Effect Size Does Matter
The Long Road to Mechanistic Insight From Genome-Wide Association

Calum A. MacRae, MD, PhD; Martin R. Pollak, MD

In the last decade, several thousand genome-wide associations (GWAs) have been described, and the technique has proven to be a remarkably robust approach to the identification of common alleles that contribute to disease phenotypes. However, to date, few of these highly statistically significant associations have translated into mechanistic insights. This translational gap is a consequence of several factors, each of which is relevant to a variable extent for any specific association. Together they conspire to prevent a reproducible trajectory from genotype to phenotype. To some degree, the translational gap has been partly obscured by the tendency to associate GW A signals with neighboring genes whose presumed function makes them appealing as candidates. However, the very nature of the GWA approach and emerging data on the consequences of too rapid a jump to conclusions suggest that the ‘nearest gene’ and ‘nearest appealing gene’ strategies are unlikely to offer generalizable solutions.

Perhaps the most obvious hurdle to establishing causality is the fact that in most instances, the effect sizes at individual loci are in the range of 1.1- to 1.2-fold changes in the prevalence of disease or the magnitude of a quantitative trait. Such small effect sizes are difficult to model in experimental systems in which typical biological noise cannot be overcome even with enormous sample sizes. In many instances in which putative biology has been inferred from the study of local genes, the published observations are a consequence of loss of function or overexpression of a candidate gene and may not in any way reflect the subtle biology of the locus.

The GWA study design is an efficient means of detecting linkage disequilibrium but does not allow the definition of locus boundaries. Indeed, so-called fine mapping may facilitate the detection of the most tightly associated variant at a GW A locus but does not readily allow the localization of the causal variant that, depending on the population history of the local haplotype, may lie outside the most common linkage disequilibrium block in the cohort under study. Indeed, linkage disequilibrium mapping may reflect ancestral alleles present in only a very small proportion of the population. For example, the differences in haplotype blocks in African Americans with and without kidney disease helped lead to the identification of the ApoL1 gene as the underlying causal factor in the association and in turn led rapidly to evolutionary and mechanistic insights. It is likely that different natural selective pressures conferred by different environmental circumstances will be found to have altered local linkage disequilibrium at many disease-influencing loci. The fact that the majority of GWA signals lie in intergenic regions has likely prevented many misattribution events, but largely speculative word association is able to act across many megabases of intervening DNA.

The difficulties in moving from genetic markers to mechanism are only amplified by our limited understanding of the functions of nonprotein coding genes. Notwithstanding recent advances in our understanding of the regulatory functions of the genome, there is still no systematic method for moving from a sequence change to a functional insight. Clearly, a single nucleotide may act in many different ways, on many different genes, and at many different times or in different cells, so definitive attribution of a population effect is likely to require a very detailed understanding of the biology of each variant. To date, the correlation between common variants and peripheral or tissue-specific expression quantitative trait loci (eQTL) has identified some potential mechanistic connections but does not appear to offer a generalizable solution. More extensive tissue-specific eQTL data are accumulating. Newer experimental approaches that begin to define at a genome level the interactions between different regulatory sequences will likely shed light on the function of some variants, at least in some domains. However, there may be many remote effects in cis or trans that are undetected using current assays. The lack of conservation of regulatory sequences (although not necessarily of regulatory functions) in relevant model organisms has also placed an undue burden (and perhaps too much faith) on experiments in human inducible pluripotent stem cell–derived differentiated cell lines. Although such experiments may offer remarkable insights, they are biased toward cell-autonomous effects in single differentiated cell types that reflect only a small fraction of the biology of the homeostatic pathways that are likely to be detected in GWA.

Ultimately, system-level insight into how individual loci interact with each other across multiple mechanisms is likely to be necessary. Several lines of evidence point to the possibility that GWA loci are likely misconceived as individual factors: the modest effect size, the limited proportion of inherited variance explained by additive models, examples of extensive...
gene coregulation through effects in *cis* or *trans*, and the identification of the same loci in numerous unrelated traits. Several computational strategies can inform investigators using extant data on potential biological links across multiple loci for the same trait. These strategies are not constrained by artificial genomic boundaries, can help to identify genes with closely related biology even if they are at the same locus, and inform potential intermediate cellular phenotypes. Nevertheless, one is still left with the fact that traditional scientific logic would like to establish that changing a specific aspect of the function of a gene or gene network is both sufficient and necessary to explain the observed population effect. It is likely that this degree of rigor may never be feasible with small effect sizes. Perhaps the most reproducible approach may turn out to be efforts to complete comprehensive genetic architectures for specific traits through combinations of common alleles, larger-effect-size alleles from mendelian loci, or resequencing studies and formal screens in model organisms.

What then might the field consider reasonable evidence for the causal role of a gene in a specific GWA signal? In the first instance, there should be some unbiased rationale or direct experimentation that links the genetic signal with the gene or genes being studied. It is important to consider that the causal variant may lie well outside the linkage disequilibrium block in which the primary GWA signal sits. Once a gene or genes have been identified, evidence of a relevant biological activity for the specific genomic change (in contrast to extreme loss of function or overexpression) in an in vitro or in vivo assay will be necessary to allow the chain of causation to be constructed. The final step must be an attempt to demonstrate through this intermediate assay that the directionality and magnitude of effect are consistent in the original human cohort. Indeed, the correct in vivo biology should allow investigators to determine the proportion of the observed effect of the locus on the phenotype that is explained by the intermediate functional phenotype. For peripheral expression quantitative trait loci, this may be readily feasible, but in most other instances, this will require assay development and translation back into the original cohort. This last step alone will require the assembly of cohorts amenable to iterative deep phenotyping.

