this established method, the slight genetic similarity between mates disappeared (their *SI Text* and table S1). However, this result is not reported in the main text—the authors instead reported results of three alternative controls that did not completely eliminate spousal genetic similarity. These methods are nonstandard and, we argue, inadequate.

The first method restricted analyses to individuals with less variability on the genetic PCs; this method will reduce but not eliminate heterogeneity in ancestry. Even within single European populations from geographically small regions, PCs reflecting the relatively small ancestry differences show significant spouse correlations (5).

The second method was to control for geographical region of birth as a proxy for ethnicity. However, the census division used does not correspond geographically with the clustering of different European or global ancestries in the United States (3).

The third method was to restrict analyses to SNPs the authors claim show very little evidence of population stratification. However, using PCs obtained from the European populations in the 1000 Genomes project, Fig. 1 shows that that even the most conser-

vatively restricted set of SNPs captures European ancestry differences just as well as the rest of the SNPs (both PCs showed correlations of >0.9 between the two SNP sets).

Last, Domingue et al. replicate their finding in a “geographically/genetically homogeneous” population, without accounting for population stratification. Using a large subset of the same sample ($n = 515$ spouse pairs), we found that significant spousal genetic similarity is eliminated after correcting for two ancestry-informative PCs projected from the HapMap-3 dataset; the empirical P value changed from 0.001 to 0.343.

In conclusion, we believe that Domingue et al. did not provide sufficient evidence for spousal genetic similarity beyond that due to population stratification.

Abdel Abdellaoui^{a,1}, Karin J. H. Verweij^{a,b,1}, and Brendan P. Zietsch^{b,c,2}

^aDepartment of Biological Psychology, Vrije Universiteit Amsterdam, 1081 BT, Amsterdam, The Netherlands; ^bSchool of Psychology, University of Queensland, St. Lucia, Brisbane, QLD 4029, Australia; and ^cGenetic Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD 4006, Australia

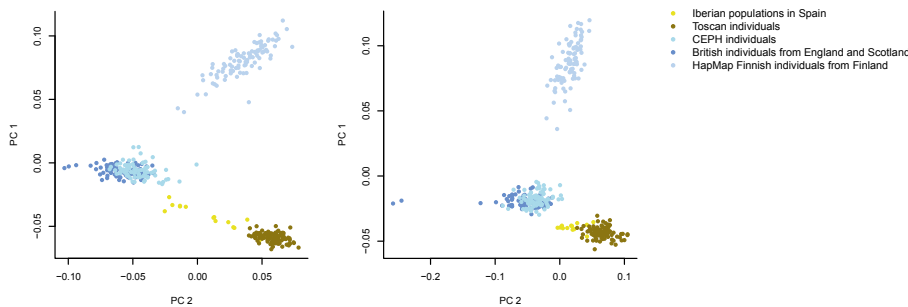


Fig. 1. The first two ancestry-informative PCs from the PC analysis (PCA) conducted on the European populations from the 1000 Genomes project using Eigenstrat (4). (Left) PCs from a PCA on 1,185,924 of the 1,250,013 ancestry SNPs that correlated significantly ($\alpha = 0.05$) with any of the first five PCs from Domingue et al. (Right) PCs from the PCA on 433,450 of 457,201 SNPs that were not individually significantly associated with any of the first five PCs, showing very little evidence of population stratification according to Domingue et al. PC1 correlates 0.97 between the two SNP sets, and PC2 correlates 0.92.

- 1 Domingue BW, Fletcher J, Conley D, Boardman JD (2014) Genetic and educational assortative mating among US adults. *Proc Natl Acad Sci USA* 111(22):7996–8000.
- 2 Bossard JH (1932) Residential propinquity as a factor in marriage selection. *Am J Sociol* 38(2):219–224.
- 3 Brittingham A, De La Cruz GP (2004) *Ancestry: 2000* (US Department of Commerce, Economics and Statistics Administration, US Census Bureau, Washington, DC).
- 4 Price AL, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909.
- 5 Abdellaoui A, et al. (2013) Population structure, migration, and diversifying selection in the Netherlands. *Eur J Hum Genet* 21(11):1277–1285.

Author contributions: A.A., K.J.H.V., and B.P.Z. designed research; A.A. analyzed data; and A.A., K.J.H.V., and B.P.Z. wrote the paper. The authors declare no conflict of interest.

¹A.A. and K.J.H.V. contributed equally to this work.

²To whom correspondence should be addressed. Email: zietsch@psy.uq.edu.au.