Editorial

Large-Scale Genome-Wide Association Studies Consortia: Blessing, Burden, or Necessity?

Erik Ingelsson, MD, PhD

In the current issue of *Circulation: Cardiovascular Genetics*, Preuss and colleagues describe the design of the Coronary ARtery Disease Genome-wide Replication And Meta-Analysis (CARDIoGRAM) consortium. CARDIoGRAM is one of several consortia that have formed in the past few years in order to combine data from different genome-wide association studies (GWAS) in large-scale metaanalyses. Many of these consortia have focused their efforts on traits from the cardiovascular and type 2 diabetes domains. In addition to CARDIoGRAM, other examples of trait-specific consortia in this field include the Meta-Analysis of Glucose and Insulin-Related Traits Consortium, the Global Lipids Genetics Consortium, the Diabetes Genetics Replication and Meta-analysis consortium, and the Genetic Investigation of Anthropometric Traits consortium. There are also examples of consortia that have a wider scope, with phenotypes ranging over the whole cardiovascular field, such as the Cohorts for Heart and Aging Research in Genomic Epidemiology, European Network of Genomic and Genetic Epidemiology, and Candidate Gene Association Resource consortia. The large-scale GWAS consortia have many things in common, such as: (1) rigorous and robust methods with greater statistical power than single-center studies, as exemplified in the current article from the CARDIoGRAM consortium; (2) researchers being involved in several consortia, leading to cross-fertilization, quicker development and implementation of new methods; and (3) study samples being used for different traits in different consortia for increased cost-effectiveness.

The foremost impetus for this development is science: the first successful example of a GWAS reporting an association between the *CFH* gene and age-related macular degeneration in 2005, via the landmark study from the Wellcome Trust Case Control Consortium in 2007, to the current situation has been quite remarkable. Independently of whether one thinks that the GWAS era has been a blessing or a burden for the cardiovascular genetics field, it is hard to argue against the fact that the number of robustly and consistently replicated genotype-phenotype associations has gone from almost none during the period of candidate gene-based studies to increasing exponentially in the current period of larger and larger GWAS metaanalysis consortia. However, this tremendous increase in the number of robustly replicated loci has not convinced all about the usefulness of the GWAS methodology. Common criticisms include that the genetic variants found so far only explain a tiny fraction of the total heritability and that the mechanistic understanding or clinical applicability in terms of risk prediction or treatment have been limited thus far. Some critics even go so far as to consider the whole GWAS era a total failure because expectations of “individualized medicine” have not been realized in the short run relative to the substantial economic investments in this line of research. In my opinion, a more nuanced way of looking at it is to note that the GWAS design (and metaanalyses of GWAS) has been extremely efficient in establishing associations between common gene variants and phenotypes, which may help us to understand more about genes and mechanisms involved in these phenotypes. It is not realistic to expect full mechanistic knowledge within 2 to 3 years after the first GWAS studies. Only through laborious and time-consuming work, including downstream studies of gene and protein expression, use of cellular or animal models, detailed human physiological studies, and other designs for further characterization, will we understand the full value of findings from the GWAS era. That said, there are already several good examples where the initial GWAS discoveries have been followed by studies characterizing the mechanisms in detail, improving our understanding of the biological processes involved.

The CARDIoGRAM consortium consists of multiple study samples with GWAS data, where, in most cases, researchers have published their own results first and then joined forces in larger and larger constellations. This development is shared with other trait-specific consortia where individual study investigators first published their own results (provided that the sample was large enough and there was anything novel and robust to report) and then joined forces with some other studies in a consortium, which in turn have merged with other consortia studying the same trait. This development of gradually increasing consortia sizes with a growing number of collaborators and an escalating number of study participants has for many cardiovascular traits led to (or is about to lead to) the development of one “megaconsortium” per trait, often with discovery stages of >100,000 DNA samples. So, what have been the driving forces behind this development? The foremost impetus for this development is science: the drive to explore and make novel scientific discoveries. The
majority of these GWAS consortia are investigator initiated, often without any dedicated funding, where different groups of researchers have come together and decided to collaborate in the interest of science (and, of course, to benefit in terms of publications, funding, and other worldly matters). The simple question as to why this has happened is that it just had to happen in order to identify the large number of common loci with small effects that are involved in the pathophysiology of complex diseases. In other words, the large-scale GWAS consortia turned out to be a necessity for performing this kind of research.

