

Inviting in the Exome for Alcohol and Smoking Traits

Joel Gelernter

In this issue of *Biological Psychiatry*, Brazel *et al.* (1) report the first large meta-analysis of array-based exome genotype data including rare variant (RV) content for phenotypes related to smoking and alcohol use. While there have been numerous genome-wide association studies (GWASs) of these traits and of related traits, the custom—at least for smaller studies up to this point—has been to sweep RVs under the rug (where they formed only a very small pile) rather than take notice of them. This has been a necessity, for the most part. Rare variants are, after all, rare. A variant with a minor allele frequency (MAF) of 0.1% would be expected to be observed in about 20 individuals in a sample of 10,000 individuals, for example; not enough total observations to attain the power needed for any conclusion, unless the variant had a large effect on the phenotype. For phenotypes distributed in the population like alcohol use and smoking—that is, common phenotypes with wide observed ranges of values—this is, practically speaking, difficult to establish with a small number of observations. Genotyping microarrays with RV content, including arrays designed specifically to include exomic RV content, are available from both major genotyping microarray suppliers, Illumina (San Diego, CA) and Affymetrix (Santa Clara, CA), and have been popular for years now. So while many investigators have acquired RV (and other exome-wide) genotype information over the years, they have not been able to do much with this information. Many GWASs have used MAF cutoffs in the range of 0.01 to 0.03. That is, RV content (allele frequency <1%), directly genotyped and imputed, is typically removed before analysis. There are two obvious solutions to this problem—either acquire the data in a study sample large enough for RV content to be analyzable, or combine many studies with such data meta-analytically. Brazel *et al.* (1) do the latter while making great use of data from the UK Biobank, which does the former. This follows on their recent article considering common variant content for the same traits (2). Depending on the phenotype, this study includes roughly between 150,000 and 430,000 subjects—the UK Biobank provides a large majority of analyzable subjects, from more than 70% to nearly 80%, depending on phenotype. It is a recent development in the field that such huge GWAS samples are no longer especially unusual.

Brazel *et al.* (1) confirm several well-known risk loci for smoking and alcohol use traits with high statistical significance. The strongest findings by far implicate *ADH1B**rs1229984 for drinks per week and *CHRNA5**rs16969968 for cigarettes per week—neither of which is actually “rare.” Further down the list of associated genes, there is much that is novel, including significant associations mapped to *NAV2* for smoking initiation and *SERPINA1* protective with respect to drinks per week. This

is fodder for future studies, for example, evaluating the biology of these variants in animal models.

Brazel *et al.* (1) also report that RVs with MAF <1% explain 11% to 18% of the observed single nucleotide polymorphism heritability. There has been considerable debate about the extent to which rare variation will explain what is commonly termed “missing heritability,” or the difference between heritability as determined by genetic epidemiology studies and summing up the heritability explained by common variants identified in GWASs. RVs in the article by Brazel *et al.* (1) provide a sizable increment in the heritability that can be accounted for. Recent work (3) based on whole genome sequencing (WGS) shows that the “missing heritability” for height and body mass index can be virtually completely accounted for by RVs, in particular those in regions of low linkage disequilibrium. There are several key differences between the approach taken by the authors of that article and Brazel *et al.* (1). First, the traits are different; the genetic architecture of height and body mass index may differ substantially from the genetic architecture of traits related to smoking and drinking. Smoking and drinking, unlike height and body mass index, have a key pharmacogenomic component. Second, Wainschein *et al.* (3) considered genome-wide data, whereas Brazel *et al.* (1) considered exome-wide data. Third, the WGS study considered sequence data, whereas Brazel *et al.* (1) considered microarray data. Microarray data are only useful for genotyping previously observed variants that have been seen enough times either to make it onto a microarray or to be imputable. New mutations can never be captured by array data. Sample sizes like those obtained by Brazel *et al.* (1) are presently infeasible to undergo WGS because of its expense and the needed computational bandwidth. But the increase in heritability explained by means of exomic rare variation, considered in the context of the greater increase for other complex traits using WGS data, suggests that datasets with much more genomic detail may allow substantial additional progress in explaining the genetic basis of these traits quite fully.

This is a meta-analysis, and the meta-analysis format requires compromise in terms of genetic variants that are included and the phenotypes that are studied. The analyzed phenotypes must be reasonably consistent across the analyzed constituent samples. For some smoking and alcohol use phenotypes, such phenotypes can be found quite readily. That is not the case for other phenotypes of urgent interest for public health, such as opioid use disorder (much lower prevalence than alcohol and tobacco use and less well captured in biobank data) and other illegal drug dependencies; nor is it the case for alcohol use disorder (AUD) per se, which is less likely to be collected for

SEE CORRESPONDING ARTICLE ON PAGE 946

biobank use or for studies based on other traits, such as cardiac phenotypes or cancer, where basic alcohol use data are quite likely to be collected. Well-powered studies of traits such as illegal substance dependencies require larger samples than can be assembled at present, even in meta-analysis; this will require more directed recruitment of affected subjects. Measures of quantity and frequency of substance use (such as alcoholic drinks per week) are easier to obtain in large samples. Such traits have great medical importance but are not identical to dependence traits. Recent work has shown that alcohol quantity/frequency measures are quite different genetically from AUD (4,5).

