Review of: "Signals of Human Polygenic Adaptation: Moving Beyond Single-Gene Methods and Controlling for Population-Specific Linkage Disequilibrium"

Françoise Clerget-Darpoux[1]

1 French Institute of Health and Medical Research

Potential competing interests: No potential competing interests to declare.

This study is intended to show that a substantial proportion of variation in polygenic scores can be found between populations for educational achievement (or IQ) and schizophrenia. It concludes to polygenic adaptation for these phenotypes.

For the author, the polygenic score of a trait represents a measure of additive genetic variance and Qst allows to quantify the genetic part of phenotypic differences. These views are widely spread in the literature; but unfortunately they are erroneous (1, 2) as polygenic scores and Qst are based on :

1) **An inappropriate underlying genetic model to describe the trait**

By construction, polygenic scores assume an additive polygenic transmission model for the trait under study. Among the many hypotheses of this model, it is ruled out that an environmental factor can play an important role in the expression of the trait. And yet, when it comes to assessing - for example - the IQ, we know the extent to which access to education is a determining factor. This model also assumes that there is no covariation or interaction between genetic and environmental factors. Indeed, environmental factors are assumed to have a random impact on the individual, irrespective of family, social and occupational conditions.

Also, when adopting the polygenic additive model for disease's liability, one assumes an homogeneous pathological processes to be at the root of the disease (3) … whereas human diseases - schizophrenia being no exception - are highly heterogeneous.

2) A misinterpretation of the associations with genetic markers

While the association between a trait and a genetic marker may indicate the role of a genetic factor, it may also reflect environmental or cultural factors. For example, a GWAS study comparing people in France who consume salted butter (mainly French from Brittany) with those who consume unsalted butter would show a large number of genetic markers associated with this trait. Not because it would reveal genetic factors conferring a particular taste for salted butter … but because these markers differ between Brittany and other French regions. The problem of interpreting associations also arises for traits that are due to complex interactions between genetic and environmental factors. For example, associations observed between markers and Body Mass Index(BMI) may reflect socio-cultural differences, particularly

differences in eating habits. Similarly, a difference can be observed for the IQ variable when comparing two groups with different access to a given cultural environment. All the more so since the IQ variable - defined in a standardized way with an average of 100 for a given school learning process - has been inappropriately transformed into a universal measure for totally heterogeneous learning processes (in time, space and even between sexs, at certain periods or in certain countries).

The interpretation of an association as reflecting the role of a genetic variant must go beyond the use of population information alone. It must be confirmed by subsequent familial and functional studies. Although geneticists have the tools to avoid any misinterpretation of an association or correlation in terms of causality, the idea that a genome-wide association study alone can measure genetic variance has become dogma. Admittedly, some people have contested this (4,5,6), but they have been no match for the enormous quantity of literature that affirms it as a truth!

This is how the author of this study, following his teacher Richard Lynn, explains by genetic differences the differences observed between continents, between regions of the same country and also between sexes. With all these genetic differences, one wonders what the author means by 'pure' Native American polygenic score !!!

The misinterpretation of differences observed between groups for IQ measurement is nothing new.Unfortunately, despite the extraordinary progress in biological knowledge – in particular with epigenetics which shows that the expression of our genes depends on our environment from the moment we are conceived until our death - the return to a model which denies gene-environment interactions is particularly harmful. This assumes from the outset that it is possible to separate the effects of genetic factors from those of the environment … ultimately leading to the idea that all our traits, pathological or otherwise, are predetermined.

https://theconversation.com/does-our-dna-really-determine-our-intelligence-and-health-199266

Many scientists, whether sociologists, psychologists or clinicians, are unaware of the impact that false hypotheses can have on their conclusions (7). It is less likely for the group led by Wray and Visscher which has been fueling this drift for many years, in many ways:

- by adopting the additive polygenic model as the architecture of complex trait to compute polygenic risk scores and heritability of disease's liability (G1);

- by recognizing that, for complex human traits, G and E covariation and G*E interaction exist and that heritability estimates using twin studies are very sensitive to it (G2) … while continuing to support a model that assumes both no such covariation nor interaction;

- by interpreting the contrast between genomic profile of affected and unaffected in terms of genetic variance (G3) … whereas, by construction, the difference between the PRS curves in affected and unaffected people only reflects the fact that the draw of marker alleles is done with different frequencies between the two groups (it is therefore another way of expressing the associations observed between the genetic markers and the disease under study).

- by denying the importance of familial information (G3) … whereas it is so critical to discriminate between different familial recurrence risks for heterogeneous diseases and for the interpretation of association studies

- by making people believe that GWAS make it possible to extract genetic factors independently of environmental factors

- by spreading the idea that going back to a model proposed over a century ago was more modern and realistic than assuming the existence of interactions between genes or significant effects of certain genes in subgroups or gene pathways (G4). One of the arguments given for returning to the polygenic additive model was the small OR observed for GWAS ...whereas for heterogeneous diseases, this does not at all rule out the strong effect of certain genetic factors in small subgroups. Another argument is that one does not show any interaction on GWAS data … when one already has little power to show marginal effects.

