PERSPECTIVE

Meta-Analyses of Genome-Wide Association Data Hold New Promise for Addiction Genetics

ABSTRACT. Meta-analyses of genome-wide association study data have begun to lead to promising new discoveries for behavioral and psychiatrically relevant phenotypes (e.g., schizophrenia, educational attainment). We outline how this methodology can similarly lead to novel discoveries in genomic studies of substance use disorders, and discuss challenges that will need to be overcome to accomplish this goal. We illustrate our approach with the work of the newly established Substance Use Disorders workgroup of the Psychiatric Genomics Consortium. (*J. Stud. Alcohol Drugs, 77,* 676–680, 2016)

SUBSTANCE USE DISORDERS (SUDS) ARE common, complex, and moderately heritable traits with obligate environmental components—exposure to and ingestion of the substance. Early candidate gene studies, coupled with strong biochemical knowledge, identified variants with protective effects against alcoholism that mapped to genes encoding alcohol-metabolizing enzymes (*ADH1B, ALDH2*), with effect sizes approximately 3.0 for a single protective allele. The functional alleles are common in Asian populations but uncommon to rare in European populations (Edenberg, 2007)[1]. It took large meta-analyses of European-ancestry subjects to confirm the role of rs1229984, a missense single-nucleotide polymorphism in *ADH1B,* as similarly protective against alcohol use disorders in those populations (Bierut et al., 2012; Gelernter et al., 2014).

For nicotine use, a candidate gene approach (Saccone et al., 2007)—followed by large genome-wide association study (GWAS) meta-analyses of cigarette smoking—has confirmed rs16969968 and its proxy, rs1051730, as highly replicable risk loci despite a relatively modest effect size (approximately one cigarette/day for each copy of the risk allele; 0.5% of the variance) (Liu et al., 2010; Thorgeirsson et al., 2010; Tobacco and Genetics Consortium, 2010). Other variants in cholinergic nicotinic receptor genes have also been identified (e.g., Saccone et al., 2010), and these discoveries have spurred functional, neuroimaging, and pharmacogenetic studies (e.g., Fowler et al., 2011; Jensen et al., 2015). Variants in other genes have been sporadically implicated for a variety of substance use phenotypes, albeit with smaller effects sizes and less robustness (Agrawal et al., 2012).

[1]Here and elsewhere, we only provide exemplar references. The psychiatric and addiction genetics literature is too vast to provide a comprehensive list of references. Thus, omission of references reflects convenience and not relevance.

GWASs have the advantage of allowing discovery of loci without *a priori* hypotheses. Common risk alleles that are genome-wide significant are anticipated to exert small effects; odds ratios for most alleles identified by GWASs are less than 1.2 (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Individual GWASs have generally been underpowered to identify such variants; therefore, it is not surprising that results of such studies have often not been replicated (Heath et al., 2011; Sullivan, 2012). Meta-analyses are the most economical, and therefore the most feasible, approach to aggregating the large sample sizes needed for complex diseases. Recent increases in sample size, made possible through meta-analyses, have produced landmark findings for a variety of complex traits including height (e.g., Wood et al., 2014), body mass index (e.g., Locke et al., 2015), type 2 diabetes (e.g., Morris et al., 2012), educational attainment (e.g., Rietveld et al., 2013), and, importantly, for psychiatric disorders such as schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Despite their modest effect size, loci from GWASs have altered our etiological understanding of complex traits and can potentially suggest targets for drug repositioning (e.g., Sanseau et al., 2012). SUDs are likely to follow a similar path.

Pooling data from different GWASs allow large increases in sample size—to the extent that there are many reasonably large studies to agglomerate—but there will inevitably be heterogeneity because of differences in population, ascertainment, and assessment. However, with sufficient numbers and attention to these sources of heterogeneity, replicable findings are likely. For SUDs, there are four main sources of such heterogeneity: (a) definition of cases, (b) definition of controls, (c) comorbidity, and (d) genomic sources of heterogeneity.

(a) Definition of cases

Individual studies ascertain subjects with SUDs from a variety of sources. Study design (case-control or family-based), severity of SUD in affected individuals, whether diagnoses reflect lifetime or current psychopathology, and developmental effects on transitions into and out of SUD during adolescence and young adulthood (e.g., maturing out of problem drinking; Chassin et al., 2013) are major sources of heterogeneity. Many can be modeled into the analyses, and, as subject numbers increase, power to detect loci of overarching effect on the phenotype, despite the heterogeneity, increases. The use of quantitative indices related to SUD (e.g., maximum drinks in a single 24-hour period, SUD criterion count) can also circumvent some of these concerns—including those related to diagnostic heterogeneity—and increase power (Markon et al., 2011), while also providing descriptive information regarding across-study differences.

