Phenotype and Environment Matter: Discovering the Genetic and Epigenetic Architecture of Alcohol Use Disorders

Jonathan D. Pollock, Ph.D., Amy C. Lossie, Ph.D.

Alcohol use disorder (AUD) is a chronic, relapsing brain disease characterized by periods of compulsive alcohol use in which one experiences a loss of control over drinking and a negative emotional state when not drinking. According to the National Institute on Alcohol Abuse and Alcoholism, approximately 6% of adults in the United States meet the criteria for AUD, which demonstrates a heritability of approximately 50% (1). Genome-wide association studies (GWASs) have been productive in discovering hundreds of variants associated with AUD (1), including single-nucleotide polymorphisms (SNPs) associated with ethanol metabolizing genes. However, it has been difficult to identify specific genes that contribute major effects to the phenotype, and many questions remain about the genetic architecture, the missing heritability, and the role that gene-by-environment interactions play in complex genetic conditions such as AUD.

Genetic Architecture Studies

A study by Sanchez-Roige et al. (2), in this issue of the *Journal*, along with previously published work by Walters et al. (3), teases out differences in the genetic architecture of alcohol consumption compared with problem drinking. One way to parse out these differences is to conduct genetic correlation studies, which are statistical analyses that are used to determine whether phenotypes or traits share the same genes. These studies show that the problematic consequences of drinking (2) and DSM-IV diagnoses (3) are polygenic and are genetically correlated with schizophrenia, major depressive disorder, and attention deficit hyperactivity disorder (ADHD), while consumption is not. In addition, their analyses suggest that higher alcohol consumption and the problematic consequences of drinking are influenced by numerous genetic variants that partially overlap with those that influence the risk to develop alcohol dependence as defined by DSM-IV.

Sanchez-Roige et al. (2) conducted a GWAS meta-analysis of the Alcohol Use Disorders Identification Test (AUDIT) that investigated population-based samples from the UK Biobank (N=121,604) and 23andMe (N=20,328). AUDIT is a screening tool designed to identify hazardous alcohol use in the past year. AUDIT consists of 10 items (AUDIT-T), which Sanchez-Roige et al. parsed into two dimensions, one pertaining to alcohol consumption (items 1–3, AUDIT-C) and another pertaining to dependence symptoms and harmful alcohol use (items 4–10, AUDIT-P). Genetic analyses of AUDIT-T and AUDIT-C revealed associations with variants proximal to *ADH1B* (alcohol dehydrogenase 1B), *ADH1C* (alcohol dehydrogenase 1C), *KLB* (Klotho beta), and *GCKR* (glucokinase regulatory protein) (2), which replicated prior studies of alcohol consumption. These analyses also identified new associations at variants proximal to *JCAD* (junctional cadherins associated with coronary artery disease) and *SLC39A13* (zinc transporter *ZIP13*). As expected, AUDIT-C was more strongly genetically correlated with alcohol consumption (r_g=0.92) compared with AUDIT-P (r_g=0.76). Importantly, AUDIT-T and AUDIT-C were only moderately genetically associated with DSM-IV alcohol dependence (r_g=0.39 and r_g=0.33, respectively). However, AUDIT-P showed stronger genetic association with alcohol dependence (r_g=0.63) compared with AUDIT-T or AUDIT-C. Walters et al. (3) conducted a meta-analysis of 14,904 individuals with AUD and 37,944 control subjects, which also identified variants proximal to *ADH1B* and *ADH1C*. They also reported a genetic correlation between DSM-IV alcohol dependence and alcohol consumption, which was strong in the Cohorts for Heart and Aging Research in Genomic Epidemiology Plus Consortia (r_g = 0.69) but weaker in the UK Biobank (r_g = 0.37).

Many questions remain about the genetic architecture, the missing heritability, and the role that gene-by-environment interactions play.

The findings from these genetic correlations may reflect ascertainment bias. The UK Biobank cohort is a population-based sample in which only five percent of individuals agreed to participate, and only a small number of these exhibited problem drinking. The 23andMe sample is self-selected and reflects a group of individuals with higher levels of education and socioeconomic status. Additionally, this sample probably has a low percentage of individuals who meet criteria for AUD. In contrast, ascertainment in the Walters et al. study was (3) on DSM-IV diagnosis.

