Cardiovascular Genomics, Personalized Medicine, and the National Heart, Lung, and Blood Institute

Part I: The Beginning of an Era

Christopher J. O’Donnell, MD, MPH; Elizabeth G. Nabel, MD

The inaugural issue of *Circulation: Cardiovascular Genetics* arrives at a remarkable time in the history of genetic research and cardiovascular medicine. Despite tremendous progress in knowledge gained, cardiovascular disease (CVD) remains the leading cause of death in the United States,¹ and it has overcome infectious diseases as the leading cause of death worldwide.² In addition, rates of CVD remain higher in black and Hispanic populations in the United States.¹ The recent Strategic Plan of the National Heart, Lung, and Blood Institute (NHLBI) emphasizes research areas to fill the significant knowledge gaps needed to improve the diagnosis, treatment, and control of known risk factors and clinically apparent disease. Simultaneously, the NHLBI Strategic Plan recognizes a tremendous opportunity that is available for use of genetic and genomic research to generate new knowledge that might reduce the morbidity and mortality from CVD in US populations.³ Public availability of vast amounts of detailed sequence information about the human genome, completed sequence data on dozens of other animal genomes, and private sector development of high-throughput genetic technologies has transformed in a few short years the conduct of cardiovascular genetics and genomics research from a primary focus on mendelian disorders to a current emphasis on genome-wide association studies (GWAS; Figure 1). In this review, we describe the rationale for the current emphasis on large-scale genomic studies, summarize the evolving approaches and progress to date, and identify immediate-term research needs. The National Institutes of Health (NIH) and the NHLBI are supporting a portfolio of large-scale genetic and genomic programs in diverse US populations with the longer-term objective of translating knowledge into the prediction, prevention, and preemption of CVD, as well as lung, sleep, and blood disorders. Underlying this portfolio is a strong commitment to make available participant-level data and aggregate research results to the broad community of investigators, while protecting the privacy and confidentiality and respecting the informed consent of study participants.

**Rare Genetic Variation in CVD**

Much of our current understanding of the genetic predisposition of human diseases derives from detailed studies during the “pregenome” era of large pedigrees harboring rare, mendelian conditions. The critical role of clinical observation in such studies cannot be overemphasized, because the identification of rare conditions usually begins with the assembly of pedigrees in which an autosomal dominant or recessive pattern of inheritance is identified. Pedigree studies that used positional cloning approaches have led to a comprehensive body of information on the role of deleterious mutations, often in exons (the protein-coding region of human genes), which lead to altered protein function or number. These studies have uncovered rare (minor allele frequency generally <0.1%) genetic mutations in the sequence encoding key sarcomere and storage-disorder proteins underlying hypertrophic and dilated cardiomyopathy, in ion channels underlying long-QT syndrome and ventricular tachycardia, and in *FBN1* underlying thoracic aortic aneurysms and aortic dissection.⁴ Knowledge of causal mutations has contributed tremendously to our understanding of the pathophysiology of these conditions, and in some cases, such as familial hypercholesterolemia, it has led to the development of new treatments, such as lipid-lowering drug therapies, that are now widely used in general populations. However, although genetic testing for some of the above conditions is now available in Clinical Laboratory Improvement Amendments–certified laboratories, in most cases, there is not yet a clear role in clinical practice for the genetic testing of mutations in cardiovascular diagnosis or treatment.

One example of remarkable progress in the past several years has been our understanding of the molecular genetics of the rare, premature aging condition, Hutchinson-Gilford progeria syndrome (progeria). Progeria results from a spon-
Timeline of NHLBI Population Genomic Studies for Common Diseases

Figure 1. Timeline of NHLBI genetic and genomic investigations for rare and common diseases, taking advantage of advances in the Human Genome Project and other technologies. ENCODE indicates Encyclopedia of DNA Elements.

