GENETICS

Potential bias in genetic correlations

Mating patterns across two traits can inflate estimates of genetic overlap

By **Andrew D. Grotzinger**[1,2] *and* **Matthew C. Keller**[1,2]

Genome-wide association studies (GWASs) identify genetic variants associated with a trait. Most traits are associated with thousands of variants, and many variants are pleiotropic, meaning they are associated with multiple traits. Pervasive pleiotropy makes it impractical to assess genetic overlap between two traits by tallying the shared variants. For example, traits such as major depression and anxiety are likely associated with a shared set of thousands of variants. Genetic correlation (r_g) estimated from GWASs of a pair of traits is typically interpreted as an overall measure of genetic overlap, providing a useful metric for quantifying shared biology between traits. On page 754 of this issue, Border *et al.* (*1*) report simulation-based and empirical findings that challenge this interpretation.

Border *et al.* investigate the effect on r_g of cross-trait assortative mating (xAM), which occurs when individuals scoring highly for trait Y mate with partners who score high (or low) for a separate trait, Z. The level of xAM will vary depending on the pair of traits being considered and is quantified in their study using cross-trait correlations across spouses. They show that xAM can increase r_g in two ways (see the figure). They demonstrate that xAM will increase r_g due to genetic variants for trait Y being coinherited with the variants for trait Z (and vice versa). This first potential effect of xAM on r_g estimates has been understood for decades as an interpretive caveat in family-based studies (*2*–*4*), although it is perhaps not widely discussed. The authors point out that this first effect of xAM does not cause an actual bias in r_g estimates because the polygenic scores—the cumulative set of genetic variants that affect each trait—are legitimately correlated. This is despite there being no pleiotropic genetic variants that affect both traits. If this were the only issue, the interpretation of r_g as strictly an index of shared biology would require updating, but r_g would still be a valuable metric with etiological and clinical implications for understanding risk for trait Y given trait Z.

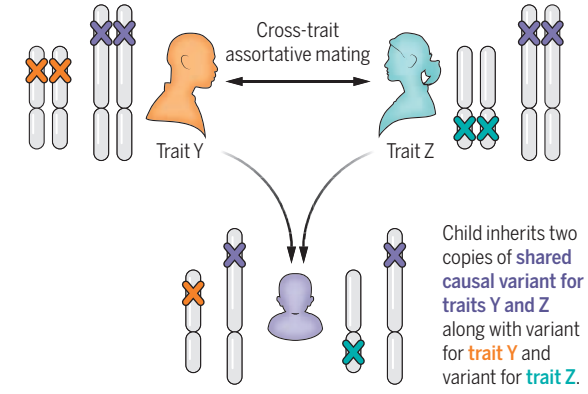
The second effect that xAM has on r_g is potentially more problematic. Commonly used methods for estimating r_g, such as linkage disequilibrium (LD) score regression (*5*), Haseman-Elston regression (*6*), and genomic residual maximum likelihood (REML) (*7*), assume that there is no long-range correlation across causal variants. Border *et al.* find that xAM violates this assumption by inducing long-range correlations between causal variants that are on average positive or negative (i.e., sign-consistent). This causes an upward bias in r_g estimates after even a single generation of xAM. Unlike in the first case where only the interpretation of r_g would shift, this second issue causes bias in the true sense of the term: The inflation of r_g is a statistical artifact that distorts the correlation between polygenic scores. Somewhat reassuringly, the first effect of xAM on r_g will typically be greater than the influence of the second effect, particularly when genomic REML is used.

In the worst-case scenario, xAM for two traits with no overlapping causal variants will produce r_g estimates that are not due to pleiotropy and are upwardly biased. Border *et al.* use simulation findings and empirical r_g and cross-mate, cross-trait correlation estimates to show that, for example, r_g estimates between alcohol use disorder–schizophrenia or height–educational attainment may fall into this worst-case scenario. However, for other pairs of traits, the r_g estimates are too high relative to the estimated levels of xAM to realistically reflect a by-product of xAM alone. Within psychiatric phenotypes, these trait pairs with large r_g estimates include bipolar disorder–schizophrenia and major depressive disorder–anxiety. In addition, estimates of r_g between traits that show little or no xAM, such as certain diseases or late-onset disorders, are still likely to reflect pleiotropy. Even for traits that do show evidence for xAM, r_g will typically reflect some combination of three alternative contributions—pleiotropy, correlated polygenic scores not due to pleiotropy, and bias due to long-range correlation. Therefore, the findings of Border *et al.* should not be interpreted as implying that r_g estimates are wholly unrelated to pleiotropy.

