Qingres | JPBS

The Omnigenic Model: Response from the Authors

Evan A Boyle[1], Yang I Li[1, 2], Jonathan K Pritchard[1, 3, 4*]

[1] Department of Genetics, Stanford University;

[2] Section of Genetic Medicine, University of Chicago;

[3] Department of Biology, Stanford University;

[4] Howard Hughes Medical Institute, Stanford University.

*Correspondence: Jonathan K Pritchard, Email: pritch@stanford.edu.

We thank all of the commenters—and many others who have discussed the paper online or directly with us in person—for their thoughtful responses and critiques of our paper.

One of our goals in writing this paper was to highlight an apparent paradox in human genetics: most of the heritability for a typical complex trait is driven by genetic variation at loci that seem unrelated to the trait in question. We argued that the lack of a clear explanation for this seeming paradox is a major conceptual gap in modern human genetics, and we proposed one possible model to fill the gap. More generally we believe this is a question the field needs to grapple with.

Historical views of complex trait architecture

One of the key lessons from genetic studies of model organisms and human monogenic disease has been that specific mutations can drive organismal phenotypes, and that it is possible (at least in principle) to dissect precise mechanisms by which each mutation acts. Prior to the GWAS era, many researchers conceptualized complex traits in a very similar paradigm. In contrast to monogenic traits, they expected that complex traits would be driven by variants in multiple genes, each with proportionally smaller effect sizes. Nonetheless, there was a clear expectation that if those genes could be found, they would lead directly to disease-relevant biology.

However, in the past decade many groups have shown that typical complex traits are hugely polygenic, such that (1) the largest-effect variants confer only modest risk and, together, explain only a small fraction of the heritability, and (2) huge numbers of variants make non-negligible contributions to heritability. In our review we further emphasized several related points: (3) the signal is spread surprisingly broadly across the genome—so that, for example, most 100kb windows contain variants that measurably affect height; (4) there is only weak enrichment of heritability in genes with putatively relevant gene functions; and (5) while signals are strongly enriched in chromatin that is active in relevant cell types, there is little difference between the enrichment of cell type-specific chromatin vs. generically active chromatin such as constitutive promoters of housekeeping genes.

Thus, since around 2006, our shared understanding of the architecture of complex traits has been completely transformed. Yet there has been little or no discussion in the literature about conceptual frameworks for understanding this new reality. It is no longer tenable to expect that every variant that affects disease risk has a straightforward mechanistic effect on disease. In our view this implies a need to rethink how we conceptualize the cellular and organismal processes linking genetic variation to complex traits and diseases.

The omnigenic model and alternatives

In Boyle *et al.*, we coined the term "omnigenic" to refer to the hypothesis that "essentially any gene with regulatory variants in at least one tissue that contributes to disease pathogenesis is likely to have nontrivial effects on risk for that disease. Furthermore, the relative effect sizes are such that, since core genes are hugely outnumbered by peripheral genes, a large fraction of the total genetic contribution to disease comes from peripheral genes that do not play direct roles in disease."

Nancy Cox and Barbara Franke commented that the existing term "polygenic" already encompasses the omnigenic extreme. That may be true, but in our experience "polygenic" means different things to different people, and may be used to imply anything from 20 genes up to and including omnigenic, depending on context. We believe there is value in a new term that refers precisely to the extreme scenario where essentially every expressed gene can contribute to a trait.

The other key aspect of the omnigenic model is that we proposed a mechanism: namely that "gene regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes".

In our view, the connected-network aspect of the omnigenic model is a parsimonious model that can potentially explain the major observations, although we agree with Xin He that more data are needed to understand the nature of long-range interactions in cellular networks. While these types of data are currently very limited, new technologies for high throughput knockdown screens and expression profiling may soon plug this gap.

At the present time, however, the best hints about network effects of variation come from data on cis and trans-eQTLs. Various studies have shown that most of the heritability in gene expression is determined by trans effects. But (in contrast to cis-eQTLs) trans-eQTLs are very hard to find—suggesting that there must be huge numbers of trans-eQTLs, each with small effect sizes. We suspect that the networks underlying trans-variation in gene expression likely have heavy overlap with the networks linking peripheral genes to complex traits.

We also welcome other conceptual models that may explain the omnigenic scenario. In this context, Xin He proposes that some variants may affect cellular states and thereby affect disease risk, rather than affecting disease risk through core genes per se. We see this model as a plausible (and non-exclusive) alternative to the network model—indeed our reference to Preininger *et al.* (2013) in the Future Directions paragraph of Boyle *et al.*, was intended to refer to this type of idea.

Elliot Gershon and Ney Alliey-Rodriguez discussed the possibility that multiple diseases that are etiologically distinct may be clumped together as a single diagnosis such as schizophrenia. This may increase the apparent complexity of the diagnosis. We agree with this point to some extent—however it seems unlikely to us that by merging together a small number of discrete traits we can plausibly account for the huge number of involved genes observed in GWAS data, if the underlying traits themselves do not already span most of the genome.

Definition of core genes, and the utility of GWAS

Another question raised by Nancy Cox and others is what exactly we mean by "core" genes. Barbara Franke suggests an interesting hierarchy in the context of psychiatric diseases.

In Boyle *et al.*, we tried to leave the definition of core genes open, because it seems unlikely to us that a single definition can cover all cases in all complex traits. But if a precise definition is needed, we suggested that core genes may be defined as the (minimal) set of genes such that "conditional on the genotype and expression levels of all core genes, the genotypes and expression levels of peripheral genes no longer matter." We anticipate that in the near future, the combination of GWAS data and expression data in large case-control samples will enable tests to distinguish core and peripheral genes by this definition.

Francis McMahon, and others in the community, have pushed back against our paper as being unnecessarily gloomy, or overly critical of the

GWAS enterprise. But this is not our intention. It is reasonable to expect from our model (and from GWAS experience thus far) that many of the top hits in a genome-wide scan do indeed implicate core genes, even though they only contribute a small fraction of the variance. In various cases, functional follow-up of top hits from GWAS studies has indeed led to important biological insights, as may be expected for core genes. In their comments, Chunyu Liu and Stephen Faraone also touch on the challenges and opportunities in using core genes for drug discovery.

What remains unclear is what fraction of the top hits are core genes. Furthermore, we need better understanding of, and better tools for measuring, cellular networks to effectively triage peripheral genes that have large GWAS effect sizes. We expect that this will lead to functional follow-ups of genes that matter.

Summary

We contend that the findings of the last decade of GWAS require us to rethink the biological processes that link genetic variation to complex traits. The intuition gained from classical genetics, in which mutations or variants do something to cause disease, is not especially helpful for understanding the bulk of the heritability for complex traits. We believe the GWAS data are hinting at important new principles of biological organization, and how genetic perturbations percolate through cellular systems. Understanding these must be a major goal for human genetics in the coming years.