Recently, investigators conducting sequencing studies in general populations have begun to estimate the contribution to common CVD risk of variants in genes known to underlie mendelian CVD or other candidate genes. For example, a recent sequencing study of genetic variation in exons of sarcomere protein genes and storage cardiomyopathy–causing genes was conducted for determinants of increased left ventricular wall thickness in \( \approx 1800 \) persons from the community-based Framingham Heart Study. There were 7 mutations found in 5 sarcomere protein genes and 1 mutation in the \( \alpha \)-galactosidase A (GLA) gene; at least 1 of these mutations was detected in 1% of all persons in the community and in 18% of those with increased left ventricular wall thickness.\(^9\) In a separate community-based sequencing study, at least 1 of 64 subjects was found to carry a functional mutation in 1 of 3 genes (\( NCCT, \) \( NKCC2, \) and \( ROMK) \) previously known to cause rare recessive diseases that feature large reductions in blood pressure.\(^\) In a sequencing study of the adipokine \( ANGPTL4 \) gene in \( >3500 \) subjects from the Dallas Heart Study, nonsynonymous (amino acid–altering) variants were found in 13 persons with low triglycerides versus 2 with high triglycerides, and more nonsynonymous variants were found in European Americans than in African Americans.\(^9\) The role of human \( ANGPTL4 \) has been unknown, although in mice, a genetic deletion of \( ANGPTL4 \) leads to lower plasma triglyceride levels, and overexpression of \( ANGPTL4 \) in the liver leads to hypertriglyceridemia. A similar approach led to the discovery of variants of \( PCSK9 \) that occur in \( \approx 3\% \) of the population and are associated with low LDL and low risk of incident coronary heart disease.\(^9\) Reliable estimates of the overall contribution of rare and uncommon variants to the burden of common CVDs await large-scale sequencing studies of not only genes implicated in mendelian forms of CVD but also all other known genes. Candidate-gene sequencing studies are under way for CVD candidate genes supported by both the National Human Genome Research Institute’s Medical Sequencing Program\(^\) and the NHLBI’s Resequencing and Genotyping Program.\(^\) Through a recent request for applications, the NHLBI will support the development of technologies to enable the accurate, high-throughput, genome-wide sequencing of all protein-coding gene exons and other important regulatory regions in large numbers of human subjects, with the ultimate goal of defining the association of such variants with cardiovascular, lung, and blood diseases.

As we await results of sequencing studies, our estimates of the role of mendelian versus more complex forms of inheritance in CVD derive primarily from available family history studies. One clue may be found in the results of a very large survey of 122,155 Utah families, which underscore the significant burden of familial risk to coronary heart disease and stroke but the relatively small contribution of rare conditions. A positive family history accounted for 14% of all families with coronary heart disease but nearly three fourths of persons with early-onset coronary heart disease; a positive family history of stroke was noted in 11% of families and accounted for 86% of early-onset strokes. On the other hand, only 1% of families carry a strongly positive family history, although these families account for 17% of cases of
early-onset coronary heart disease and 19% of cases of early-onset stroke. These data generate the hypothesis that rare “private” mutations in families with a strongly positive family history may account for a small but significant minority of the cases of early-onset CVD.

**Common Genetic Variation in CVD**

There is substantial evidence that heritable factors underlie the variation in clinical CVD, subclinical CVD, and its risk factors in human populations. However, although pedigree studies often provide strong evidence for a mendelian mode of transmission for rare familial CVD, the impact of a familial predisposition and the mode of transmission is often not apparent from studies of individual pedigrees for most cases of common, “complex” CVD, such as myocardial infarction. Studies of large groups of families or observational studies of family history have provided evidence for a familial predisposition to myocardial infarction, atrial fibrillation, and congestive heart failure. Heritability (the proportion of interindividual phenotypic variation that is attributed to genetic variation among individuals) can be estimated in groups of siblings or other related family members. There is substantial evidence from multiple studies for moderate heritability (eg, 30% to 50%) in the quantitative measurement of subclinical atherosclerosis, such as coronary artery calcification and carotid intima-media thickness. Similarly, there is moderate heritability for traditional risk factor measures, such as systolic blood pressure, total cholesterol, and body mass index. Thus, the totality of evidence from studies of the familial predisposition for CVD and from heritability studies of the independent precursors of CVD suggests that genetic variation underlies a substantial proportion of CVD and its risk factors.

The public availability of the complete human genome sequence brought greater attention to the hypothesis that common genetic variants may underlie most common, complex diseases such as CVD. Among the most common sources of genetic variation are single-nucleotide polymorphisms (SNPs), which occur approximately every 1000 base pairs in the approximately 3 billion base pairs that make up the human genome sequence. A complete, publicly available catalogue of SNPs is being updated by the NIH-supported international HapMap Project (http://www.hapmap.org/). Although there are likely >10 000 000 SNPs in the human genome, SNPs that reside near each other are highly correlated owing to a lack of recombination (also known as “linkage disequilibrium”) within chromosomal segments. From detailed studies of linkage disequilibrium patterns, it appears that genotype scans of 300 000 to 500 000 SNPs spaced across the entire genome provide sufficient statistical power to define the majority of common genetic variation. Greater numbers of SNPs are needed to define variation for populations of older ancestry, such as those of African ancestry, than for more recent populations, such those of European ancestry. Technology development and competition have allowed the genotyping of thousands of SNPs with high-throughput, multiplex SNP assays on genotyping chips.

