GWAS deems parents guilty by association

Arbel Harpak[a,b,1] and Michael D. Edge[c]

Almost 150 y ago, Galton (1) pitted "nature" against "nurture." In asking whether our biological or environmental inheritance is more important in shaping our traits, Galton implicitly suggested that their effects can be separated. Galton's division has always been too neat. Our genes and environments intertwine to shape our bodies, capacities, and personalities. The "nature vs. nurture" dichotomy is particularly badly strained in interpreting gene-by-environment interactions, in which the effects of genetic variants depend on the environment in which they are expressed (2). Another challenge is gene–environment covariation, in which, for example, people's genotypes make them more likely to be exposed to particular environments (3). For example, skin tone—a genetically influenced trait—affects access to health care, socioeconomic exposures, and more (4–6).

In PNAS, Wu et al. (7) analyze a fascinating source of gene–environment covariation: Our biological parents, who are the sources of our genes, can also shape our environment. Parental effects on environments—on prenatal environment, diet, socioeconomic circumstances, home environments during childhood, and more—may in turn be influenced by the parents' genes. The alleles of parents can therefore affect a child's traits either via the effects they have when inherited in the genome ("direct genetic effects") or via their effect on the environments that parents create ("indirect parental genetic effects"; Fig. 1). Wu et al. provide new tools for studying such indirect parental genetic effects using summary data from genome-wide association studies (GWASs) (i.e., marker-level summaries that authors of GWASs typically release, even when they do not share raw data) and use them to study genetic effects on birth weight and on educational attainment.

Indirect genetic effects—which need not be parental and can in principle occur between any pair of individuals, related or unrelated—have been studied by quantitative geneticists since at least the 1940s (8–11). Indirect genetic effects are a sometimes-crucial influence on the evolution of trait variation (12), with special relevance for the evolution of social behavior (13). Recently, Kong et al. (14) energized the study of indirect genetic effects in humans with analyses showing their importance for several human traits. Kong et al.'s approach was pathbreaking in its consideration of indirect genetic effects in human GWASs. However, their approach required genotype and phenotype data from members of the same families; and although family-level data are experiencing a blissful resurgence in human genetics (15), they are still not as common as standard GWAS samples of unrelated people. Wu et al. introduce an approach based on standard GWASs, which they call "DONUTS," for estimating indirect parental genetic effects.

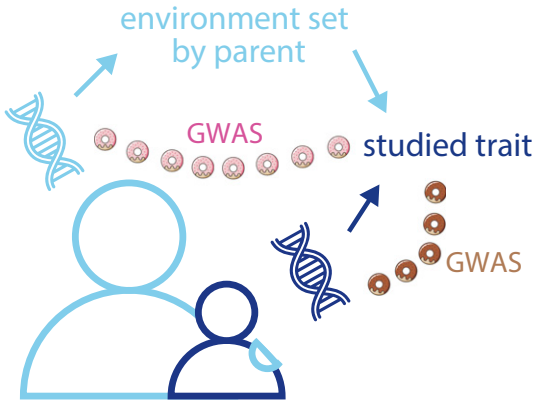

Fig. 1. Alleles of biological parents can affect a trait of their child either when inherited and expressed in the child's genome ("direct genetic effects"; dark blue arrow) or when they are expressed in a parent and modify environmental effects on the child's trait ("indirect parental genetic effects"; light blue arrows). DONUTS (7) uses summary statistics from two GWASs to estimate direct and indirect effects separately: a standard GWAS of genotype and phenotype of the same individuals (denoted by the string of chocolate donuts) and a GWAS of a parent's genotype and their child's phenotype (strawberry donuts).

[a]Department of Population Health, University of Texas at Austin, Austin, TX 78712; [b]Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712; and [c]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089

Author contributions: A.H. and M.D.E. wrote the paper.

The authors declare no competing interest.

See companion article, "Estimating genetic nurture with summary statistics of multigenerational genome-wide association studies," 10.1073/pnas.2023184118.

[1]To whom correspondence may be addressed. Email: arbelharpak@utexas.edu.

Published June 30, 2021.

The dissection of genetic effects into direct and indirect is central to several ongoing conversations in human genetics. Most fundamentally, in understanding how trait differences between people emerge, direct vs. indirect effects suggest distinct sets of mechanisms. Secondly—and even in settings where direct effects are of primary interest, such as in some clinical applications of polygenic scores or in more controversial applications to embryo selection (16)—problems arise if estimated effects are assumed to be direct when they are in fact partly indirect. Thirdly, distinguishing direct and indirect effects may be important for the "portability" problem: Why polygenic scores based on GWASs tend to drop in prediction accuracy in groups that differ in their composition—genetic ancestry, socioeconomic, or otherwise—from the GWAS sample (17, 18). Although the various factors hypothesized to affect portability are relevant for both direct and indirect effects, the fact that indirect effects are, by definition, environmentally mediated—and for some traits plausibly culture-bound—may in itself suggest they would port poorly across groups of people (18).

