CellPress

Current Biology
Dispatches

Genetics: From Molecule to Society

Michel G. Nivard and Dorret I. Boomsma
Vrije Universiteit, Department of Biological Psychology, Netherlands Twin Register, Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands
Correspondence: m.nivard@vu.nl (M.G.N.), di.boomsma@vu.nl (D.I.B.)
http://dx.doi.org/10.1016/j.cub.2016.09.064

A genome-wide association study of neighborhood characteristics and family income finds heritability, identifies single nucleotide polymorphisms and shows genetic correlates of these traits with numerous other health and cognitive traits. Different mechanisms behind genetic correlations imply different interpretations of association and causality.

> Hairless dogs have imperfect teeth; long-haired and coarse-haired animals are apt to have, as is asserted, long or many horns; pigeons with feathered feet have skin between their outer toes; pigeons with short beaks have small feet, and those with long beaks large feet. Hence if man goes on selecting, and thus augmenting any peculiarity, he will almost certainly modify unintentionally other parts of the structure, owing to the mysterious laws of correlation
>
> — Charles Darwin, *The Origin of Species*, 1859

There hardly is a human trait for which a twin study has not been carried out [1] and the same now holds true for genome-wide association studies (GWAS). With the advent of the genomic era and affordable dense genotyping, GWAS results in humans are available for 'OMIC' traits and biomarkers (telomere length, gene expression, metabolites), brain structure and function, visible characteristics, numerous somatic and mental diseases and disorders, and traits like personality, temperament, entrepreneurship, educational attainment and subjective wellbeing. In this endeavor the UK Biobank plays a highly prominent role. The UK Biobank is a large repository of over 500,000 participants with survey and physical measures and blood, urine and saliva samples who are followed through health-related records (http://www.ukbiobank.ac.uk/). In this issue of *Current Biology*, Hill *et al.* [2] report a genome-wide association analysis of social deprivation and household income, which both index socio-economic status, in 112,151 participants from UK Biobank.

The study represents a milestone as it is the first genome-wide analysis of a cluster of variables that are not measured at the level of the individual, but at a family or neighborhood level. The study looked at 7 million single nucleotide polymorphisms (SNPs), household income and the Townsend deprivation index (a neighborhood level variable, which reflects the deprecation of a neighborhood as a function of unemployment, car and home ownership, and household overcrowding). Based on all SNPs, the heritability for social deprivation was estimated at 21% and for household income at 11%; thus, both variables were found to be heritable. We note that while the estimated heritabilities are not especially high, this does not have a direct bearing on the question of whether a trait is directly or indirectly influenced by the genome. Traits such as gene expression levels or epigenetic regulation of the genome are heritable traits [3,4] and their heritability often does not exceed the heritability reported for socio-economic status in this study.

Partitioned heritability analyses revealed that the genetic effects on socio-economic status are abundantly located in evolutionarily conserved regions of the genome, and in the case of social deprivation in genomic regions likely to affect the central nervous system. As the authors clearly convey, there is no reason to believe the relationship between genotype and socio-economic status reflects a direct effect of the genome on the phenotype. Rather it is likely that heritable traits mediate the relationship between genome and outcome. This implies that the SNPs significantly associated with the two measures of socio-economic status are not 'SNPs for socio-economic status'. Rather, it is likely that their effects are mediated through multiple other heritable phenotypes.

Beyond genome-wide evidence of genetic effects, which establishes genomic influence but does not pinpoint any specific genomic locations or genes, Hill *et al.* [2] obtain evidence that two loci associate with socio-economic status at a significance level which is genome-wide significant.

Next, in a series of bivariate genetic analyses, which make use of publically available results from GWAS meta-analyses for other traits, Hill *et al.* [2] estimate negative genetic correlations between socio-economic status and disease (obesity, type 2 diabetes), mental health (major depressive disorder, ADHD), addictive behaviors and personality (neuroticism). They estimate positive genetic correlations between socio-economic status and (childhood) IQ and longevity. These correlations offer some suggestion of the mediating phenotypes that give rise to the observed heritability of measures of social economic status.

Given that the estimation of genetic correlations no longer requires two traits to be measured within a single sample [5,6], genetic correlations between all traits subjected to genome-wide analyses can be estimated. Proliferation of genetic correlations suggests a careful consideration of the etiology of genetic correlations is in order. We consider the tantalizing possibility of leveraging the genome to distinguish causation from correlation and the implications of the detection of causal relations and genetic correlations between different traits. Hill

et al., for example, correlate the Townsend index (an interpersonal measures reflecting one's relative position in society) with body mass index (a personal measure) and HOMA-B and HOMA-IR (biomarkers related to β-cell function and insulin resistance). We discuss possibilities and pitfalls which may arise when correlating measures from the molecular to the individual and societal levels.