Another reason that this field has moved along so rapidly in the past few years is the opportunities it has provided to younger investigators. In a new research area with novel methods, the systems often can be less rigid, providing young investigators opportunities that they would not have had otherwise. In the case of different GWAS consortia, a substantial proportion of the hands-on work with data checking and quality control, meta-analyses, and drafting of manuscripts has been done by postdoctoral fellows and early career investigators. For this group of researchers in the early stages of their careers, the large-scale GWAS consortia have provided excellent opportunities for exciting research collaborations, exchange of experiences, contact with and travel to other groups, countries, and cultures, and chances to step up and take a lead in large projects that often also have ended up in top-tier journals. The involvement of early career investigators has most certainly increased the flexibility, swiftness, and development of the field. Additionally, the more senior investigators obviously have played an important role in this development not only by both giving their younger colleagues these opportunities and by mentoring them in their endeavors, but also by their leadership in coordinating and facilitating collaborations within the consortia environment.

Now, what can we expect to see in the next few years in terms of large-scale GWAS consortia and population genetics in general? One thing that definitely is happening is that the genetic consortia are growing even larger, as many additional study samples are currently being genotyped by GWAS technology or with the MetaboChip. There are several reasons for doing additional GWAS genotyping, including (1) the fact that many well-phenotyped study samples, as well as study samples of ethnicities other than those of northern European descent, have not been previously genotyped; (2) the evidence that probably hundreds of common genetic loci associated with most complex traits is gathering both from actual examples in the literature and from simulation-based methods; and (3) the falling prices for GWAS genotyping in response to the current interest in sequencing methods, which still are too expensive to do in a larger scale for most research groups. Additionally, many consortia now are incorporating data from the MetaboChip, a custom Illumina iSelect genotyping array designed to analyze cost-effectively ~185 000 single nucleotide polymorphisms identified through GWAS metaanalyses of cardiovascular and metabolic traits. These data, together with the new GWAS data becoming available, will further increase the data available to the genetic consortia, which should be expected to identify even more novel genetic loci. The Metabochip also is designed to study rare genetic variants in selected regions, and studies of rare variants with different methods, including resequencing, exon or whole-genome sequencing, or imputation based on data from the 1000 Genomes project (http://www.1000genomes.org/), are among the main avenues for population-based genetic research in the coming years. Even if the effect sizes may be larger, at least in theory, for rare variants than for common ones, it is nevertheless likely that studies of rare variants also will need to join forces in consortia to boost statistical power given that the variants are rare in the population.

In summary, the development of large-scale GWAS consortia has resulted in an impressive number of robustly replicated associations between common variants and cardiovascular phenotypes. I would call these collaborations neither a blessing nor a burden but, rather, a necessity driven by the scientific curiosity and energy of good scientists, younger as well as older. The nurturing of scientists who are used to and interested in working in an international and collaborative manner is likely to be beneficial to biomedical science in general. The different omics methods that are becoming increasingly available to research groups throughout the world provides exciting new opportunities, and the positive experiences of collaborative efforts in the GWAS field should influence our plans of how to best make use of these other kinds of massive data. It should not come as a total surprise if large-scale collaborations will be a necessity also for other large-scale methods and in epidemiological studies in general.

Disclosures

Dr Ingelsson is involved in many of the large-scale GWAS consortia in the cardiovascular field, including CARDIoGRAMPlus, a recent expansion of the CARDioGRAM consortium that occurred after the CARDioGRAM study described in this issue of Circulation: Cardiovascular Genetics. He is supported by grants from the Swedish Research Council, the Swedish Foundation for Strategic Research, the Swedish Diabetes Foundation, the Swedish Society of Medicine, the Novo Nordisk Foundation, and the Royal Swedish Academy of Science.

References


8. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007;447:661–678.


**Key Words:** Editorials • genetic association studies
Large-Scale Genome-Wide Association Studies Consortia: Blessing, Burden, or Necessity?
Erik Ingelsson

_Circ Cardiovasc Genet_. 2010;3:396-398
doi: 10.1161/CIRCGENETICS.110.958397
_Circulation: Cardiovascular Genetics_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2010 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/3/5/396

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/