The difference in the strengths of genetic correlation between consumption and DSM-IV diagnosis emphasizes the importance of phenotype and highlights that alcohol consumption and problematic drinking are partially genetically distinct. Gene-based analysis of AUDIT-P scores conducted by Sanchez-Roige et al. (2) showed that the dopamine receptor *DRD2*, *CADM2* (an adhesion molecule also associated with lifetime cannabis use and other phenotypes associated with risky behavior), and *KLB* (Klotho beta, the FGF21 coreceptor) are associated with problem drinking and can be separated from AUDIT-T and AUDIT-C. This is consistent with the observation that, like other drugs of abuse such as nicotine, the pharmacological actions of alcohol are exerted through distinct genetic and neurobiological mechanisms (4). These findings illustrate that the quantity of alcohol use alone may not be a good proxy for AUD and that AUDIT-P alone captures a high proportion of the genetic risk for AUD.

Missing Heritability

Missing heritability remains a challenge. The SNP heritability for AUDIT-T (12%), AUDIT-C (11%), AUDIT-P (9%), and AUD (9%) (2) is smaller than what is observed in twin studies of AUD (49%) (5). The underlying heterogeneity and phenotypic complexity associated with aspects of problem alcohol use may explain the need to recruit massive cohorts to identify genetic variants that chip away at the missing heritability. The genetic correlations between AUDIT-P (2) and DSM-IV diagnosis of alcohol dependence (3) with schizophrenia, major depressive disorder, ADHD, and use of cannabis and cigarettes indicates that the genetic variants underlying these disorders are pleiotropic.

Both studies (3, 4) use an additive genetic linear model that does not include epistatic modeling (i.e., gene-gene interactions) (6). The SNP heritability or genetic risk score is typically computed by adding all the SNPs in a logistic regression until the maximum heritability is found. The associated SNPs are not necessarily the causal SNPs, but they exist at an allele frequency less than one. Assuming that these common SNPs are inherited independently, the probability of a single individual inheriting all of them is low. In the Sanchez-Roige study, the risk allele frequencies for the genome-wide significant SNPs associated with the genes *SNX17* and *GCKR*, *LINCO1833*, *RPL6P8*, *KLB* (2 SNPs), *METAP1*, *RP11–696N14.12*, *RP11–588P81*, *ADH1C*, *SLC39A8*, *RP11–700E23.3*, *JCAD*, *SLC39A13*, *MAPT*, and *FUT2* are 0.59, 0.69, 0.71, 0.61, 0.22, 0.99, 0.01, 0.01, 0.87, 0.91, 0.33, 0.68, 0.31, 0.24, and 0.56, respectively. Thus, the probability of one person inheriting all these SNPs is 1.98×10^{-8}. The probability of inheriting four of these SNPs (e.g., *SNX17* and *GCKR*, *LINCO1833*, *RPL6P8*, and *MAPT*) is $0.59\times0.69\times0.71\times0.24=6.9\%$, a frequency that is close to the population frequency of AUD (6%).

These SNPs explain less of the heritability that is computed using an additive model. Therefore, the genome-wide significant SNPs represent a population average in which the variants in common rise to genome-wide significance. There are likely additional alleles within any one individual that contribute to the phenotype but are not detected because their contribution to the phenotype on a population level is minimal, even though their contribution on an individual level is high. When the genetic complexity is low (e.g., in recombinant inbred mouse strains), the contribution of two or three alleles can account for 15%−30% of the variance in quantitative trait locus studies, and repeating the analysis on different strains often leads to the identification of different variants. This indicates that genetic background plays a significant role in phenotypic outcomes, and if these could be studied in a larger population, the results may yield different outcomes and the amount of the heritability explained by each locus associated with each SNP may be larger than what is predicted by the linear additive model.

Recent studies have incorporated epigenetic analyses, chromatin studies, and epistatic modeling to account for the missing heritability. One way to demonstrate this is through examining physical interactions between SNPs on the same chromosome. SNPs in weak disequilibrium are located near each other on the chromosome, because linkage disequilibrium increases with distance, with a maximum score of 0.5, which corresponds to a random (e.g., 50%) chance of association as a result of meiotic recombination. Elegant studies of hematopoietic cells show that SNPs in weak linkage disequilibrium to the risk SNP physically interact with the risk SNP to explain a two- to 10-fold increase in the genetic variance (7). In addition, proximal SNPs can be associated with enhancer clusters that interact with *trans*-acting factors in the three-dimensional genome (8). DNA is tightly wound around histone octamers in the nucleus, which forms the traditional chromosome structure. Chromosomes change position in the three-dimensional nucleus and can physically interact with one another. One way this is seen is through enhancer clusters, which interact at a distance to regulate a target gene. Different alleles of a SNP may strengthen or weaken this interaction. *Cis*-regulatory elements are discrete units of DNA that are proximal to the genes they regulate, such as promoters and enhancers. They often bind and interact with *trans*-acting factors, such as transcription factors, RNAs, and other chromosomes. The three-dimensional nucleome model illustrates how one SNP identified by a GWAS could interact with multiple target genes through various interactions. These targets can produce profound changes in gene networks by altering gene expression of coding and noncoding RNAs, histone modifications, DNA methylation, and mRNA splicing.

