Editorial

Next-Generation Genome-Wide Association Studies
Time to Focus on Phenotype?

Calum A. MacRae, MD, PhD; Ramachandran S. Vasan, MD

As investigators plan the next round of genome-wide association studies (GWAS), cohorts of more than 100,000 individuals are being proposed as the solution to the “missing heritability” from first-generation studies. Al-though such studies will undoubtedly reveal many additional common alleles contributing to human disease, consideration of the intrinsic design of GWAS, our knowledge of the genetic architecture of disease and the results of GWAS to date suggests that complementary strategies will be necessary to uncover all of the heritability.

First-Generation GWAS
The outcomes of GWAS to date were largely anticipated. By design, the approach offers an unbiased assessment of the role of common alleles in disease syndromes, and there are numerous examples of successful GWAS in which multiple loci have been unequivocally associated with specific diseases. Even in genetically heterogeneous syndromes with major mendelian contributions, shared common alleles that operate downstream to affect trait expression may be detectable. In less heritable traits, GWAS may offer the only evidence of any genetic contribution. Many of the most successful studies to date have explained only a proportion of the heritability, whereas for some key disease phenotypes no loci have been identified that achieve genome-wide statistical significance. Serendipitous identification of large-effect size common alleles has occurred, but these may represent common modifiers of multiple mendelian genes, genes with large interactions with environmental factors or simply very homogeneous phenotypes.

Although the “small effect” alleles identified in GWAS may be important at a population level, a rigorous understanding of the role of such contributions will require much more comprehensive definition the genetic architecture for each trait. Epistatic interactions among common alleles, or between common alleles and a range of rarer mendelian alleles that vary between individuals, may contribute substantially to the unexplained heritability. For the majority of human traits, information on genetic architecture is scant because of the absence of unbiased family studies.

The simultaneous identification of multiple novel GWAS loci has highlighted a need for innovative approaches to define the causal gene(s) at each locus and then to explore the fundamental mechanisms of disease. Pathway entry has not yet been achieved for most GWAS loci. This is a result of the barriers to identification of the causal genes responsible for the association signal, including the lack of definitive locus boundaries, small effect sizes, and limited understanding of the regulatory functions of intergenic regions in cis or trans. Even where biological effects have been identified, it can be difficult to define the proportion of the effect at each locus that is attributable to a specific mechanism.

These predictable outcomes from GWAS serve to reinforce the implication from both human mendelian genetics and clinical medicine that many common disorders subsume substantial unidentified etiologic heterogeneity. Clearly, this etiologic (genetic) heterogeneity must be resolved if the comprehensive architecture of these traits is to be understood. In this context, increasing the size of human disease cohorts is likely only to scale the heterogeneity in parallel. Although such enormous studies will offer the statistical power to detect larger numbers of alleles of even smaller effect size, they are not likely to be powered to delineate gene-gene or gene-environment interactions.