(b) Definition of controls

Use of exposed controls (i.e., those who report using the substance but do not develop problems) can lead to novel discoveries (e.g., Nelson et al., 2013). First, genetic liability to SUD cannot be appropriately assessed in individuals lacking exposure (Kendler et al., 1999). Second, exposed controls allow identification of loci whose effects might be contingent on exposure or loci that protect against progression from exposure to SUD. Third, unexposed controls can be a significant source of confounding. For example, individuals with high genetic liability to alcoholism may not consume alcohol for cultural or health reasons or personal choice. If included as controls, these unexposed individuals can dilute the overall genetic signal. GWASs of alcohol and nicotine typically use exposed controls. For illicit drugs, this severely reduces the sample size of the control group, yet GWASs have primarily focused on exposed controls. The impact of including unexposed controls varies based on the extent to which genetic liability to initiation of use and development of SUDs is correlated (Neale et al., 2006), opportunities for exposure to the substance, and on the addictive potential of the substance. Thus, although unexposed/population controls generate larger sample sizes, exposed controls yield more interpretable results in the context of the conditional nature of SUDs.

(c) Comorbidity

The genetic overlap across SUDs is considerable (Hicks et al., 2011). Going beyond individual SUDs, phenotypic models that disentangle the effects of loci with substance-specific influence from those that confer vulnerability to a broader spectrum of addictive and related externalizing behaviors (e.g., latent factors) are necessary (e.g., Derringer et al., 2015). SUDs also commonly co-occur with other psychiatric illnesses. Some of this comorbidity is attributable to shared genetic effects (e.g., alcohol use disorder and major depressive disorder; e.g., Prescott et al., 2000). However, the etiology of such comorbidities can vary: for example, alcohol use disorder can increase negative affect and result in major depressive disorder, or alcohol use disorder can result from "self-medication" for major depressive disorder (Schuckit, 1994). Whether this difference translates to important variations in the underlying genetics and biology remains to be determined; GWASs large enough to allow comparisons of subgroups can help address this. Resources like the Psychiatric Genomics Consortium (PGC), which have large numbers of comorbid cases, can be used to augment sample sizes for SUD and also to explore their inter-relationships (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013) and outline this genomic architecture (The Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium, 2015).

(d) Genomic sources of heterogeneity

Pipelines that can systematically process GWAS data, either raw genotypes or summary statistics, from diverse sources reduce technical variability and increase power to detect real effects. Filters range from person-centered approaches to ascertaining population substructure, to variant-centered methods for identifying loci that fail to impute with confidence. The validated analytic pipeline of the PGC, which has primarily relied on quality control of raw genotypic data (but has more recently expanded to include summary statistics), has led to important discoveries for schizophrenia when sufficient sample sizes were attained. When meta-analyses included fewer than 10,000 cases, few loci reached genome-wide significance, and some oscillated in and out of significance. At ~18,000 cases, a much larger number of loci were significant, and, when the number of cases reached ~37,000, 108 independent loci were identified. This polygenic liability is now being linked to schizophrenia chronicity (e.g., Meier et al., 2016) and treatment resistance (e.g., Frank et al., 2015) as well as a variety of other psychiatric and medical conditions (e.g., Lee et al., 2015).

The GWAS meta-analytic finding of the association between schizophrenia and genes in the Major Histocompatibility Complex has now resulted in the identification of the role of complement component C4 haplogroups in the etiology of schizophrenia using a combination of translational techniques (Sekar et al., 2016). We are optimistic that GWAS meta-analyses of SUDs will identify similar actionable variants for a wide range of substances in the next decade. This enthusiasm stems not only from comparisons with successes for other complex traits but also from recent discoveries for SUDs, including genome-wide significant findings for alcohol (e.g., Gelernter et al., 2014), opioids (e.g., Nelson et

al., 2016), cannabis (e.g., Sherva et al., 2016), and nicotine dependence (e.g., Hancock et al., 2015).

