

COMMENTARY

Integration of innovative statistical methods using genetic data provides pharmacological insight and facilitates drug development

Andrew J. Grant 

MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

CorrespondenceAndrew J. Grant, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.
Email: andrew.grant@mrc-bsu.cam.ac.uk

The use of genetic data can be of great benefit in drug development. When analysed with appropriate statistical methods, such resources can be leveraged to identify potential drug targets and inform experimental trials.¹ It has been shown that drug development done with the backing of genetic data is more likely to be successful.² Increasingly, pharmacological studies are able to harness the results of genome-wide association studies (GWAS), which test for associations between a phenotype and genetic variation across the entire genome. Such studies are rapidly expanding in terms of both size of samples and range of phenotypes.³ Although GWAS are able to identify many genetic variants that are associated with a phenotypic trait of interest, they are not able to provide, on their own, evidence as to which of these associations are causal, or by which mechanisms these associations come about. New statistical methodology is being developed which uses genetic data to help to answer these questions.

The study of Li et al.⁴ uses an integrative framework which combines a number of analytical techniques with the aim of identifying genes that may play a role in clozapine-related neutropenia. This is an important problem because clozapine can be prescribed for treatment-resistant schizophrenia but is underutilized due to potential side effects such as neutropenia.⁴ Little is known about what makes a patient susceptible to neutropenia while taking clozapine, and so by identifying genes that may play a causal role in this process, insight can be gained into the underlying biological mechanisms involved. Li et al. use two sources of genetic data: GWAS summary statistics from a study that measured neutrophil count in individuals during clozapine treatment and summary statistics from a GWAS using expression quantitative trait loci (eQTL) data, which identified associations between genetic variants and gene expression levels. They apply Mendelian randomization (MR) and colocalization analyses using these datasets to identify relevant genes of interest and then

investigate these candidate genes further using gene set enrichment analysis.

MR is a technique that uses genetic data to assess the effect of an exposure on an outcome.⁵ A genetic variant that is associated with the exposure of interest is used as a proxy, or instrument, for varying that exposure. Because the genetic variants in an individual are determined randomly at conception, they are typically independent of environmental factors that can confound the exposure-outcome association. Thus, under certain assumptions, MR gives stronger evidence, compared with typical observational studies, that detected associations represent causal effects.

In a pharmacological context, genetic variants that are associated with the expression of a protein can be used as a proxy for perturbing that protein, for example, via a drug. Methodological innovations allow MR analyses to be performed by combining multiple genetic variants associated with the same exposure and by using summary level statistics from GWAS (that is, without requiring individual level data).⁵ Although most polygenic MR methods require the genetic instruments to be independent, applications in pharmacology typically involve multiple genetic variants from the same gene region which will be correlated (that is, in linkage disequilibrium). Recently, *cis*-MR techniques have been developed to allow for such cases.⁶ The technique applied by Li et al., probabilistic MR-Egger,⁷ was developed to perform *cis*-MR in a transcriptome-wide framework, testing for the effect of each gene region on the outcome.

Colocalization is a technique used to determine whether genetic associations with two traits from separate studies are due to a shared causal variant. The method of Giambartolomei et al.⁸ provides a test of the colocalization hypothesis using GWAS summary statistics. Evidence of a shared causal mechanism between a gene expression level and a disease outcome identifies that gene as a potential target for

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Author. *British Journal of Clinical Pharmacology* published by John Wiley & Sons Ltd on behalf of British Pharmacological Society.

pharmacological intervention. By applying this method using the gene regions shown by MR to be potentially causal, Li et al. identified those for which there is evidence of colocalization with neutrophil count during clozapine treatment. Gene set enrichment analysis then mapped these genes to biological pathways with which they are associated. Their results give insight into the possible mechanisms underlying clozapine-related neutropenia.