Gene-by-Environment Interactions

All alcohol and substance use disorders are by definition gene-by-environment interactions: alcoholism cannot occur without exposure to alcohol. Gene-by-environment interactions occur when individuals respond to their unique environment based on their individual background genotype. These sources of genetic variance are mediated by a variety of

epigenetic modifications. Adverse events during adolescence are associated with increased risk for developing a substance use disorder. However, not all adolescents who experience psychosocial stress develop a substance use disorder. It is likely that there are gene-by-environment interactions caused by long-term changes in gene expression through epigenetic modifications, such as histone modifications or DNA methylation at a specific genetic locus.

To address gene-by-environment interactions, Tay et al. (9), in this issue of the *Journal*, analyze the relationship between alcohol use, adverse events in adolescence, and epigenetic modifications of whole-blood DNA, with the presumption that the epigenetic modifications in the blood serve as a surrogate for nervous system tissue. Previously, neuroinflammation through activation of microglia was proposed to mediate the association between psychosocial stress and psychiatric and substance use disorders (8, 10). To test the role of epigenetic modifications in gene-by-environment interactions, Tay et al. conducted a genome-wide DNA methylation analysis of psychosocial stress and its association with alcohol and nicotine abuse using whole blood from 14-year-old adolescents (N=1,287). They identified and replicated a genome-wide significant methylated domain in the promoter of the sterile alpha motif/pointed domain epithelial transcription factor (*SPDEF*) gene. They found a gene-by-environment interaction at SNP rs2233631 in the *SPDEF* promoter. This SNP contains a 5′-CpG-3′ (cytosine nucleotide connected to a guanine by a phosphate bridge) dinucleotide (cg01395541). CpG dinucleotides are dynamic epigenetic signposts that can be subject to DNA methylation. Numerous studies have implicated DNA methylation in gene regulation. Although CpGs near gene promoters are typically hypomethylated, promoter methylation is often detected in gene silencing, and alterations in DNA methylation are implicated in alternative splicing, genomic imprinting, X-chromosome inactivation, and regulation of transposable elements (11).

Tay et al. (9) report an association between the frequency of stressful events and enhanced methylation at cg01395541. There are three alleles at this locus: GG, GA, and AA. The GG and GA variants are subject to DNA methylation, but the AA allele cannot be methylated. Adolescents carrying the GG (8.9% allele-specific frequency) and GA (36.3% allele-specific frequency) alleles are more susceptible to adverse life events and to engage in substance use compared with those carrying the AA allele (41.3% allele-specific frequency), indicating that genetic background can influence reactions to adverse life events through epigenetic mechanisms. Stratification by genotype increases the amount of the variance explained between lifetime alcohol consumption from 1.3% to 3.3% and alcohol binge drinking from 1.4% to 2.3%, without any change in the amount of the variance in lifetime smoking (4.4%). Surprisingly, methylation levels correlated with the frequency of smoking and drinking are associated with a reduction in activation in the right caudal cuneus gyrus, an area involved in vision.

To demonstrate the relevance of these biomarkers of psychosocial stress across blood and brain, Tay et al. also report a near-perfect correlation between blood-based DNA methylation at cg01395541 and methylation at the same site across multiple brain regions, including the prefrontal cortex, entorhinal cortex, superior cortex, and superior temporal gyrus, using the Blood Brain DNA Methylation Comparison Tool. They could not validate this observation experimentally as a result of the limited availability of adolescent postmortem brain samples. However, there are several caveats. Given the different cell types in blood (macrophages, monocytes, eosinophils, neutrophils, mast cells, T cells, and B cells, etc.), it is impossible to determine which cell types in blood and brain are methylated at cg01395541. The short lifespan of leukocytes (6–20 days), except for memory B cells and T cells, which are long-lived, suggests that methylation at cg01395541 occurs in stem cells or in memory B cells or T cells (12). Otherwise, the temporal relationship between adverse life events and methylation at cg01395541 would fall apart unless another tissue continues to send a signal of longer duration than the lifespan of the leukocytes.