Next-Generation Phenotypes
At the core of many of the problems with GWAS and underlying much of the missing heritability is the issue of phenotype resolution. Most common diseases phenotypes suffer from remarkably low resolution and imprecision. Indeed, phenotype resolution probably is a major determinant of the success or failure of GWAS to date. For traits where the phenotype is precise or serendipitously homogeneous (eg, macular degeneration), large effect sizes have been observed for select loci, and GWAS in small cohort have proven successful. In contrast, in situations in which the phenotype is less precise or subject to greater confounding (eg, blood pressure), GWAS have had limited success even with impeccable design and large cohorts. Clearly, selection pressures play a role in the genetic architecture for a given phenotype, and, for phenotypes in which there is little selection pressure on reproductive efficiency, genetic heterogeneity is less likely. However, in the setting of more stringent selection, even quite precise phenotypes may exhibit extensive genetic and allelic heterogeneity.
New diagnostic tools to discriminate homogeneous disease subsets and to identify causally related but more penetrant “endophenotypes” are urgently required.17–20 “Unbiased” phenotypes are emerging from model organism genetics, but large-scale phenotype discovery in human disease has not been undertaken.21–23 For many disease syndromes, the addition of even a limited number of orthogonal phenotypes may suffice to resolve the underlying etiologic heterogeneity.15,19,20,24 Functional genomic profiling of serum or tissue offers the possibility of discrete phenotypes, but, to date, the main successes have been in clonal neoplastic disorders. Innovative approaches to define etiology-specific functional assays range from tomographic electric mapping to metabolomics and cellular profiling.25–27 Maximizing the information content of phenotypic assays through dynamic testing will not only enable more rigorous genetics but will also facilitate systems level analyses in biology and medicine.28 Importantly, in order to redefine disease, candidate phenotypes must be readily scalable and robust enough for routine clinical use. Understanding the expanding “phenome” will require careful correlation with classic disease entities in well-characterized kindreds and in extended populations.18 This new wave of clinical investigation will exploit translational research facilities and major population studies such as the National Heart, Lung, and Blood Institute–funded Framingham Heart Study, ideally capturing multiple phenotypic axes in parallel. Many different phenotypes might be used in downstream integrative studies, but efficient translation to and from model systems is likely to be a critical attribute for the empirical definition of genetic architecture and for pathway entry or exploration.29

Phenotypic homogeneity may be conferred by simple orthogonal features, such as the presence of an additional associated trait or the restriction to a very specific demographic subgroup. Indeed, where true etiologic homogeneity exists, surprisingly small cohort studies may be sufficient to detect an underlying common allele. In this issue of Circulation: Cardiovascular Genetics, Horne et al30 present a tantalizing report of GWAS in a cohort of only 40 patients with peripartum cardiomyopathy (PPCM). In this small cohort, they were able to identify a locus with genome-wide significance on chromosome 12, near the gene encoding parathyroid hormone–like hormone. This locus was validated in a second series of PPCM patients, as well as in a cohort of individuals with pregnancy-associated cardiomyopathy that did not meet criteria for PPCM. Should we ignore these results simply on the basis of the size of the discovery cohort? Although it is conceivable that, in this condition and in this population, a single common allele is present and detectable in a study of this size, there are many reasons for caution. The limited number of unusual subjects brings with it high false-positive rates in GWAS, risks of population stratification, and problems with the identification of robust controls. Superimposed on these confounders is existing evidence that PPCM is not a homogeneous entity. Several reports suggest that PPCM is found in the context of several distinct autosomal dominant forms of dilated cardiomyopathy, and the data presented here are not inconsistent with a founder effect with reduced penetrance.31 All of these issues could be directly addressed by additional replication studies in diverse cohorts with PPCM, but it may not be possible to convince skeptics without a biological mechanism.

Next-Generation GWAS

Larger GWAS studies will be performed in the near future, but the trade-off between statistical power and the effect size may limit their utility. Complementing these studies in the search for missing heritability will be next-generation mendelian studies exploiting small kindreds and whole-genome sequence analysis. However, to fully define the genetic architecture of disease, it will be vital to add hybrid approaches capable of identifying alleles with intermediate effect sizes, and ultimately powerful enough to detect gene–gene and gene-environment interactions. Deconvoluting the genetic basis of complex disease at this resolution will require the integration of population- and family-based studies in kin-cohort type designs32 with novel multidimensional and dynamic phenotypes that can be translated efficiently to and from model organisms for parallel pathway exploration.

Sources of Funding

Dr MacRae was supported in part by National Heart, Lung, and Blood Institute grant HL098938; Dr Vasan was supported in part by National Heart, Lung, and Blood Institute contract NO1-HC 25195.

Disclosures

None.

References

3. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447:661–678.


Next-Generation Genome-Wide Association Studies: Time to Focus on Phenotype?
Calum A. MacRae and Ramachandran S. Vasan

doi: 10.1161/CIRCGENETICS.111.960765
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2011 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/4/4/334

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Genetics 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: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org//subscriptions/