To understand how indirect genetic effects are estimated, it is helpful to think about the sources of GWAS effect estimates (19). In a standard GWAS, participants' genotypes are associated with their phenotypes at hundreds of thousands of markers. If an association is observed between a given marker and a given phenotype, there are several possible explanations. First, there may be one or more genetic variants nearby in the genome that directly affect the phenotype. [Here, "direct" means only that the variant affects an individual's phenotype because the individual inherits it. Direct effects in this sense might nonetheless be mediated via biological, environmental, and social factors. For example, some of the variants most strongly associated with lung cancer risk appear to operate via their effects on smoking behavior (see box 1 in ref. 20).] This is the main signal that GWASs were designed to capture. Second, the marker may be near variants that have familial indirect effects on the genotype. Because the genotypes of GWAS participants are correlated with those of their biological relatives, indirect effects will lead to an association between the genotype and phenotype of participants. Additional sources of association signals are genetic or environmental confounding—as might be caused, for example, by population stratification or assortative mating (21)—and sampling error. Statistical approaches aim to remove confounding—sometimes with less than complete success (22–24)—and quantify sampling error. Among these sources, direct and indirect effects are of greatest interest, but a standard GWAS alone does not distinguish them from each other.

Methods for estimating direct effects and parental indirect genetic effects separately (14, 25–28) take advantage of the fact that the allele that a parent transmits to a child has both a direct and indirect effect on the child's phenotype, whereas the allele that the parent does not transmit has only an indirect effect. Thus, if genotypes are available for parents and a child, as well as phenotype data from the child, direct and indirect effects can be distinguished by comparing the association of the parents' transmitted and nontransmitted alleles with the child's phenotype.

DONUTS relies on a clever approach that, under some assumptions, can perform the same estimation using only summary statistics from two or three distinct GWASs, for example a GWAS associating one's own genotypes with their phenotype, and a second GWAS associating parental genotype with their child's phenotype (Fig. 1). DONUTS also accommodates complications such as assortative mating in the parental generation, sample overlap in the GWASs that are sources of the summary statistics, and indirect effects that differ by parent of origin.

In PNAS, Wu et al. analyze a fascinating source of gene–environment covariation: Our biological parents, who are the sources of our genes, can also shape our environment.

Although samples of parental genotypes alongside their children's phenotypes are currently uncommon, they are likely less burdensome to collect than full family data. Wu et al. therefore expand the possibilities for studying direct and indirect genetic effects in humans.

With the example of birth weight, Wu et al. show that DONUTS' estimates of indirect parental effects agree closely with those from methods using individual-level data (26). They also use their framework to explore previously observed genetic correlations between educational attainment and other traits. Educational attainment is of special interest in part because of prior evidence of sizeable parental indirect effects (14, 25). Wu et al. show that many of the known genetic correlations with educational attainment that they examine—including with taller height, lower body mass index, less active smoking behavior, and better health outcomes—are mostly explained by indirect effects on educational attainment. In turn, the genetic correlation between educational attainment and autism spectrum disorder is mostly attributable to direct effects on educational attainment.

DONUTS extends the study of indirect genetic effects to settings closer to traditional GWASs, and at the same time Wu et al.'s empirical findings highlight the value of family-based designs. Direct and indirect genetic components, as well as indirect genetic components from distinct relatives—for example, mothers and fathers—can in principle behave differently, suggesting distinct explanations for observed genetic correlations involving overlapping sets of traits. This information is inaccessible in a typical GWAS, which confounds direct and indirect effects.

As Wu et al. carefully note, systematic biases like those due to population stratification or assortative mating remain a threat for any GWAS with "unrelated" participants. Statistical methods have been successful in alleviating the bias at any given marker, but the threat is aggravated when signals from many markers are summed in the same analysis, such as when working with polygenic risk scores (18, 24). Family-based designs can eliminate such confounding from population stratification and assortative mating. For example, if genetic differences between full biological siblings are associated with differences in phenotype, then stratification is not a possible explanation: Such genetic differences are caused by random Mendelian segregation alone, which is not plausibly confounded by other causal factors.

The main drawback of such family designs relates to statistical power. The largest samples available are composed of sibling pairs and remain an order of magnitude smaller than standard GWAS samples of unrelated individuals in the biobank era [although see a recent advance toward closing this gap by Howe et al. (15)]. For this and other reasons, such as lower genetic variation within families than across unrelated individuals, family-based GWASs are likely to remain less powerful than standard GWASs for the foreseeable future (15, 18).

