1. **Challenging the Utility of Polygenic Scores for Social Science: Environmental Confounding, Downward Causation, and Unknown Biology**

2. Benjamin C. Nephew, Christopher Murgatroyd, Hudson Santos, Angela Rodriguez

Abstract

Genetic studies in the social sciences could be augmented through the additional consideration of functional (transcriptome, methylome, metabolome) and/or multimodal genetic data when attempting to understand the genetics of social phenomena. Understanding the biological pathways linking genetics and the environment will allow scientists to better evaluate the functional importance of polygenic scores.

Commentary

The article by Burt is timely in that it raises the importance of needing "dialogue between social and behavioral scientists about the scientific value of adding genetics to social science at the current state of knowledge." We agree with many of the issues raised and the complications of incorporating polygenic scores (PGSs) and genetics into models without fully understanding "the scientific costs." It is clear that PGSs need to be controlled for "population stratification, familial confounding, and downward (socio-environmental) causation" and the exposome of an individual but treating these solely as environmental confounders neglects the important impact of biology.

For example, though exposure to adversity at a particular timepoint increases the risk for emotional and behavioral symptoms and stress-related disorders, not everyone exposed develops these symptoms. Individuals respond to environments in different ways, and we now have methods which can provide insight into this previously "unknown biology." Although we agree that PGSs might not capture this biological risk across different studies because of many of the important points raised in this article, it does not mean that it is not valuable. PGSs and how they relate to social science need to be presented with the appropriate limitations or with additional, more functional assessments of genetic context.

Genetic studies in the social sciences could be augmented through the additional consideration of functional (transcriptome, methylome, metabolome) and/or multimodal genetic data when attempting to understand the genetics of social phenomena. PGSs should be embedded within multiple omics approaches, which could be further augmented with specific measures of physiological effectors such as functional neuroimaging or endocrine assays. Understanding the biological pathways potentially linking genetics and the environment will allow scientists to better evaluate the functional importance of PGSs.

Methylome studies have identified changes in DNA methylation as markers of overall brain health (Gadd et al., Reference Gadd, Hillary, McCartney, Shi, Stolicyn, Robertson and Marioni2022). More direct epigenetic research has demonstrated links between DNA methylation and educational attainment, indicating that the methylome of lower-educated people was suggestive of exposure to pollution (van Dongen et al., Reference van Dongen, Bonder, Dekkers, Nivard, van Iterson and Willemsen2018). DNA

methylation has also been linked to chronic cannabis use with associated changes in cognitive performance (Wiedmann et al., Reference Wiedmann, Kuitunen-Paul, Basedow, Wolff, DiDonato, Franzen and Golub2022). Additional inclusion and considering of epigenetic data in PGS studies may enhance our understanding of how the environment, especially during early life, impacts our genome to induce lasting effects, allowing us to progress from environmental confounding to environmental mediation and/or modulation.

Transcriptome data have provided valuable insight into how genes play a role in complex traits and disease (Hatcher, Relton, Gaunt, & Richardson, Reference Hatcher, Relton, Gaunt and Richardson2019). Neuroimaging-based research has yielded associations between transcriptome-wide genes for brain structures and complex traits in different domains (Zhao et al., Reference Zhao, Shan, Yang, Yu, Li, Wang and Zhu2021). Cortical transcriptome changes have been specifically linked with educational attainment (Bartrés-Faz et al., Reference Bartrés-Faz, González-Escamilla, Vaqué-Alcázar, Abellaneda-Pérez, Valls-Pedret, Ros and Grothe2019). Combining transcriptome data with PGSs can provide a clearer picture of which specific genes are having the most significant effects on social factors at discrete points in time. However, there are substantial challenges with transcriptome data because of tissue and temporally specific gene expression that limit its application.

The metabolome has been a topic of expanding interest in how genes affect change. Studies examining the metabolome have highlighted social-to-biological processes resulting in heath inequalities (Karimi et al., Reference Karimi, Castagné, Delpierre, Albertus, Berger, Vineis and Chadeau-Hyam2019). Using metabolic profiles, other investigations have revealed that social and economic factors have measurable impact on human physiology (Robinson et al., Reference Robinson, Carter, Ala-Korpela, Casas, Chaturvedi, Engmann and Vineis2021). Metabolic impairment has been associated with the apolipoprotein E4 and insulin resistance in type 2 diabetes, which is often mediated by socioeconomic factors and is a major risk factor for late onset of Alzheimer's disease. These results could guide development of socioeconomic-based preventive measures and therapies for cognitive decline (Johnson et al., Reference Johnson, Torres, Impey, Stevens and Raber2017). Similar to transcriptome data, metabolome information could provide crucial temporally specific functional insight into how environmental and social factors interact with the genome to induce change.