Tay et al. describe a *trans*-expression quantitative trait locus (*trans*-eQTL) for cg01395541, suggesting that *SPDEF* or its promoter regulates expression in other genes. Trans eQTLs are relatively rare; 15% show overlap between blood and brain (13). In the Tay et al. study, methylation levels of the GG and GA alleles at cg01395541 were associated with expression of 159 genes, including *DRD2* and *OPRM1*. Hi-C analysis of the physical interaction of the methylated promoter of *SPDEF* at cg01395541 with other regions of the genome indicates that *DRD2* is physically linked with other genes. Many of these genes, such as mu opioid (*OPRM1*) and *DRD2*, have been implicated in substance use disorders. Interestingly, the Genotype-Tissue Expression (GTEx) project shows an eQTL for rs2233631 and *DRD2* in the cerebellum but not for *OPRM1* in any brain region. This may represent an eQTL that is not well represented in GTEx. Further analysis of the relationship between allele-specific methylation in the *SPDEF* gene promoter and gene expression in specific cell types is needed in postmortem brain tissue.

In conclusion, the study conducted by Sanchez-Roige (2) in this issue and the previous study conducted by Walter et al. (3) in another publication provide valuable insights into the genetic architecture of alcohol use disorder and emphasize the importance of phenotype. Quantity frequency may not be the best proxy for substance use disorders with the distinct variants contributing to the different phenotypes associated with DSM-IV criteria. Assumptions about viewing alcohol use disorder as a homogeneous disorder need to be examined. The study by Tay et al. (9) in this issue points to variation in a gene that may be important in conveying environmental influences related to risk and also illuminates some challenges of using blood as a proxy for the brain when studying gene-environment interactions. It is noteworthy that the current additive models using the analysis of complex traits

such as alcohol use do not account for epistatic interactions, complex gene regulatory interactions occurring at a single locus, or gene-environment interactions. While important steps in understanding the genetic underpinnings of substance abuse, these studies highlight the need for research using postmortem brain tissue, as well as the need for new statistical and computational methods (14).

AUTHOR AND ARTICLE INFORMATION

From the Genetics, Epigenetics, and Developmental Neuroscience Branch, National Institute on Drug Abuse, National Institutes of Health, Bethesda, Md. (Pollock and Lossie).

Send correspondence to Dr. Pollock (jpollock@nida.nih.gov).

The views and opinions expressed in this editorial are those of the authors only and do not necessarily represent the views, official policy, or position of the U.S. Department of Health and Human Services or any of its affiliated institutions or agency.

The authors report no financial relationships with commercial interests.

Accepted December 11, 2018.

Am J Psychiatry 2019; 176:92–95; doi: 10.1176/appi.ajp.2018.18121364

REFERENCES

1. Deak JD, Miller AP, Gizer IR: Genetics of alcohol use disorder: a review. Curr Opin Psychol 2018; 27:56–61
2. Sanchez-Roige S, Palmer AA, Fontanillas P, et al: Genome-wide association study meta-analysis of the Alcohol Use Disorders Identification Test (AUDIT) in two population-based cohorts. Am J Psychiatry 2019; 176:107–118
3. Walters RK, Polimanti R, Johnson EC, et al: Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat Neurosci 2018; 21:1656–1669
4. Changeux JP: Nicotine addiction and nicotinic receptors: lessons from genetically modified mice. Nat Rev Neurosci 2010; 11:389–401
5. Verhulst B, Neale MC, Kendler KS: The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. Psychol Med 2015; 45:1061–1072
6. Yang J, Benyamin B, McEvoy BP, et al: Common SNPs explain a large proportion of the heritability for human height. Nat Genet 2010; 42: 565–569
7. Corradin O, Cohen AJ, Luppino JM, et al: Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry. Nat Genet 2016; 48:1313–1320
8. Calcia MA, Bonsall DR, Bloomfield PS, et al: Stress and neuroinflammation: a systematic review of the effects of stress on microglia and the implications for mental illness. Psychopharmacology (Berl) 2016; 233:1637–1650
9. Tay N, Macare C, Liu Y, et al: Allele-specific methylation of *SPDEF*: a novel moderator of psychosocial stress and substance abuse. Am J Psychiatry 2019; 176:146–155
10. Lo Iacono L, Catale C, Martini A, et al: From traumatic childhood to cocaine abuse: the critical function of the immune system. Biol Psychiatry 2018; 84:905–916
11. Ferreira HJ, Esteller M: CpG islands in cancer: heads, tails, and sides, in Methods in Molecular Biology. New York, Springer, 2018, pp 49–80
12. Daniels VG, Wheater PR, Burkitt HG: Functional Histology: A Text and Colour Atlas. Edinburgh, United Kingdom, Churchill Livingstone, 1979
13. McKenzie M, Henders AK, Caracella A, et al: Overlap of expression quantitative trait loci (eQTL) in human brain and blood. BMC Med Genomics 2014; 7:31
14. Joubert W, Nance J, Climer S, et al: Parallel accelerated custom correlation coefficient calculations for genomics applications. Ithaca, NY, arXiv, Cornell University, 2017. https://arxiv.org/abs/1705.08213