Approaches like those of Wu et al. thus extend the scope of GWASs for understanding complex trait variation. Human geneticists struggle to address two features of genetic associations with complex traits: on the one hand, small marginal effect sizes that are only detectable with large association studies, and on the other, a subtle interplay of genetics and environment, calling for more informative designs. DONUTS represents a step toward marrying the power of standard GWASs with the articulation and control of family designs.

Acknowledgments

We thank Frances Champagne, Hans Hofmann, Tom Juenger, Mark Kirkpatrick, and Molly Przeworski for comments and helpful discussions. M.D.E. acknowledges support from NIH Grant R35 GM137758.

1 F. Galton, *English Men of Science: Their Nature and Nurture* (Macmillan and Company, 1875), vol. 27.
2 J. Tabery, R. A. Fisher, Lancelot Hogben, and the origin(s) of genotype-environment interaction. *J. Hist. Biol.* **41**, 717–761 (2008).
3 D. M. Dick, Gene-environment correlation. Wiley StatsRef: Statistics Reference Online (2014). https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat06727. Accessed 1 June 2021.
4 E. P. Monk, Jr, Skin tone stratification among Black Americans, 2001–2003. *Soc. Forces* **92**, 1313–1337 (2014).
5 A. H. Goldsmith, D. Hamilton, W. Darity, Jr, Shades of discrimination: Skin tone and wages. *Am. Econ. Rev.* **96**, 242–245 (2006).
6 E. Telles, *Pigmentocracies: Ethnicity, Race, and Color in Latin America* (UNC Press Books, 2014).
7 Y. Wu *et al.*, Estimating genetic nurture with summary statistics of multigenerational genome-wide association studies. *Proc. Natl. Acad. Sci. U.S.A.*, 10.1073/pnas.2023184118 (2021).
8 G. E. Dickerson, Composition of hog carcasses as influenced by heritable differences in rate and economy of gain. *Iowa Agric. Home Econ. Exp. Stn. Res. Bull.* **28**, 1 (1947).
9 R. L. Willham, The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* **19**, 18–27 (1963).
10 D. S. Falconer, Maternal effects and selection response. *Genet. Today* **3**, 774 (1965).
11 B. Griffing, Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Aust. J. Biol. Sci.* **20**, 127–139 (1967).
12 J. B. Wolf, E. D. Brodie Iii, J. M. Cheverud, A. J. Moore, M. J. Wade, Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.* **13**, 64–69 (1998).
13 J. Schneider, J. Atallah, J. D. Levine, Social structure and indirect genetic effects: Genetics of social behaviour. *Biol. Rev. Camb. Philos. Soc.* **92**, 1027–1038 (2017).
14 A. Kong *et al.*, The nature of nurture: Effects of parental genotypes. *Science* **359**, 424–428 (2018).
15 L. J. Howe *et al.*, Within-sibship GWAS improve estimates of direct genetic effects. *BioRxiv* [Preprint] (2021). https://www.biorxiv.org/content/10.1101/2021.03.05.433935v1 (Accessed 1 June 2021).
16 E. Karavani *et al.*, Screening human embryos for polygenic traits has limited utility. *Cell* **179**, 1424–1435.e8 (2019).
17 A. R. Martin *et al.*, Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
18 H. Mostafavi *et al.*, Variable prediction accuracy of polygenic scores within an ancestry group. *eLife* **9**, e48376 (2020).
19 A. I. Young, S. Benonisdottir, M. Przeworski, A. Kong, Deconstructing the sources of genotype-phenotype associations in humans. *Science* **365**, 1396–1400 (2019).
20 S. H. Gage, G. Davey Smith, J. J. Ware, J. Flint, M. R. G. Munafò, G = E: What GWAS can tell us about the environment. *PLoS Genet.* **12**, e1005765 (2016).
21 B. J. Vilhjálmsson, M. Nordborg, The nature of confounding in genome-wide association studies. *Nat. Rev. Genet.* **14**, 1–2 (2013).
22 J. J. Berg *et al.*, Reduced signal for polygenic adaptation of height in UK Biobank. *eLife* **8**, e39725 (2019).
23 M. Sohail *et al.*, Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* **8**, e39702 (2019).
24 N. Barton, J. Hermisson, M. Nordborg, Why structure matters. *eLife* **8**, e45380 (2019).
25 S. Selzam *et al.*, Comparing within- and between-family polygenic score prediction. *Am. J. Hum. Genet.* **105**, 351–363 (2019).
26 N. M. Warrington *et al.*; EGG Consortium, Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).
27 A. I. Young *et al.*, Mendelian imputation of parental genotypes for genome-wide estimation of direct and indirect genetic effects. *BioRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.07.02.185199 (Accessed 25 June 2021).
28 S. Trejo, B. W. Domingue, Genetic nature or genetic nurture? Quantifying bias in analyses using polygenic scores. *BioRxiv* [Preprint] (2019). https://doi.org/10.1101/524850 (Accessed 25 June 2021).