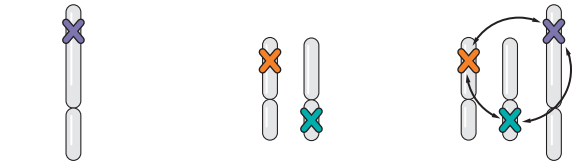
Their results for case-control traits, which compare individuals with a disorder (a case) to those who do not have the disorder (a control), should also be interpreted with caution. As is standard, Border *et al.* quantify xAM between case-control traits using tetrachoric correlations, which estimate the correlation between the continuous risk scores assumed to underlie two disorders. Because patterns of xAM across all levels of genetic risk affect r_g, this approach assumes that the same degree of xAM inferred from mating patterns for cases with extreme risk holds across all levels of risk,

Measuring genetic correlations

Cross-trait assortative mating could result in genetic correlations with certain traits that are explained by three scenarios, two of which inflate the score, r_g, upward even in the absence of pleiotropy.



If assortative mating occurs in the population, the measure of genetic correlation, r_g, would reflect three scenarios:



Shared causal variant reflects typical pleiotropy interpretation of r_g.

Separate causal variants are inherited together. r_g increases owing to correlated genetic scores but not pleiotropy.

Upward bias in r_g score owing to long-range correlation between variants.

[1]Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO, USA. [2]Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, CO, USA. Email: andrew.grotzinger@colorado.edu

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

including for mates who are "subthreshold" for a disorder. However, increased mating between individuals who receive a diagnosis may occur for reasons, such as an elevated chance of meeting a partner in a clinical setting, that have no bearing on mating that occurs between individuals with subthreshold risk. Thus, tetrachoric correlations between disorders, and perhaps especially between psychiatric disorders, may overstate the influence of xAM on r_g estimates.

The findings of Border *et al.* make it clear that more realistic models for why mates correlate within and between multiple traits need to be developed and tested. It may be that xAM occurs only on a few traits of central importance to mating and that most other mate correlations are consequences of the phenotypic and genetic correlations between those central traits and the traits that are actually measured (indirect AM). It is also possible that some xAM estimates are due to mates becoming more similar over time (convergence). Assortment can also reflect social homogamy, which occurs when mate choice is based on environmental components of the trait that are not due to genetics. This could occur, for example, if mates are chosen on the basis of religious beliefs and religion is both due to the environment and affects the traits being studied. Convergence and social homogamy are not expected to affect r_g estimates. Deriving more complete models of the mechanisms through which observed levels of xAM manifest will be important for obtaining a better understanding of downstream effects on r_g.

Border *et al.* demonstrate that xAM is likely to be pervasive and affect many r_g estimates to some degree. This bias will inherently carry forward to results from any methods that use genetic correlations as input, such as Mendelian randomization (*8*) or genomic structural equation modeling (*9*). Therefore, complex trait genetics ignores these problems at its peril. These issues can be addressed by increased care in interpreting r_g as well as through the development of methods that can disentangle the various contributions to r_g estimates. ■

REFERENCES AND NOTES

1. R. Border *et al.*, *Science* **378**, 754 (2022).
2. L. J. Eaves, A. C. Heath, N. G. Martin, *Behav. Genet.* **14**, 371 (1984).
3. M. C. Keller *et al.*, *PLOS Genet.* **9**, e1003451 (2013).
4. G. P. Vogler, J. C. DeFries, *Behav. Genet.* **15**, 111 (1985).
5. B. Bulik-Sullivan *et al.*, *Nat. Genet.* **47**, 1236 (2015).
6. J. K. Haseman, R. C. Elston, *Behav. Genet.* **2**, 3 (1972).
7. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, *Am. J. Hum. Genet.* **88**, 76 (2011).
8. F. P. Hartwig, N. M. Davies, G. Davey Smith, *Genet. Epidemiol.* **42**, 608 (2018).
9. A. D. Grotzinger *et al.*, *Nat. Hum. Behav.* **3**, 513 (2019).