There is a critical need to go beyond simple PGSs and institute more comprehensive social and genetic data collection (which is becoming more readily available) to strengthen associations and improve causal conclusions on how genes and environment interact to affect behavior. As has been observed with both genome- and brain-wide association studies, bigger is not always better, and an increased focus on smaller, more thoroughly characterized populations with functional genetic data will lead to stronger conclusions. A critical factor that needs to be considered in all studies is the growing awareness of the plasticity of genetic mechanisms of behavior, particularly the role of epigenetics.

The line between what was traditionally seen as genetic and environmental effects is increasingly blurred. Rather than environmental confounding, these may be epigenetic effects, and the discussion of PGSs in social science would be informed by a greater understanding of and appreciation for animal studies of behavioral genetics, where the bar for causal conclusions may be much higher. This is an especially important consideration in discussions of using PGSs, or any other type of genetic data, to control for genetic effects and focus on environmental factors. This is a problematic notion at the very least. Even if other types of genetic data are beyond the primary focus of the target article, we argue

that consideration of functional genetic outputs is critical for future genetic studies in the social sciences, whether or not these data are collected in a particular PGS study.

References

Bartrés-Faz, D., González-Escamilla, G., Vaqué-Alcázar, L., Abellaneda-Pérez, K., Valls-Pedret, C., Ros, E., & Grothe, M. J. (2019). Characterizing the molecular architecture of cortical regions associated with high educational attainment in older individuals. The Journal of Neuroscience, 39(23), 4566. https://doi.org/10.1523/JNEUROSCI.2370-18.2019

Gadd, D. A., Hillary, R. F., McCartney, D. L., Shi, L., Stolicyn, A., Robertson, N. A., … Marioni, R. E. (2022). Integrated methylome and phenome study of the circulating proteome reveals markers pertinent to brain health. Nature Communications, 13(1), 4670. https://doi.org/10.1038/s41467-022-32319-8

Hatcher, C., Relton, C. L., Gaunt, T. R., & Richardson, T. G. (2019). Leveraging brain cortex-derived molecular data to elucidate epigenetic and transcriptomic drivers of complex traits and disease. Translational Psychiatry, 9(1), 105. https://doi.org/10.1038/s41398-019-0437-2

Johnson, L. A., Torres, E. R. S., Impey, S., Stevens, J. F., & Raber, J. (2017). Apolipoprotein E4 and insulin resistance interact to impair cognition and alter the epigenome and metabolome. Scientific Reports, 7(1), 43701. https://doi.org/10.1038/srep43701

Karimi, M., Castagné, R., Delpierre, C., Albertus, G., Berger, E., Vineis, P., … Chadeau-Hyam, M. (2019). Early-life inequalities and biological ageing: A multisystem biological health score approach in Understanding Society. Journal of Epidemiology and Community Health, 73(8), 693. https://doi.org/10.1136/jech-2018-212010

Robinson, O., Carter, A. R., Ala-Korpela, M., Casas, J. P., Chaturvedi, N., Engmann, J., … Vineis, P. (2021). Metabolic profiles of socio-economic position: A multi-cohort analysis. International Journal of Epidemiology, 50(3), 768–782. https://doi.org/10.1093/ije/dyaa188

van Dongen, J., Bonder, M. J., Dekkers, K. F., Nivard, M. G., van Iterson, M., Willemsen, G., … consortium, B. (2018). DNA methylation signatures of educational attainment. NPJ Science of Learning, 3(1), 7. https://doi.org/10.1038/s41539-018-0020-2

Wiedmann, M., Kuitunen-Paul, S., Basedow, L. A., Wolff, M., DiDonato, N., Franzen, J., … Golub, Y. (2022). DNA methylation changes associated with cannabis use and verbal learning performance in adolescents: An exploratory whole genome methylation study. Translational Psychiatry, 12(1), 317. https://doi.org/10.1038/s41398-022-02025-6

Zhao, B., Shan, Y., Yang, Y., Yu, Z., Li, T., Wang, X., … Zhu, H. (2021). Transcriptome-wide association analysis of brain structures yields insights into pleiotropy with complex neuropsychiatric traits. Nature Communications, 12(1), 2878. https://doi.org/10.1038/s41467-021-23130-y