Capitalizing on the PGC model and these recent GWAS discoveries for SUDs, an international team of addiction geneticists (currently 26 studies across 9 countries) with genome-wide association data formed the Substance Use Disorders Workgroup of the PGC. The Substance Use Disorders Workgroup is focused on those problematic aspects of substance involvement in which a majority of heritable variation appears to be concentrated. The initial analysis of the Substance Use Disorders Workgroup of the PGC includes nearly 13,000 alcohol-dependent cases and 30,000 controls of European American, African American, and Hispanic ancestry. Upcoming analyses will focus on other major diagnostic phenotypes (including nicotine, cannabis, cocaine, opioid, and other drug use disorders) as well as quantitative indices of addiction liability including measures of heavy use (e.g., maximum drinks per 24-hour period) and criterion counts. Analyses that disaggregate genetic effects on the shared liability to all SUDs from those that are substance specific will be undertaken in this large collection of data sets, which are, for the most part, richly phenotyped. The vast array of psychiatric data in the other PGC workgroups (Schizophrenia, Bipolar Disorders, Major Depressive Disorders, Autism Spectrum Disorders, Attention Deficit Hyperactivity Disorders, Anorexia Nervosa and Eating Disorders, Obsessive-Compulsive Disorder and Tourette's Syndrome, and Post-Traumatic Stress Disorders) as well as close collaborations with other consortia will allow us to estimate bivariate genetic relationships. These cross-cutting collaborations are further enhanced by the PGC's policy of sharing full-summary statistics for all completed GWAS, even those without genome-wide loci, with the entire community. Although we expect effect sizes associated with these loci to be small, common risk alleles that explain less than 1% of the variance can still lead to important biological and pharmacotherapeutic discoveries (Hirschhorn, 2009; Visscher et al., 2012).

The PGC has already gone well beyond expectations in identifying genetic risk factors for major psychiatric disorders, demonstrating pleiotropy between some of those disorders, and providing the impetus for major biological discoveries regarding disease mechanism (Sekar et al., 2016). Arguably, SUDs, despite their rather well-characterized pharmacokinetic and pharmacodynamic aspects, may be challenging phenotypes because of clinical heterogeneity and obligate environmental effects. Therefore, careful phenotype harmonization will need to accompany any gene-finding efforts. Nonetheless, with these experiences with GWASs for complex traits as our guide, we predict similar accomplishment in the field of SUD research (in both genetics and basic pathophysiology). Our progress will be commensurate with our success in adding as many subjects as possible to our analyses. This will require wide collaboration within the SUD research community, and most likely further National Institutes of Health–supported recruitment for those traits and populations whereby the world's extant DNA collection resources still fall short.

Arpana Agrawal, ph.d.[a]
arpana@wustl.edu
Howard J. Edenberg, ph.d.[b]
edenberg@iu.edu
Joel Gelernter, m.d.[c,d]
joel.gelernter@yale.edu

[a]*Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri*

[b]*Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana*

[c]*Department of Psychiatry, Genetics, and Neuroscience, Yale University School of Medicine, New Haven, Connecticut*

[d]*Veterans Affairs Connecticut Healthcare Center, West Haven, Connecticut*

Acknowledgments

PGC is funded by MH094421. We thank the National Institute on Drug Abuse for supporting us via an administrative supplement and the National Institute on Alcohol Abuse and Alcoholism for support via the Collaborative Study on the Genetics of Alcoholism (AA008401). Arpana Agrawal acknowledges K02DA32573 and Joel Gelernter acknowledges R01DA12690.

Conflict of Interest Statement

Arpana Agrawal previously received peer-reviewed grant funding and travel ad honorarium from the Alcoholic Beverage Medical Research Foundation (ended 2012), which receives support from the brewing industry. Joel Gelernter and Howard J. Edenberg have no conflicts of interest to disclose.

References

Agrawal, A., Verweij, K. J., Gillespie, N. A., Heath, A. C., Lessov-Schlaggar, C. N., Martin, N. G., . . . Lynskey, M. T. (2012). The genetics of addiction—A translational perspective. *Translational Psychiatry, 2,* e140. doi:10.1038/tp.2012.54

Bierut, L. J., Goate, A. M., Breslau, N., Johnson, E. O., Bertelsen, S., Fox, L., . . . Edenberg, H. J. (2012). ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Molecular Psychiatry, 17,* 445–450. doi:10.1038/mp.2011.124

Chassin, L., Sher, K. J., Hussong, A., & Curran, P. (2013). The developmental psychopathology of alcohol use and alcohol disorders: Research achievements and future directions. *Development and Psychopathology, 25,* 1567–1584. doi:10.1017/S0954579413000771

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *The Lancet, 381,* 1371–1379. doi:10.1016/S0140-6736(12)62129-1

Derringer, J., Corley, R. P., Haberstick, B. C., Young, S. E., Demmitt, B. A., Howrigan, D. P., . . . McQueen, M. B. (2015). Genome-wide association study of behavioral disinhibition in a selected adolescent sample. *Behavior Genetics, 45,* 375–381. doi:10.1007/s10519-015-9705-y

Edenberg, H. J. (2007). The genetics of alcohol metabolism: Role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Research & Health, 30,* 5–13.