This study shares a limitation with almost all meta-analytic studies of psychiatric traits: the study sample is almost entirely of patients of European ancestry (<9000 African ancestry subjects are included). Populations—e.g., European, African, Asian, Native American, and others—can differ in genetic risk for many traits. Famously for alcohol use phenotypes, European, African, and Asian subjects share a major overall risk mechanism: variation in the genes encoding alcohol-metabolizing enzymes. For all three populations, *ADH1B* is a major risk gene. Europeans and Asians share their major risk variant [also identified in the Brazel *et al.* (1) study], but Africans have a completely different major risk variant, absent in these other populations. And Asians also have a major large effect risk variant that maps to *ALDH2* that is absent in Europeans and Africans (6). RVs tend to be even more population specific than common variants. In focusing exclusively on subjects of European ancestry, we restrict our discovery to a subset—possibly a small subset—of the genetic variation and the genetic risk that is present worldwide. Medication development based on genetic findings may be specific to the discovery population; the same goes for risk prediction (7). Further, post-GWAS analyses that rely on the comparison of GWAS results with publicly available GWASs of other traits [like linkage disequilibrium score regression (8) used by Brazel *et al.* (1)] and transcriptomics depend on those comparison datasets being available in similar populations. Accordingly, even when GWASs of substantial non-European populations are published, post-GWAS analysis and biological understanding is limited by the lack of availability of these other resources. It is our obligation as a research community to address this lack. While only a few findings emerged from the African ancestry portion of the sample studied (1), the publication of this GWAS information will allow for progress in understanding alcohol use and smoking in this population, and in identifying pleiotropy with other traits studied by future investigators.

In large biobanks with appropriate genotyping and imputation, the study of RV association comes into play naturally, and indeed in a recent analysis of alcohol use disorder and AUDIT-C (a quantity-frequency measure of alcohol use), the MAF cutoff for analysis was 0.0005 in European-Americans (5). (Numerous RV associations were identified, most near *ADH1B*.) This was feasible because the sample, from the Million Veteran Program, was large (>270,000 subjects) and was genotyped uniformly. With the advent of samples like the Million Veteran Program and the UK Biobank, we may expect to gain greatly in our understanding of the role that (array-genotypable or imputable) RVs play in risk for complex genetic traits in future. Relatively few people in a population are carriers with respect to particular rare

variants, but RVs are still important because they often have a greater effect on disease risk than common variants; and they can also be valuable indicators of disease pathophysiology. The classic example is RVs at the *PCSK9* locus (9,10) implicating a mechanism involving its protein product in lipid metabolism leading to a new treatment target for modulating low-density lipoprotein levels. Novel findings in the present study could similarly point to new understandings of the biology of cigarette smoking and habitual drinking behaviors. And in future, we will all be paying careful attention to the role of RVs in complex psychiatric traits.

Acknowledgments and Disclosures

This work was supported by National Institute on Drug Abuse Grant No. R01DA012690 and National Institute on Alcohol Abuse and Alcoholism Grant No. R01AA026364.

I thank Renato Polimanti for his helpful comments.

JG is named as an inventor on the Patent Cooperation Treaty patent application #15/878,640, filed January 24, 2018 and entitled “Genotype-guided dosing of opioid agonists.”

Article Information

From the Yale University School of Medicine and the Veterans Affairs Connecticut Healthcare Center, West Haven, Connecticut.

Address correspondence to Joel Gelernter, M.D., Yale University School of Medicine, Veterans Affairs Connecticut Healthcare Center, 950 Campbell Ave, West Haven, CT 06516; E-mail: joel.gelernter@yale.edu.

Received Apr 3, 2019; accepted Apr 5, 2019.

References

1. Brazel DM, Jiang Y, Hughey JM, Turcot V, Zhan X, Gong J, *et al.* (2019): Exome chip meta-analysis fine maps causal variants and elucidates the genetic architecture of rare coding variants in smoking and alcohol use. *Biol Psychiatry* 85:946–955.
2. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, *et al.* (2019): Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 51:237–244.
3. Wainschein P, Jain DP, Yengo L, Zheng Z, Cupples LA, Shadyab AH, *et al.* (2019): Recovery of trait heritability from whole genome sequence data [published online ahead of print Mar 25]. *bioRxiv*.
4. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, 23andMe Research Team, the Substance Use Disorder Working Group of the Psychiatric Genomics Consortium, *et al.* (2019): Genome-wide association study meta-analysis of the Alcohol Use Disorder Identification Test (AUDIT) in two population-based cohorts. *Am J Psychiatry* 176:107–118.
5. Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, *et al.* (2019): Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat Commun* 10:1499.
6. Edenberg HJ, Gelernter J, Agrawal A (2019): Genetics of alcoholism. *Curr Psychiatry Rep* 21:26.
7. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, Freedman BI, *et al.* (2010): Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* 329:841–845.
8. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, *et al.* (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47:291–295.
9. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, Devillers M, *et al.* (2003): Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 34:154–156.
10. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH (2005): Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* 37:161–165.