Using nature to understand nurture

Philipp D. Koellinger, Ph.D., Vrije Universiteit Amsterdam

K. Paige Harden, Ph.D., University of Texas at Austin

Parents transmit advantages and disadvantages to their children, perpetuating inequalities in health, wealth, and well-being across generations (*1*). Disentangling the roles that nature and nurture play in intergenerational transmission is one of the most enduring and difficult problems in social science and medicine. On page 424 of this issue, Kong and colleagues use genetic data from trios of parents and offspring to address this challenge in a novel, intriguing way. They show that parents' non-transmitted genetic alleles (i.e., the part of the parental genotype the children *didn't* inherit) can nonetheless predict children's educational attainment. They call this effect "genetic nurture" -- a indirect link between parental genotypes and children's characteristics that is not caused by the children's own biology, but is rather caused by the family environment provided by the parents (which was, in part, inherited from their parents, *et cetera*, back through a lineage). In contrast to results for educational attainment, parental alleles associated with height and BMI predict their children's bodies only insofar as the children actually inherit those genes. These results are consistent with twin studies, which also find evidence for moderate family environmental effects on educational attainment (*2*) but not height or BMI (*3*, *4*). Kong et al.'s ingenious analysis of family data reminds us, yet again, of the methodological problems that plague social scientists as we try to understand individual differences in complex human behavior, but also illustrates how understanding nature can provide us with new tools for studying nurture.

This study comes at a time when collecting genetic data has become so affordable that millions of people have already been genotyped. Scientists have studied the genetic architecture of hundreds of traits and have identified thousands of replicable genetic associations (*5*, *6*). Although any single gene tends to have only very small effects on traits such as body height or educational attainment, polygenic scores that aggregate the effects of many genes have begun to predict appreciable variation in the population.

Yet, interpreting these results remains difficult. Genetic associations may be environment-specific (*7*), and genetic effects can exert their influence via environmentally-mediated channels (*8*, *9*). For all of these reasons, it is wrong to interpret the results of genetic association studies as evidence for a biologically reductionistic account of human individual differences. Most social scientists who work with genetic data continue to stress this important qualification (*10*, *11*). Now, genetic nurture provides another compelling example of how tightly genetic and environmental influences are entangled.

At the same time, it raises an important question: If a genetic association is identified only because it correlates with an unobserved, environmental factor that is the actual cause, isn't this a statistical confound that needs to be eliminated? In one sense, yes. Geneticists have long been concerned with the problem of population stratification (*12*), which is defined as the existence of systematic differences in gene frequency between groups that might also differ in environment and culture. Population stratification can induce a genotype-phenotype correlation that is, in fact, due to the group environment. The classic example is the "chopsticks gene" (*13*) -- any gene that is more common in Asians than Europeans would come to be correlated with using chopsticks if population stratification were uncontrolled. If we substitute "family" differences for "group" differences, we have a definition of genetic nurture. Consequently, if the research goal is to identify only those genetic effects that could

be causally manipulated by changing an individual's *own* DNA, then genetic nurture is, like population stratification, a confound to be contained.

Genetic nurture also presents a new challenge for research designs, such as Mendelian Randomization (*14*), that aim to identify causal effects by using genes as naturally occurring experiments. Mendelian randomization is rapidly gaining popularity in the medical and social sciences at the moment (*15*). One of the assumptions of this approach is that the genes used for identification are not correlated with unobserved confounds, such as parental environment. Kong et al.'s study clearly shows that this assumption can be strongly violated in practice, which may yield misleading results.

Yet, genetic nurture does not undermine the value of genetic associations for prediction purposes. For example, researchers who want to use a polygenic score to obtain more precise estimates of a treatment effect in an expensive randomized controlled trial (*16*) do not care *why* the polygenic score is correlated with the outcome. All that matters is that the score is as predictive as possible.

Furthermore, for many scientists, nurture *is* the phenomenon of interest, while nature is the confound to be contained. Kong et al.'s study provides these scientists with a fascinating new tool for investigating the effects of nurture. Until now, the primary tools to disentangle the effects of a parent's genes from their actual parenting were adoption studies (*17*) and children-of-twin studies (*18*). However, collecting such data is difficult, and these naturally occurring experiments are rarely representative of the entire range of environments (*19*). Kong et al capitalize on the same logic with different data (**Figure 1**). The non-transmitted parental alleles function like the genotype of an adoptive parent, in that they help to shape the rearing environment, but they are independent from the offspring's genotype.

Datasets with genotyped trios are, unfortunately, still rare. But thanks to low genotyping costs, the trio design developed by Kong et al. could become an attractive and cost-effective research paradigm with great value to scientists who are interested in understanding the impact of family environments on human flourishing.

References

1. J. Ermisch, M. Jäntti, T. Smeeding, Eds., *From Parents to Children: The Intergenerational Transmission of Advantage* (Russell Sage Foundation, 2012; http://www.jstor.org/stable/10.7758/9781610447805).
2. A. R. Branigan, K. J. Mccallum, J. Freese, Variation in the heritability of educational attainment: An international meta-analysis. *Soc. Forces*. **92**, 109–140 (2013).
3. K. Silventoinen *et al.*, Differences in genetic and environmental variation in adult body mass index by sex , age , time period , and region : an individual-based pooled analysis of 40 twin cohorts (2017), doi:10.3945/ajcn.117.153643.Differences.
4. A. Jelenkovic *et al.*, Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts. *Sci. Rep.* **6**, 1–13 (2016).
5. P. M. Visscher *et al.*, 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
6. T. J. C. Polderman *et al.*, Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).

7. R. de Vlaming *et al.*, Meta-GWAS Accuracy and Power (MetaGAP) Calculator Shows that Hiding Heritability Is Partially Due to Imperfect Genetic Correlations across Studies. *PLOS Genet.* **13**, e1006495 (2017).

8. C. Jencks, Heredity, environment, and public policy reconsidered. *Am. Sociol. Rev.* **45**, 723–736 (1980).

9. E. M. Tucker-Drob, D. A. Briley, K. P. Harden, Genetic and environmental influences on cognition across development and context. *Curr. Dir. Psychol. Sci.* **22**, 349–355 (2013).

10. C. A. Rietveld *et al.*, GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science (80-.).* **340**, 1467–1471 (2013).

11. A. Okbay *et al.*, Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. **533**, 539–542 (2016).

12. A. L. Price *et al.*, Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

13. D. Hamer, L. Sirota, Beware the chopsticks gene. *Mol. Psychiatry*. **5**, 11–13 (2000).

14. D. A. Lawlor, R. M. Harbord, J. A. C. Sterne, N. Timpson, G. Davey Smith, Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).

15. G. Davey Smith, G. Hemani, Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89-98 (2014).

16. D. J. Benjamin *et al.*, The promises and pitfalls of genoeconomics. *Annu. Rev. Econom.* **4**, 627–662 (2012).

17. J. M. Horn, J. C. Loehlin, L. Willerman, Intellectual resemblance among adoptive and biological relatives: The texas Adoption Project. *Behav. Genet.* **9**, 177–207 (1979).

18. B. M. D'Onofrio *et al.*, The role of the children of twins design in elucidating causal relations between parent characteristics and child outcomes. *J. Child Psychol. Psychiatry Allied Discip.* **44**, 1130–1144 (2003).

19. M. Stoolmiller, Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior–genetic adoption studies. *Psychol. Bull.* **125**, 392–409 (1999).