**An Example From Atherosclerosis**

In the current issue of *Circulation*, Almontashiri and colleagues build a case for disruption of specific transcription factor (TF) binding sites as part of the causal mechanism by which the 9p21 locus mediates its effects on coronary artery disease. The authors used a computational approach to predict those variants in the risk haplotype that perturb known TF binding sites or might create de novo TF binding sites. They identified 2 single-nucleotide polymorphisms that would be predicted to disrupt TEA domain (TEAD) binding sites, and then proved that the these single-base-pair substitutions are sufficient and necessary to interfere, at least in human aortic smooth muscle cells in vitro, with TEAD induction of p16, a protein with a reduced expression that has already been associated with the same risk allele. Finally, the authors were able to show that transforming growth factor-β is not able to induce p16 in aortic smooth muscle cells bearing the homozygous 9p21 risk allele or to inhibit the proliferation of these cells. This large a body of work succeeds in building a robust association between some of the variants in the risk allele and biology that was previously implicated in atherosclerosis. However, the final steps that would be necessary to complete the causal chain are, through no fault of the authors, unattainable. The location and timing of the TF effects and the subsequent demonstration that these effects are sufficient and necessary to generate a specific component of the modest effect observed in a human population cannot be directly undertaken. Model organisms do not represent these specific noncoding sequences. Fundamentally, because there is no definitive way to refute the hypothesis that this specific observed effect of these alleles contributes to the observed allelic risk, it is difficult to discriminate true from false functional associations.

This problem is intrinsic to the GWA study design because in the absence of definitive locus boundaries it is difficult to associate individual observations with population biology. Perhaps the most salutary conclusion is that it is possible only to relate the genetics to human disease through pathways that are already known to be involved in the disease in question. This philosophical point suggests that only through more comprehensive multidimensional phenotypic annotation of the genome will we be able to begin to make mechanistic inferences for most alleles of smaller effect size.

**Do We Need a New Direction in GWA?**

It is possible that the GWA study design happens to select for alleles that function in a very particular manner and that as a result there will emerge broadly generalizable solutions that will allow investigators to move efficiently from locus to mechanism. However, experience even with Mendelian loci teaches that many genes within a locus may affect a specific phenotype in an experimental system, yet typically only 1 gene is the actual cause. It will require a much larger experience in GWA studies to allow broader inferences to be drawn, but with tens or even hundreds of loci with very similar effect sizes, it may be difficult to discriminate signal from noise. One must also consider the distinct possibility that some complex traits may be beyond the limits of genetic dissection. A back-of-the-envelope calculation based on the conservative estimate that each of, say, half of the genes in the genome have 2 functionally distinct alleles would mean there are 2^1000 genetically distinct human biologies, vastly larger than the number of humans on earth.

In an era when we are rapidly moving toward ever larger GWA studies, these problems are likely only to be compounded. With larger cohorts, everything scales: the underlying etiologic heterogeneity, the diminishing effect size of the loci identified, and the cost. If we are to overcome such hurdles, we need to focus more coherently on effect size. Several of the most successful GWA studies have identified genome-wide signals with modest numbers of cases and control subjects. Although there are undoubtedly concerns about false-positive associations, independent replication is also feasible with modest numbers. Why were these and other smaller studies feasible? Successful smaller GWA studies appear to share certain attributes. First, the phenotypes were perhaps serendipitously quite uniform biological entities such as senile macular degeneration, lone atrial fibrillation, or renal failure in blacks. These conditions are relatively late in onset and do not appear to affect reproductive efficiency to any major degree. Finally, they each appear to reflect quite distinctive genetic
architectures, although our knowledge of the clinical genetic architecture of most human diseases is embarrassingly superficial. The possibility of exploiting modern electronic health records to capture proband-based kin cohorts that might systematize this study design is now within reach.

To move toward large-effect-size genetic studies in common disease requires more homogeneous core phenotypes. Indeed, this would likely also aid ongoing resequencing studies that have been rather disappointing, presumably because of the etiologic heterogeneity alluded to above. As we note, core phenotypes may be identified serendipitously, but there are several strategies that could be used to identify moderate- or even quite large-effect-size alleles that currently escape our detection. Systematically searching for ordered subsets has already been undertaken in efforts to salvage negative GWA studies. However, these post hoc efforts are constrained by the narrow focus of extant phenotyping. The structured collection of orthogonal traits that parse traditional diseases into biologically distinct subclasses is long overdue, although there are undoubtedly difficulties in identifying truly orthogonal traits. Such an effort would not only rescue precision medicine from an over-reliance on genetic data but also accelerate the annotation of the etiologic heterogeneity alluded to above. As we note, core phenotypes may be identified serendipitously, but there are several strategies that could be used to identify moderate- or even quite large-effect-size alleles that currently escape our detection. Systematically searching for ordered subsets has already been undertaken in efforts to salvage negative GWA studies. However, these post hoc efforts are constrained by the narrow focus of extant phenotyping. The structured collection of orthogonal traits that parse traditional diseases into biologically distinct subclasses is long overdue, although there are undoubtedly difficulties in identifying truly orthogonal traits. Such an effort would not only rescue precision medicine from an over-reliance on genetic data but also accelerate the annotation of the etiologic heterogeneity alluded to above.

Disclosures
C.A.M.'s work in this area is funded by NIH grants OD017870 and HG006500. C.A.M. and M.R.P are funded by NIH grant DK088826.

References
Effect Size Does Matter: The Long Road to Mechanistic Insight From Genome-Wide Association
Calum A. MacRae and Martin R. Pollak

Circulation. 2015;132:1943-1945; originally published online October 20, 2015; doi: 10.1161/CIRCULATIONAHA.115.019306

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/132/21/1943

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/