**Genome-Wide Association Studies**

A GWAS is a study of genetic variation across the entire human genome designed to identify genetic associations with observable clinical traits or with the presence or absence of a disease. A first demonstration that a GWAS could be used to identify genes predisposing to common disease was reported in 2005, when a GWAS using a first-generation chip of 100 000 SNPs was used in a case-control study of fewer than 150 subjects to identify variation in the complement factor H (CFH) predisposing to age-related macular degeneration. The pooled estimate of multiple subsequent replication studies demonstrated unequivocally a consistent association of the CFH variant with age-related macular degeneration. Since 2005, technology development and price reductions have led to several generations of DNA chips with higher densities, ranging from 300 000 to >1 000 000 SNPs. Between 2006 and 2007, approximately 10 published GWAS reports had appeared. From 2006 to the present, >175 GWAS reports have been published in the peer-reviewed literature, reporting associations of >300 SNP variants with a wide range of disease phenotypes.

Although many GWAS for CVD are in progress and have yet to be reported, there are already several remarkable discoveries, through GWAS reports, of genetic variants associated with CVD and its risk factors. CVD outcomes studied by GWAS have included peripheral arterial disease, myocardial infarction, stroke, and atrial fibrillation. Other GWAS phenotypes have included traditional CVD risk factors (including body mass index and obesity, type 2 diabetes mellitus and fasting glucose, hypertension and systolic blood pressure, nicotine dependence, and levels of LDL, HDL, and triglycerides), QT-interval length on ECG, inflammatory and hemostatic biomarkers, imaging measures of subclinical atherosclerosis (including coronary artery calcium and carotid intima-media thickness), and left ventricular hypertrophy. A few selected GWAS reports highlight some surprising findings. For example, QT-interval length on ECG was among the first CVD phenotypes to be studied by GWAS, revealing a strong association with SNPs in the CAPON gene in a staged study that involved 2 independent cohorts and was subsequently independently replicated in another large cohort. Unlike most genes known to be associated with QT-interval length, CAPON does not encode an ion channel but rather is a regulator of nitric oxide synthase; its mechanism of prolonging the QT interval is unknown. In a second example, a recent GWAS was undertaken to identify genetic determinants of alterations in LDL cholesterol, HDL cholesterol, or triglycerides in 3 independent cohorts in which collaborators had recently identified new loci underlying type 2 diabetes mellitus. Using an *in silico* meta-analysis of GWAS data from all 3 cohorts, common SNPs from 18 loci were associated with at least 1 of the lipid phenotypes. Of these loci, 6 were unexpected and had not been implicated previously in lipid physiology. A risk score, which used 9 of the previously identified LDL- or HDL-related SNPs, was tested for association with incident CVD; the risk score discriminated as well as but not better than clinical risk factors. Finally, 3 independent GWAS studies recently reported strong evidence for an association of a region in
chromosome 9p21 near the cyclin-dependent kinase inhibitor genes CDKN2A and CDKN2B.37–39 However, the signal for association is in a region free of known protein-coding sequence, and interestingly, one new loci associated with type 2 diabetes mellitus is also in chromosome 9p21.37 Several other loci are implicated but have somewhat lower strength of evidence for association with myocardial infarction.40

The number of GWAS reports is increasing in an exponential manner, and several common themes are emerging. First, because of the risk of false-positive results from multiple testing, SNP associations must meet rigorous statistical thresholds to be considered significant on a genome-wide level. Remarkably, many of the initial GWAS association findings have been replicated, in a convincing manner, with modern, stringent guidelines.41 Second, the magnitude of most SNP associations in GWAS is modest (relative risks of 1.2 to 1.4), as is the proportion of variation explained by individual SNPs, which emphasizes the need for large sample sizes to discover true positive associations. Third, although a growing number of GWAS associations have confirmed previously reported associations found by use of other methods, many loci discovered in GWAS are located neither in or near protein-coding genes, nor are they near genes previously known to be associated with the disease of interest. Thus, it appears that some genetic associations may be attributed to genetic mechanisms other than alterations in the sequence of protein-coding genes. Of interest, in a recent analysis of genetic mechanisms other than alterations in the sequence of protein-coding genes, many loci discovered in GWAS are located neither in or near protein-coding genes, nor are they near genes previously known to be associated with the disease of interest. Thus, it appears that some genetic associations may be attributed to genetic mechanisms other than alterations in the sequence of protein-coding genes. Of interest, in a recent analysis of genetic mechanisms other than alterations in the sequence of protein-coding genes, many loci discovered in GWAS are located neither in or near protein-coding genes, nor are they near genes previously known to be associated with the disease of interest. Thus, it appears that some genetic associations may be attributed to genetic mechanisms other than alterations in the sequence of protein-coding genes.