Fowler, C. D., Lu, Q., Johnson, P. M., Marks, M. J., & Kenny, P. J. (2011). Habenular 5 nicotinic receptor subunit signalling controls nicotine intake. *Nature, 471,* 597–601. doi:10.1038/nature09797

Frank, J., Lang, M., Witt, S. H., Strohmaier, J., Rujescu, D., Cichon, S., . . . Rietschel, M. (2015). Identification of increased genetic risk scores for schizophrenia in treatment-resistant patients [Letter to the editor]. *Molecular Psychiatry, 20,* 150–151. doi:10.1038/mp.2014.56. Erratum in *Molecular Psychiatry, 20,* 913. doi:10.1038/mp.2015.52

Gelernter, J., Kranzler, H. R., Sherva, R., Almasy, L., Koesterer, R., Smith, A. H., . . . Farrer, L. A. (2014). Genome-wide association study of alcohol dependence: Significant findings in African- and European-Americans including novel risk loci. *Molecular Psychiatry, 19,* 41–49. doi:10.1038/mp.2013.145

Hancock, D. B., Reginsson, G. W., Gaddis, N. C., Chen, X., Saccone, N. L., Lutz, S. M., . . . Stefansson, K. (2015). Genome-wide meta-analysis reveals common splice site acceptor variant in *CHRNA4* associated with nicotine dependence. *Translational Psychiatry, 5,* e651. doi:10.1038/tp.2015.149

Heath, A. C., Whitfield, J. B., Martin, N. G., Pergadia, M. L., Goate, A. M., Lind, P. A., . . . Montgomery, G. W. (2011). A quantitative-trait genome-wide association study of alcoholism risk in the community: Findings and implications. *Biological Psychiatry, 70,* 513–518. doi:10.1016/j.biopsych.2011.02.028

Hicks, B. M., Schalet, B. D., Malone, S. M., Iacono, W. G., & McGue, M. (2011). Psychometric and genetic architecture of substance use disorder and behavioral disinhibition measures for gene association studies. *Behavior Genetics, 41,* 459–475. doi:10.1007/s10519-010-9417-2

Hirschhorn, J. N. (2009). Genomewide association studies—illuminating biologic pathways. *The New England Journal of Medicine, 360,* 1699–1701. doi:10.1056/NEJMp0808934

Jensen, K. P., DeVito, E. E., Herman, A. I., Valentine, G. W., Gelernter, J., & Sofuoglu, M. (2015). A CHRNA5 smoking risk variant decreases the aversive effects of nicotine in humans. *Neuropsychopharmacology, 40,* 2813–2821. doi:10.1038/npp.2015.131

Kendler, K. S., Neale, M. C., Sullivan, P., Corey, L. A., Gardner, C. O., & Prescott, C. A. (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychological Medicine, 29,* 299–308. doi:10.1017/S0033291798008022

Lee, S. H., Byrne, E. M., Hultman, C. M., Kähler, A., Vinkhuyzen, A. A. E., Ripke, S., . . . Wray, N. R., & the Schizophrenia Working Group of the Psychiatric Genomics Consortium and Rheumatoid Arthritis Consortium International. (2015). New data and an old puzzle: The negative association between schizophrenia and rheumatoid arthritis. *International Journal of Epidemiology, 44,* 1706–1721. doi:10.1093/ije/dyv136

Liu, J. Z., Tozzi, F., Waterworth, D. M., Pillai, S. G., Muglia, P., Middleton, L., . . . Barton, A., & the Wellcome Trust Case Control Consortium. (2010). Meta-analysis and imputation refines the association of 15q25 with smoking quantity [Letter]. *Nature Genetics, 42,* 436–440. doi:10.1038/ng.572

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., . . . Speliotes, E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature, 518,* 197–206. doi:10.1038/nature14177

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin, 137,* 856–879. doi:10.1037/a0023678