When the influence of the genome on complex outcomes such as socio-economic status reflects genetic effects of other heritable traits, one prediction that follows is that there must be a genetic correlation between these other traits and socio-economic status. As evident from Darwin's quote from *The Origin of Species*, the causes and consequences of genetic correlation have always been of interest to geneticists. The concept of genetic correlation (re)appeared in the genetics literature in 1943, when Hazel considered pleiotropic effects of genes, linkage or non-random mating as causes for genetic correlations between traits, with the last two effects expected to be less important than pleiotropy [7]. Pleiotropy is defined as the phenomenon in which a single locus affects two or more distinct phenotypic traits [8]. Pleiotropic effects of a gene, or of sets of genes, can be investigated by application of multivariate methods, which originally were mainly limited to plant and animal genetics [9]. Some of the earliest applications of multivariate genetic techniques to human data, which were seminal for twin studies, date back to the 1960s [10]. These methods relied on data from close relatives and twin pairs rather than on measured SNP effects.

Pleiotropy is not the only process that leads to genetic correlations. Others include statistical or methodological artifacts, a shared etiology between variables and direct causal relationships between traits, for example, a causal effect of body mass index on social deprivation, or a causal effect of social deprivation on brain development. These last two examples are of immediate and broad interest to the research community and policymakers.

Substantial work on disentangling causes of (genetic) correlation has been done in the field of psychiatry and psychiatric genetics. Neale and Kendler [11] proposed a taxonomy of causes of co-morbidity and their work generalizes as a framework for causes of genetic correlation. They distinguish causes of comorbidity as due to chance, population stratification, non-random sampling, shared genetic or environmental factors, and causal processes. Hill *et al.* show that the UK Biobank is a reasonable representation of the UK population by comparison to national census data. If there is evidence for non-random sampling, non-participation sometimes can be predicted from polygenic risk scores [12,13]. Recent techniques further accurately distinguish population stratification from true genetic signal [14]. In genetics the need to stringently control for multiple testing is generally adhered to, and replication of findings is broadly viewed as a prerequisite for publication. Hill *et al.* [2], for example, replicate their results in the Social Science Genetic Association Consortium and in the Scottish Family Health Study. These safeguards reduce the amount of chance findings. Other causes of genetic correlation include overlap in trait definitions, or alternative trait models. As illustrated by Neale and Kendler [11], the definitions of anxiety and depression share symptoms, ensuring any genetic effects on the shared symptoms is likely to induce a genetic correlation. A shared definition can extend beyond psychiatric symptoms; e.g. when a disease diagnosis is based on a biomarker crossing a threshold, this will induce a correlation between disease and biomarker. A genetic correlation will arise even if the biomarker merely serves as a convenient disease indicator and can be elevated in both diseased and non-diseased subjects.

An epidemiologically relevant cause of genetic correlation is direct causation. If a heritable trait causally influences another trait this must lead to a genetic correlation between them [15]. Recently the discussion of causal relations between molecular measures (gene transcripts, metabolites, epigenetic markers) and complex traits received abundant attention [16–18]. Mendelian Randomization can distinguish, given some assumptions, causation from correlation [19,20]. However, Mendelian Randomization studies of socio-economic status, which is influenced by many SNPs of small effect, requires large sample sizes. Traits with strong SNP effects (e.g. metabolites, gene transcripts, molecular and cellular measures) can be tested as causal traits in much smaller samples. The field of complex trait genetics should strive to mitigate this asymmetry. Doing so would allow GWAS results to function as a common, and unbiased, unit of analysis throughout biology, neuroscience, psychology and, as evident from the work of Hill *et al.* [2], also sociology and economy.

This conclusion brings us to a consideration of the 'levels of explanation' in the study of complex traits and diseases. Trait and disease etiology are studied at the societal level in sociology and economics, at the level of the group or individual in psychology, at individual levels in behavioral and medical neuroscience, biology and medicine, at the cellular level in biology and neuroscience, and at the molecular level in bio-chemistry, neuroscience, and physics. At all these levels causal relations between variables may exist, and in many cases causal processes could span different levels of explanation and such causal effects may be reciprocal. Contemporary complex trait genetics is quickly becoming a potent facilitator of interdisciplinary study, as exemplified in the Hill *et al.* study, its domain of application ranging beyond the individual. The success of this enterprise entirely depends on broad dissemination of the full summary statistics of GWA studies.