### Table 1. NHLBI Programs in Genomic Research

<table>
<thead>
<tr>
<th>Program</th>
<th>Objective</th>
<th>Subjects and Population(s)</th>
<th>Selected CVD, Lung, and Blood Phenotype(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CARE</td>
<td>GWAS in &gt;10 000 African-American subjects</td>
<td>GWAS in &gt;10 000 African-American subjects from 5 cohorts</td>
<td>Aging, anthropometry, atrial fibrillation, blood biomarkers, blood pressure/hypertension, coronary heart disease/myocardial infarction, cardiac ultrasound measures/heart failure, diabetes, kidney disease, lipids, pulmonary function, sleep, stroke, subclinical atherosclerosis, peripheral arterial disease</td>
</tr>
<tr>
<td></td>
<td>Candidate gene association study of ~50 000 SNPs in ~2000 genes implicated in CVD</td>
<td>Candidate gene studies in &gt;48 000 subjects from 9 cohorts. Coefficients include: ARIC, CARDIA, CFS, CHS, CSSCD, FHS, JHS, MESA, and SHHS</td>
<td></td>
</tr>
<tr>
<td>STAMPEED</td>
<td>13 GWAS projects to identify genetic variants related to heart, lung, and blood disorders and their risk factors by utilizing GWAS in existing human studies</td>
<td>GWAS studies in multiple existing human studies totaling &gt;38 000 subjects</td>
<td>Aging, asthma, coronary heart disease/myocardial infarction, stroke, platelet function, hematopoietic cell transplant outcomes, subclinical atherosclerosis, sickle cell anemia outcomes, aging</td>
</tr>
<tr>
<td>SHARe</td>
<td>GWAS project in large NHLBI cohorts studies to identify genetic variants related to heart, lung, and blood disorders and their risk factors</td>
<td>Framingham Heart Study SHARe: GWAS in &gt;9400 subjects from 3 generations of the Framingham Heart Study SHARe Asthma Research Project (SHARP), a GWAS in ~5000 subjects from asthma cohorts</td>
<td>Aging, anthropometry, atrial fibrillation, blood biomarkers, blood pressure/hypertension, coronary heart disease/myocardial infarction, cardiac ultrasound measures/heart failure, diabetes, kidney disease, lipids, pulmonary function, sleep, stroke, subclinical atherosclerosis, peripheral arterial disease, asthma and asthma-treatment response</td>
</tr>
<tr>
<td>ENDGAME</td>
<td>Consortium of 11 research groups to develop and test innovative, informative, and cost-effective study designs and analytical strategies for performing GWAS on complex diseases</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Women’s Health Genome Study</td>
<td>GWAS to identify genetic variants related to CVD, cancer, and selected risk factors</td>
<td>Up to 28 000 female participants in the Women’s Health Study, a study cohort that has been followed up for more than a decade as part of a randomized, controlled study for prevention of CVD and cancer</td>
<td>Blood biomarkers, hypertension, coronary heart disease/myocardial infarction, diabetes, lipids, stroke</td>
</tr>
</tbody>
</table>

CARE indicates Candidate Gene Association Resource; ARIC, Atherosclerosis Risk in Communities; CARDIA, Coronary Artery Risk Development in Young Adults; CFS, Cleveland Family Study; CHS, Cardiovascular Health Study; CSSCD, Cooperative Study of Sickle Cell Disease; FHS, Framingham Heart Study; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; SHHS, Sleep Heart Health Study; STAMPEED, SNP Typing for Association with Multiple Phenotypes from Existing Epidemiologic Data; SHARe, SNP Health Association Resource; and ENDGAME, Enhancing Development of Genome-wide Association Methods.
pleiotropy. Finally, to date, there have been very few studies that convincingly describe the biological function or deleterious effect of SNPs or genes identified through GWAS.