- by drawing on its dual expertise in animal and human genetics, to assert that it is equivalent to use breeding values (EBVs) in livestock and polygenic risk scores (PRS) for humans (G5). Having some experience in animal breeding (10), I totally dispute this assertion. Indeed, in livestock, it is possible to control the environment and account for interactions with the livestock genome … while this is impossible for humans. Furthermore, EBVs cannot be used to determine the phenotypic value of each animal in the herd, but only to classify them for what they will pass on - in mean - to the next generation (1). Similarly, I dispute the AUC interpretation given in G3. AUC is useful for measuring the efficacy of animal selection, but not for correctly classifying diseased and non-diseased individuals (11).

- by supporting the erroneous views of Paige Harden that measuring cognitive performance through a polygenic score would allow a fairer comparison across individuals than usual scholar tests (G6)...while ignoring potentially less generous uses.

Feldman & Lewontin's assertion (10) that *"Partitioning of the causes of variation is really illusory"* remains valid. It is wrong to pretend that genome-wide association studies make it possible. It is important to bear in mind that *"contemporary biology has demonstrated that traits are the product of interactions between genetic and non-genetic factors at every point of the development*" (11). Returning to a model that effectively separates genetic effects from environmental effects, thereby denying our biological knowledge, leads to the erroneous conclusion of this study.

References for this review

1 - Herzig AF, Clerget-Darpoux F, Génin E. The False Dawn of Polygenic Risk Scores for Human Disease Prediction. J Pers Med. 2022 Jul 31;12(8):1266. doi: 10.3390/jpm12081266. PMID: 36013215; PMCID: PMC9409868.

2 - Robette N, Génin E, Clerget-Darpoux F. Heritability: What's the point? What is it not for? A human genetics perspective. Genetica. 2022 Aug;150(3-4):199-208. doi: 10.1007/s10709-022-00149-7. Epub 2022 Jan 29. PMID: 35092541.

3 - Génin, E.; Clerget-Darpoux, F. Revisiting the Polygenic Additive Liability Model through the example of *Diabetes mellitus*. *Hum. Hered.* **2015**, *80*, 171–177. https://doi.org/10.1159/000447683.

4 - Bourgain, C.; Génin, E.; Cox, N.; Clerget-Darpoux, F. Are Genome-Wide Association Studies All That We Need to Dissect the Genetic Component of Complex Human Diseases? *Eur. J. Hum. Genet.* **2007**, *15*, 260–263. https://doi.org/10.1038/sj.ejhg.5201753

5 - Boyle EA, Li YI, Pritchard JK (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169:1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

6 - Janssens ACJW (2019) Validity of polygenic risk scores: are we measuring what we think we are? Human Molecular Genetics 28:R143–R150. https://doi.org/10.1093/hmg/ddz205

7 - Génin, E.; Clerget-Darpoux, F. The Missing Heritability Paradigm: A Dramatic Resurgence of the GIGO Syndrome in Genetics. *Hum. Hered.* **2015**, *79*, 1–4. https://doi.org/10.1159/000370327.

8 - Foulley J, Clerget-Darpoux F. Progeny group size for evaluating natural service bulls using AI reference sires. Ann Genet Sel Anim. 1978;10(4):541-56. doi: 10.1186/1297-9686-10-4-541. PMID: 22896119; PMCID: PMC3316408.

9 - Wald, N.J.; Old, R. The Illusion of Polygenic Disease Risk Prediction.*Genet. Med.* **2019**, *21*, 1705–1707. https://doi.org/10.1038/s41436-018-0418-5.

10 - Feldman, M.W.; Lewontin, R.C. The Heritability Hang-Up. *Science* **1975**, *190*, 1163–1168. https://doi.org/10.1126/science.1198102.

11- Moore, D.S.; Shenk, D. The Heritability Fallacy. *Wiley Interdiscip. Rev. Cogn. Sci.* **2017**, *8*, e1400. https://doi.org/10.1002/wcs.1400.

A small but representative sample of references from the group of Wray and Visscher

G1 - Wray, N.R.; Goddard, M.E.; Visscher, P.M. Prediction of Individual Genetic Risk to Disease from Genome-Wide Association Studies. *Genome Res.* **2007**, *17*, 1520–1528. https://doi.org/10.1101/gr.6665407.

G2 - Visscher PM, Hill WG, Wray NR. 2008 Heritability in the genomics era-concepts and misconceptions. Nat Rev Genet 9: 255-266

G3 - Wray NR, Yang J, Goddard ME, Visscher PM. The genetic interpretation of area under the ROC curve in genomic profiling. PLoS Genet. 2010 Feb 26;6(2):e1000864. doi: 10.1371/journal.pgen.1000864. PMID: 20195508; PMCID: PMC2829056.

G4 - Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. Cell. 2018 Jun 14;173(7):1573-1580. doi: 10.1016/j.cell.2018.05.051. PMID: 29906445.

G5 - Wray, N.R.; Kemper, K.E.; Hayes, B.J.; Goddard, M.E.; Visscher, P.M. Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans: Genomic Prediction. *Genetics* **2019**, *211*, 1131–1141. https://doi.org/10.1534/genetics.

G6 - Visscher PM. Genetics of cognitive performance, education and learning: from research to policy? NPJ Sci Learn. 2022 May 4;7(1):8. doi: 10.1038/s41539-022-00124-z. PMID: 35508631; PMCID: PMC9068617.

No potential competing interests to declare.