

Comments on Pritchard Paper

Nancy J. Cox^{1*}

¹ Director, Vanderbilt Genetics Institute.

***Correspondence:** Nancy J. Cox, Email: nancy.j.cox@vanderbilt.edu.

The Boyle *et al.* paper is one of those “Rorschach” papers, in which statements on the meaning of the paper may be revealing more about what the commentator believed about the biology of human disease before reading the paper than what the paper’s authors directly addressed in the science summarized in the paper. That, of course, makes both writing and reading comments about the paper much more fun, and I am still considering how much of what you are about to read is revealing my own biases and entrenched beliefs, rather than a straightforward interpretation of what this papers adds to our understanding of the biology of human disease. That said, my first reaction to the swirling discussions of the paper was to wonder what the excitement was all about. How would anyone who understands the nature of polygenic liability and is aware of what we have been learning about the contribution of regulatory variation to common disease heritability think otherwise? Given the sample sizes required to detect genome-wide significant associations of the variants making more modest contributions to the genetic liability to a common disease, it is clear we will never unambiguously identify many of them — never discern the biology that they might tell us about. But I don’t think that is the distinction that Boyle *et al.* were making by distinguishing “core” genes from others. If I have any disagreement with the paper it might be with the notion of core genes. What is a core gene for height? Even for the new biology that has been illuminated through GWAS discovery, it might be premature to conclude that the newly characterized biology is “core” in any way that is not true for other biology. Are genes involved in autophagy core genes for Crohn’s disease? Are genes involved in immune biology core genes for Crohn’s disease? Genes involved in inflammatory biology? Is the definition of a core gene simply that you can identify the big picture biological mechanism determining how variation at that gene affects risk of disease, or must that mechanism be the most important of the biological mechanisms affecting disease risk? Is it sufficient that it be one of the several most important biological mechanisms? A huge part of the problem with any of these definitions is that we remain ignorant of the bigger picture biology in which most genes function. If we have not yet understood the biology in which GWAS identified genes are functioning, does that exclude them from being core genes, or just from being identified as core genes? And, yes, I am skipping over the whole difficult question of how SNP signals will be correctly annotated to the gene or genes whose biology we are ultimately trying to get at, although that is indeed a major scientific pre-occupation of mine for many years. I remain uneasy that the concept of core genes is just the latest oversimplification of polygenicity. It is not that polygenicity is incompatible with the concept of core genes; it is that I cannot conceive of a sensible definition of “core”



<http://jpbs.qingres.com>

 OPEN ACCESS

DOI: 10.20900/jpbs.20170014(S1-S8)

Received: July 20, 2017

Accepted: September 14, 2017

Published: October 12, 2017

Copyright: ©2017 Cain *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

that we could actually define with the information we have today. Is “core” defined by the genes with just the top signals, or by the overall weight of the contributing factors related to that biology? I am in vehement agreement with any statement to the effect that what we most need to do in genetic studies of human disease is to use genetics to learn what deeper biological mechanisms impact disease risk and progression. Thus, it is less about which variants are “causal”, or even what genes are impacted by those variants, than it is about what bigger picture biology is affected by those genes. All of our efforts trying to get to variants and genes are essentially the means to that end. But I think we want to know all of that biology that we can. So if the essence of what they are saying is that not all of the variation that contributes to disease will feature in leading us to an understanding of primary driving biology, I can

agree that is likely to be true, and not very surprising to anyone who has considered the full complexity of the infinitesimal model. Indeed, I think it is a very great service to remind people of that fact, and that it is critical to be keeping our eyes on the long-term goals, as it is at least theoretically possible there will be a variety of -omics short cuts to those goals that may finesse the intermediate questions around individual variants and even genes. But I think we want to be careful about implying less complexity than there is likely to be; I would argue that many of our least productive arguments and controversies in human genetics have focused on the oversimplified extremes of generally sound paradigms. Kudos to Boyle and colleagues for making us think and discuss these issues! I am very much looking forward to reading their responses — it is after all their ink blot.