**NHLBI Genome Programs and New Research Models**

The NHLBI supports a robust portfolio of GWAS programs that provide population diversity and great depth and breadth of clinically relevant phenotypes (Table). The NHLBI’s position is that the greatest public benefit will be realized if data from GWAS are made available, under terms and conditions consistent with the informed consent provided by individual participants, in a timely manner to the largest possible number of investigators. Accordingly, the NHLBI has played a leadership role in the development of a trans-NIH data-sharing policy for GWAS studies, effective January 25, 2008 (http://grants.nih.gov/grants/gwas/), and the NHLBI adheres strongly to this policy. Concurrent with development of the policy, the National Center for Biotechnology Information (NCBI) created a genotype-phenotype NIH database called dbGaP (database of genotype and phenotype). The SNP Health Association Resource (SHARe) Program is the first NHLBI genomic program that now has data available for biomedical researchers. The SHARe scientific resource draws from genotyping of the 3 generations of participants in the NHLBI Framingham Heart Study, which includes 550,000 SNPs in >9400 consenting participants. The phenotype component includes >16,800 clinical variables, including risk factors, biomarkers, subclinical imaging measures, and clinical CVD, as well as other chronic disease areas, such as bone density and dementia. Additional cohorts are planned for SHARe, including the SHARe Asthma Research Project (SHARP), STAMPEED (SNP Typing for Association with Multiple Phenotypes from Existing Epidemiological Data) consists of 13 grants awarded to support GWAS in a wide range of cardiovascular, lung, and related diseases. Nearly 40,000 male and female participants are included, with substantial representation from minority populations. In the Candidate Gene Association Resource (CARE), a GWAS is now under way in >10,000 African American subjects from 5 NHLBI-supported population cohorts to study a range of cardiovascular phenotypes. The ENDGAME (Enhancing Development of Genome-Wide Association Methods) program supports 11 teams of investigators in the development of innovative strategies and tools for planning, initiating, conducting, and analyzing GWAS. All strategies and tools developed through these awards will be made available to the scientific community. In its totality, with ongoing or planned GWAS, the NHLBI portfolio will study well over 100,000 research participants from multiple, diverse, and well-phenotyped disease cohorts and population cohorts.

Observations from GWAS to date and the robust portfolio of ongoing GWAS suggest a number of immediate next steps for such studies (Figure 2). First, the availability of data for sharing will increase the opportunities for a broad diversity of scientists to make new discoveries. Second, collaboration between multiple GWAS will be necessary to convincingly replicate genetic associations and exploit the power of individual GWAS. The aforementioned consortia of researchers studying the genomics of type 2 diabetes mellitus and the Wellcome Trust Case Control Consortium represent 2 successful models of genomics collaboration focused on 1 or a few diseases. In one newly formed consortium, the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) Consortium, GWAS results are being pooled systematically in a series of in silico meta-analyses from 5 prospective, population-based cohorts to study GWAS associations across >20 different phenotype areas in CVD, aging, and multiple other chronic diseases. Third, the functional significance of GWAS associations must be defined, which affords great opportunity for collaboration between bench scientists, statisticians, epidemiologists, and clinical investigators. Moreover, functional genomics requires animal and human tissue models and systems biology approaches that utilize proteomics, metabolomics, and gene expression profiling. Finally, as sequencing cost decreases and accuracy improves, deep sequencing of ever-larger proportions of the human genome will follow, which provides the opportunity to define the totality of high frequency, low frequency, and rare variations that underlie CVD. The scientific opportunities have never been greater to unravel the mystery of genetic predisposition to CVD.

**Conclusions**

The NHLBI supports a large portfolio of population-based genetic and genomic programs in diverse US populations. We are strongly committed to making participant-level data and aggregate research results available to a broad community of investigators, while protecting the privacy and confidentiality of study participants. An abundance of knowledge of the mendelian forms of CVD exists, yet the current explosion of GWAS is leading to a remarkable pace of discovery with regard to highly prevalent cardiovascular risk factors and common, clinically apparent CVD. These discoveries represent only “the end of the beginning.” The next phase of discovery will require deep sequencing of the human genome, functional research in animal and human models, and application to clinical medicine.

**Disclosures**

None.


Cardiovascular Genomics, Personalized Medicine, and the National Heart, Lung, and Blood Institute: Part I: The Beginning of an Era
Christopher J. O'Donnell and Elizabeth G. Nabel

doi: 10.1161/CIRCGENETICS.108.813337
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2008 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/1/1/51

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/