



Commentary: A Gene-Based Association Method for Mapping Traits Using Reference Transcriptome Data

Udi E. Ghitza*

U.S. Department of Health and Human Services, Center for the Clinical Trials Network, National Institute on Drug Abuse, National Institutes of Health, Bethesda, MD, USA

Keywords: genetics, gene expression, bioinformatics, informatics, precision medicine, addiction, drug abuse, substance-use disorders

A commentary on

A gene-based association method for mapping traits using reference transcriptome data
by Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. *Nat Genet* (2015) 47:1091–8. doi: 10.1038/ng.3367

OPEN ACCESS

Edited by:

Mark Nicholas Ziats,
National Institute of Child Health and
Human Development, USA

Reviewed by:

Stefan Ehrlich,
University Hospital Carl Gustav
Carus, Germany
Gi Hoon Son,
Korea University, South Korea

*Correspondence:

Udi E. Ghitza
ghitzau@nida.nih.gov

Specialty section:

This article was submitted to
Behavioral and Psychiatric Genetics,
a section of the journal
Frontiers in Psychiatry

Received: 08 March 2016

Accepted: 11 April 2016

Published: 25 April 2016

Citation:

Ghitza UE (2016) Commentary:
A Gene-Based Association Method
for Mapping Traits Using Reference
Transcriptome Data.
Front. Psychiatry 7:73.
doi: 10.3389/fpsy.2016.00073

Uncovering and understanding links between gene expression, brain structure and function, and behavior may inform precision medicine by identifying more precise strategies to prevent and treat disease. Indeed, gene expression regulation mechanisms contribute to disease susceptibility, drug response, and the course of many substance-use disorders (1–3). Gamazon et al. recently reported development of an innovative gene-based association method to further understanding of gene-expression regulation mechanisms mediating relationships between single nucleotide polymorphism (SNP) variations and phenotypic variability of complex traits (4). They developed a computational method to utilize publically available gene expression datasets to estimate proportion of gene expression influenced by a person's genetic profile – genetically regulated expression – and to correlate this with complex phenotypes under investigation. Their whole-genome tissue-dependent prediction models – incorporating information on gene expression regulation from a set of markers in large high-resolution transcriptome databases – may be used to uncover mechanisms by which SNP-regulated gene expression contributes to disease susceptibility and complex traits. The investigators harnessed the power of large reference transcriptome data sets, such as the Genotype-Tissue Expression (GTEx) Project, the Genetic European Variation in Health and Disease (GEUVADIS), and Depression Genes and Networks (DGN), among others – in which both gene variation and gene expression levels have been measured – to estimate genetically regulated gene expression of SNPs. This information and genome-wide association study (GWAS) data are used to compose additive models of gene expression traits trained in reference transcriptome data sets. Thus, this novel method – termed PrediXcan – utilizes predictive algorithms to correlate associations between estimated genetically regulated gene expression of SNPs and phenotypic traits of interest using regression methods and non-parametric approaches, as a way to powerfully identify disease risk-mediating and trait-associated genes. Interestingly, unlike other gene-based testing methods, PrediXcan allows researchers to ascertain the direction of these associations. The authors demonstrated the utility of their method by identifying and replicating numerous new candidate associations within a previous data set. They explain how such an approach may be used to increase statistical power in genetics studies by reducing the multiple-testing penalty, which burdens many single-variant analysis approaches. Therefore, this method may afford researchers greater capability of reusing GWAS or whole-genome sequence data sets to detect novel trait-associated loci explaining a large

portion of disease susceptibility-associated phenotypic variability under control of genetically regulated gene expression. Research is needed to improve prediction capabilities of this approach when applying it to substance-use disorders and other behavioral health disorders, and to broaden its utility to map links between epigenetics and disease risks or traits.

There are many evidence gaps regarding how gene-regulation mechanisms may interact with environmental risk factors to contribute to disease susceptibility for substance-use disorders, especially during periods of high vulnerability such as brain and cognitive developmental windows during youth and adolescence (5). Therefore, a high-priority area for future research may be to employ this PrediXcan method or similar computational gene-testing approaches to identify human gene-expression-regulated genetic factors and mechanisms mediating risk and resilience for drug misuse and substance-use disorders, for instance, during brain and cognitive development. As the authors suggest, reanalyzing existing publicly available GWAS data sets in comprehensive biorepositories may increase the efficiency and cost-efficiency of elucidating such mechanisms (4). Another high-priority precision-medicine area is to determine whether and how prediction of expression profiles derived from genetic variance may be applied to different patient subgroups (e.g., those without substance-use problems, versus those exhibiting unhealthy drug or alcohol use but not yet considered to be on the severe end of the spectrum, versus those patients with severe substance-use disorder).