While post-hoc analysis of genetic correlation and causation based on summary statistics obtained from GWAS are valuable, the work by Hill *et al.* alludes to potential gains that can be made when raw genotype data are available. It is evident that people vary in terms of socio-economic status within and between neighborhoods. Hill *et al.* [2] applied genome-wide analysis to socio-economic status differences between individuals (household income) and between neighborhoods (Townsend deprivation index). An exciting next step that we envision is to subject individual differences within neighborhoods (relative to the neighborhood mean) and differences between neighborhoods to genome-wide analyses in a single model,

testing whether the same genes influence differences in socio-economic status within and between neighborhoods. Within- and between-neighborhood genetic effects on socio-economic status might each have unique genetic correlates. To study the between-neighborhood genetic effects on socio-economic status, concurrently with the within-neighborhood genetic effect on socio-economic status, we conceptualize a neighborhood level polymorphism (the allele count for the reference allele for all subjects in a given neighborhood), and the within-neighborhood genetic effects as the individual deviance from this neighborhood-level polymorphism. Variables subjected to genetic analysis can frequently be viewed as nested in neighborhood, subject, tissue or time. Broad access to large collections of genotyped and densely phenotyped subjects could help resolve genetic correlations and causation at multiple explanatory levels.

REFERENCES

1. Martin, N., Boomsma, D., and Machin, G. (1997). A twin-pronged attack on complex traits. Nat. Genet. *17*, 387–392.
2. Hill, W.D., Hagenaars, S.P., Marioni, R.E., Harris, S.E., Liewald, D.C.M., Davies, G., Okbay, A., McIntosh, A.M., Gale, C.R., and Deary, I.J. (2016). Molecular genetic contributions to social deprivation and household income in UK Biobank. Curr. Biol. *26*, 3083–3089.
3. Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y.-H., *et al*. (2014). Heritability and genomics of gene expression in peripheral blood. Nat. Genet. *46*, 430–437.
4. van Dongen, J., Nivard, M.G., Willemsen, G., Hottenga, J.-J., Helmer, Q., Dolan, C.V., Ehli, E.A., Davies, G.E., van Iterson, M., Breeze, C.E., *et al*. (2016). Genetic and environmental influences interact with age and sex in shaping the human methylome. Nat. Commun. *7*, 11115.
5. Bulik-Sullivan, B., Finucane, H., Anttila, V., Gusev, A., Day, F., Loh, P., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, and Duncan, L., *et al*. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.
6. Nieuwboer, H.A., Pool, R., Dolan, C.V., Boomsma, D.I., and Nivard, M.G. (2016). GWIS: Genome-wide inferred statistics for functions of multiple phenotypes. Am. J. Hum. Genet. *0*, 1236–1241.
7. Hazel, L.N. (1943). The genetic basis for constructing selection indices. Genetics *28*, 476–490.
8. Stearns, F.W. (2010). One hundred years of pleiotropy: a retrospective. Genetics *186*, 767–773.
9. Eaves, L.J., and Brumpton, J.R. (1972). Factors of covariation in Nicotiana rustica. Heredity *9*, 151–175.
10. Vandenberg, S., Clark, P., and Samuels, I. (1965). Psychophysiological reactions of twins: Hereditary factors in galvanic skin resistance, heartbeat, and breathing rates. Eugen. Q. *12*, 7–10.
11. Neale, M.C., and Kendler, K.S. (1995). Models of comorbidity for multifactorial disorders. Am. J. Hum. Genet. *57*, 935–953.
12. Martin, J., Tilling, K., Hubbard, L., Stergiakouli, E., Thapar, A., Davey Smith, G., O'Donovan, M.C., and Zammit, S. (2016). Association of genetic risk for schizophrenia with nonparticipation over time in a population-based cohort study. Am. J. Epidemiol. *183*, 1149–1158.
13. Nivard, M.G., Gage, S.H., Hottenga, J.-J., van Beijsterveldt, C.E.M., Abdellaoui, A., Baselmans, B.M.L., Ligthart, L., St Pourcain, B., Boomsma, D.I., Munafo, M.M., *et al*. (2016). Genetic overlap between schizophrenia and developmental psychopathology: a longitudinal approach applied to common childhood disorders between age 7 and 15 years. bioRxiv, 52829.
14. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Consortium, S.W.G. of the P.G., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.
15. De Moor, M.H.M., Boomsma, D.I., Stubbe, J.H., Willemsen, G., and de Geus, E.J.C. (2008). Testing causality in the association between regular exercise and symptoms of anxiety and depression. Arch. Gen. Psychiatry *65*, 897–905.
16. Pickrell, J.K., Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., *et al*. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet. *94*, 559–573.
17. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., *et al*. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
18. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., *et al*. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.
19. Smith, G., and Ebrahim, S. (2004). Mendelian randomization: prospects, potentials, and limitations. Int. J. Epidemiol. *33*, 30–42.
20. Bowden, J., Smith, G.D., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int. J. Epidemiol. *44*, 512–525.