Meier, S. M., Agerbo, E., Maier, R., Pedersen, C. B., Lang, M., Grove, J., . . . Mattheisen, M. (2016). High loading of polygenic risk in cases with chronic schizophrenia. *Molecular Psychiatry, 21,* 969–974. doi:10.1038/mp.2015.130

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, A. V., Steinthorsdottir, V., . . . McCarthy, M. I. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics, 44,* 981–990. doi:10.1038/ng.2383

Neale, M. C., Harvey, E., Maes, H. H., Sullivan, P. F., & Kendler, K. S. (2006). Extensions to the modeling of initiation and progression: Applications to substance use and abuse. *Behavior Genetics, 36,* 507–524. doi:10.1007/s10519-006-9063-x

Nelson, E. C., Agrawal, A., Heath, A. C., Bogdan, R., Sherva, R., Zhang, B., . . . Montgomery, G. W. (2016). Evidence of *CNIH3* involvement in opioid dependence. *Molecular Psychiatry, 21,* 608–614. doi:10.1038/mp.2015.102

Nelson, E. C., Lynskey, M. T., Heath, A. C., Wray, N., Agrawal, A., Shand, F. L., . . . Montgomery, G. W. (2013). *ANKK1, TTC12,* and *NCAM1* polymorphisms and heroin dependence: Importance of considering drug exposure. *JAMA Psychiatry, 70,* 325–333. doi:10.1001/jamapsychiatry.2013.282

The Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. (2015). Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature Neuroscience, 18,* 199–209. doi:10.1038/nn.3922

Prescott, C. A., Aggen, S. H., & Kendler, K. S. (2000). Sex-specific genetic influences on the comorbidity of alcoholism and major depression in a population-based sample of US twins. *Archives of General Psychiatry, 57,* 803–811. doi:10.1001/archpsyc.57.8.803

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., . . . Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science, 340,* 1467–1471. doi:10.1126/science.1235488

Saccone, N. L., Culverhouse, R. C., Schwantes-An, T.-H., Cannon, D. S., Chen, X., Cichon, S., . . . Bierut, L. J. (2010). Multiple independent loci at chromosome 15q25.1 affect smoking quantity: A meta-analysis and comparison with lung cancer and COPD. *PLOS Genetics, 6*(8), e1001053. doi:10.1371/journal.pgen.1001053

Saccone, S. F., Hinrichs, A. L., Saccone, N. L., Chase, G. A., Konvicka, K., Madden, P. A. F., . . . Bierut, L. J. (2007). Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Human Molecular Genetics, 16,* 36–49. doi:10.1093/hmg/ddl438

Sanseau, P., Agarwal, P., Barnes, M. R., Pastinen, T., Richards, J. B., Cardon, L. R., & Mooser, V. (2012). Use of genome-wide association studies for drug repositioning. *Nature Biotechnology, 30,* 317–320. doi:10.1038/nbt.2151

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature, 511,* 421–427. doi:10.1038/nature13595

Schuckit, M. A. (1994). Alcohol and depression: A clinical perspective. *Acta Psychiatrica Scandinavica, 89, Supplement s377,* 28–32. doi:10.1111/j.1600-0447.1994.tb05798.x

Sekar, A., Bialas, A. R., de Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., . . . McCarroll, S. A.. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature, 530,* 177–183. doi:10.1038/nature16549

Sherva, R., Wang, Q., Kranzler, H., Zhao, H., Koesterer, R., Herman, A.,

. . . Gelernter, J. (2016). Genome-wide association study of cannabis dependence severity, novel risk variants, and shared genetic risks. *JAMA Psychiatry, 73,* 472–480. doi:10.1001/jamapsychiatry.2016.0036

Sullivan, P., & the 96 Psychiatric Genetics Investigators. (2012). Don't give up on GWAS [Letter to the editor]. *Molecular Psychiatry, 17,* 2–3. doi:10.1038/mp.2011.94

Thorgeirsson, T. E., Gudbjartsson, D. F., Surakka, I., Vink, J. M., Amin, N., Geller, F., . . . Stefansson, K. (2010). Sequence variants at *CHRNB3–CHRNA6* and *CYP2A6* affect smoking behavior [Letter]. *Nature Genetics, 42,* 448–453. doi:10.1038/ng.573

Tobacco and Genetics Consortium. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior [Letter]. *Nature Genetics, 42,* 441–447. doi:10.1038/ng.571

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics, 90,* 7–24. doi:10.1016/j.ajhg.2011.11.029

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., . . . Frayling, T. M.. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics, 46,* 1173–1186. doi:10.1038/ng.3097