For this research to include phenotypes informative to both addiction-medicine researchers and clinical practice providers, rigorous and systematic research first needs to identify patient-centered health outcomes, which are reliably correlated with clinically meaningful changes in drug use in different populations and settings of patients with substance-use disorders. This would be particularly helpful given that substance-use disorders are chronic conditions often characterized by cycles of use, reduced use, abstinence, and relapse to use (6–8). Rigorous systematic research is also needed to establish reliable biomarkers for substance abuse, which might be used as objective measures for guiding such genetics research to predict risk for developing substance-use disorders, response to treatment, and risk for relapse. Thus, science needs to establish widely accepted clinically relevant phenotypic endpoints and biomarkers, which may help in standardizing data collection

and enable development of common data elements (CDEs) of clinically relevant phenotypic measures across studies. CDEs may, in turn, facilitate exchange of data by precisely describing semantic characteristics for a discrete piece of data, which will be collected, stored, or exchanged during the course of a study. CDEs of such measures could be key tools in building scalability in which uniform data elements with common semantic characteristics collected across studies can be incorporated and exchanged across networks and systems of networks. These benefits may be enhanced when CDEs conform to well-accepted data standards and ontologies (9). Thus, development and use of CDEs of clinically relevant phenotypic measures may facilitate this line of research by improving the efficiency and quality of data collection, as well as cross study comparisons, data aggregation, and meta-analyses.

Collectively, the above research initiatives are needed to accelerate the capacity to catalog gene expression mechanisms by which genetic variations map onto addiction risks and altered brain maturation, brain circuit function, and substance-use disorder behavioral patterns. Uncovering strong genetic evidence for such mechanisms can then be leveraged to develop an evidence-based Research Domain Criteria (RDoC) framework for substance-use disorder phenotypes, to complement one already established for other psychiatric conditions by the U.S. National Institute of Mental Health (NIMH) (10). Developing such RDoCs would expedite mechanistic precision-medicine research on genetics of substance-use disorders and psychiatric comorbidity to inform development of evidence-based approaches to improving patient-centered care for patients with multiple co-occurring psychiatric conditions.

AUTHOR CONTRIBUTIONS

UG, Ph.D., undertook a review of the literature, conceived of this general commentary, and wrote and reviewed all drafts.

FUNDING

UG is a Health Scientist Administrator, Program Official, at the Center for the Clinical Trials Network, NIDA, which is the funding agency for the National Drug Abuse Treatment Clinical Trials Network. This commentary was supported through Dr. Ghitza's duties as a Health Scientist Administrator.

REFERENCES

- Zhou Z, Enoch MA, Goldman D. Gene expression in the addicted brain. *Int Rev Neurobiol* (2014) **116**:251–73. doi:10.1016/B978-0-12-801105-8.00010-2
- Nestler EJ. Cellular basis of memory for addiction. *Dialogues Clin Neurosci* (2013) **15**:431–43.
- Kreek MJ, Levran O, Reed B, Schlussman SD, Zhou Y, Butelman ER. Opiate addiction and cocaine addiction: underlying molecular neurobiology and genetics. *J Clin Invest* (2012) **122**:3387–93. doi:10.1172/JCI60390
- Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* (2015) **47**:1091–8. doi:10.1038/ng.3367
- Enoch MA. The influence of gene-environment interactions on the development of alcoholism and drug dependence. *Curr Psychiatry Rep* (2012) **14**:150–8. doi:10.1007/s11920-011-0252-9
- DuPont RL, Compton WM, McLellan AT. Five-year recovery: a new standard for assessing effectiveness of substance use disorder treatment. *J Subst Abuse Treat* (2015) **58**:1–5. doi:10.1016/j.jsat.2015.06.024
- Volkow ND, Morales M. The brain on drugs: from reward to addiction. *Cell* (2015) **162**:712–25. doi:10.1016/j.cell.2015.07.046
- Volkow ND, Baler RD. Addiction science: uncovering neurobiological complexity. *Neuropharmacology* (2014) **76**:235–49. doi:10.1016/j.neuropharm.2013.05.007
- Ghitza UE, Gore-Langton RE, Lindblad R, Tai B. NIDA clinical trials network common data elements initiative: advancing big-data addictive disorders research. *Front Psychiatry* (2015) **6**:33. doi:10.3389/fpsy.2015.00033
- Insel TR. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *Am J Psychiatry* (2014) **171**:395–7. doi:10.1176/appi.ajp.2014.14020138

Disclaimer: The opinions in this paper are those of UG and do not represent the official position of the U.S. government.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ghitza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.