10.1126/science.ade8002

ORGANIC CHEMISTRY

Adding functions to pyridines

Chemical reactions break a pyridine ring to allow its modification

By **Jung Min Joo**

Pyridine (C_5H_5N) is a launch point for creating a wide range of chemicals, including those used in drug discovery, catalysis, and materials science (*1*). It consists of a hexagonal ring of five carbon-hydrogen (C–H) pairs and one nitrogen (N) atom. The synthesis of pyridine derivatives bearing substituents other than hydrogen requires custom-designed processes. Among them, substitutions of one substituent for another are frequently performed while maintaining the integrity of the ring. However, the efficiency of these methods and the diversity of pyridine compounds that they yield need improvement (*2*). On pages 773 and 779 of this issue, Boyle *et al.* (*3*) and Cao *et al.* (*4*), respectively, report different approaches to breaking a pyridine ring, replacing the hydrogens, and then restoring the ring. These techniques present possible new pathways to generate a wider variety of agrochemicals, pharmaceuticals, and material compounds.

"...new pathways... generate a wider variety of agrochemicals, pharmaceuticals, and material compounds."

The effects of substituents that replace the hydrogens in a pyridine is an important topic in organic chemistry. Halogenation—the replacement of these hydrogens with a halogen atom—is useful as an intermediate step for making pyridine derivatives because the installed halogen can be much more readily substituted by another atom or group of atoms as compared with hydrogen (*5*). In addition, the halogen substituents can facilitate intermolecular interactions that are essential for certain desired functions such as binding to target proteins (*6*).

Halogenation of pyridines can be accomplished through their reaction with electrophilic chemicals (*7*). These electrophiles are electron poor and can form chemical bonds by accepting an electron pair from electron-rich chemicals. However, because the nitrogen atom attracts electrons from carbon atoms of the ring, the carbon atoms themselves are electron poor and thus do not react easily with electrophiles. Furthermore, the nitrogen atom reacts preferentially with electrophilic reagents to form the corresponding pyridinium salts, which are more electron poor than pyridines and thereby less likely to react with electrophilic halogen species. As a result, halogenation requires harsh reaction conditions, usually involving strong acids and elevated temperatures, to generate very strong electrophiles that can enable the reaction. This limits the range of reactants that can be used and the products that can be created.

Positional selectivity is another important consideration in preparing pyridine derivatives. Although the most relatively electron-rich positions on the ring (positions 3 and 5) allow electrophilic substitution, it is difficult to control which of these positions is substituted or whether one or both of the positions are substituted. In addition, some substituents could alter the electronic and structural characteristics of the pyridine ring in ways that enable chemical functions but may decrease the positional selectivity of the reaction on the ring (*7*). There are also certain products that require the pyridine to react while positioned alongside other aromatic rings (such as those with delocalized electrons) that would compete with the aromatic pyridine ring for halogenation.

Different strategies addressing the halogenation of pyridines present advantages and disadvantages. For example, the use of bases instead of acids to mediate halogenation tends to generate more reactive intermediates (*8*). Some indirect methods use more reactive and accessible intermediates to avoid the use of strong electrophiles (*9, 10*). However, the requirements of these other strategies can limit the range of possible substituents. Thus, more versatile means are needed to prepare halogenated pyridines from simple pyridines or complex substituted pyridines.

Boyle *et al.* developed a strategy that transforms the electron-deficient pyridine into an electron-rich molecule, enabling it to react with electrophilic halogenat-

Department of Chemistry and Chemistry Institute for Functional Materials, Pusan National University, Busan 46241, South Korea. Email: jmjoo@pusan.ac.kr

Potential bias in genetic correlations

Andrew D. GrotzingerMatthew C. Keller

Science, 378 (6621), • DOI: 10.1126/science.ade8002