Although the approaches using genetic data discussed here can be very powerful for providing pharmacological insights, a number of considerations must be made. Each of the statistical methods make a number of assumptions, and the strength of their conclusions hinge on how well these can be justified. For example, the standard MR framework requires the genetic instruments to have no horizontal pleiotropy, that is, no association with the outcome other than via the exposure. Recent developments in MR methodology allow for some level of horizontal pleiotropy but only in return for making alternative assumptions. For example, the probabilistic MR-Egger approach allows for genetic variants to have horizontal pleiotropic effects on the outcome, but only if these effects are independent of their effects on the outcome via the exposure. This assumption will be violated, for example, if any of the genetic variants associate with a confounder of the exposure-outcome relationship. The validity of the colocalization test employed by Li et al. relies on the assumption that there is at most one association for each trait in the genetic region under consideration. Not only are many of the assumptions required in these analyses quite strong, they are also often untestable. However, consistent results from a range of approaches which make different assumptions give weight of evidence to their conclusions.⁹

A further limitation to using genetic data to identify therapeutic targets is the potential for index event bias, or collider bias. This can occur when genetic associations with a trait are estimated in a sample which has been selected based on a variable such as disease incidence. If any genetic variants cause both the disease and the trait of interest, then disease incidence is a potential collider, and conditioning the sample on it may bias the genetic association estimates.¹⁰ As a result, MR estimates may be biased. The outcome GWAS used by Li et al. was performed in a sample of patients with drug-resistant schizophrenia. If any genetic variants cause both schizophrenia and clozapine-related neutropenia, then schizophrenia is a potential collider.

Another important consideration is the study population and the ability to generalize findings. The study of Li et al. was done using datasets taken from samples of individuals of African ancestry. On one hand, this provides important information for this subpopulation, particularly in light of the fact that the under-prescription of clozapine may be particularly prevalent in this group.⁴ On the other hand, it may limit the generalizability of the findings to the wider population. Ideally, future studies would look to replicate the findings using samples from broader population groups. However, this is not always easy, particularly when using GWAS summary statistics, because the analysis relies not only on a relevant GWAS existing, but also on the results of the study being made publicly available. In the case of the study of Li et al., while there are a number of GWAS that have been performed looking at clozapine-related neutropenia, only one of them, which

studied individuals of African ancestry, has made summary statistics available.⁴ This highlights the importance of GWAS consortia to release summary statistics of their results.

Overall, the study of Li et al. demonstrates how the combination of multiple cutting-edge statistical techniques applied to genetic data can inform important questions in pharmacology, toward clinical translation. It outlines a framework for an exploratory analysis to identify genes which are potentially causally related to a disease outcome and possible mechanisms underlying these relationships. This in turn provides a basis for future in-depth studies and experimental trials.

ACKNOWLEDGEMENT

I thank Dipender Gill for his helpful comments.

CONFLICT OF INTEREST

The author has no conflicts of interest to declare.

ORCID

Andrew J. Grant  <https://orcid.org/0000-0001-6422-583X>

REFERENCES

- Gill D, Georgakis MK, Walker VM, et al. Mendelian randomization for studying the effects of perturbing drug targets [version 2; peer review: 3 approved, 1 approved with reservations]. *Wellcome Open Res.* 2021;6:16.
- Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet.* 2015;47(8):856-860.
- Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: biology, function and translation. *Am J Hum Genet.* 2017;101(1):5-22.
- Li J, Fang T, Shucheng S, Bojie W, Fuzhong X. Integration analysis of GWAS and expression quantitative trait loci identify genes and pathways for clozapine-related neutropenia. *Br J Clin Pharmacol.* 2021; <https://doi.org/10.1111/bcp.15043>
- Burgess S, Foley CN, Zuber V. Inferring causal relationships between risk factors and outcomes from genome-wide association study data. *Annu Rev Genomics Hum Genet.* 2018;19(1):303-207.
- Gkatzionis A, Burgess S & Newcombe PJ Statistical methods for cis-Mendelian randomization. arXiv. 2021; <https://arxiv.org/abs/2101.04081>
- Yuan Z, Zhu H, Zeng P, et al. Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. *Nat Commun.* 2020;11(1):3861.
- Giambartolomei C, Vukcevic D, Shadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):e1004383.
- Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol.* 2016;45(6):1866-1886.
- Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: conceptual and methodological challenges. *PLoS Genet.* 2017;13(10):e1006944.

How to cite this article: Grant AJ. Integration of innovative statistical methods using genetic data provides pharmacological insight and facilitates drug development. *Br J Clin Pharmacol.* 2021;1-2. doi:10.1111/bcp.15114