Genomics of Cardiovascular Disease
Christopher J. O'Donnell, M.D., and Elizabeth G. Nabel, M.D.

Cardiovascular disease is the leading cause of death in the United States. Considerable progress has been made in the past 50 years to define, identify, and modify risk factors for cardiovascular disease (e.g., hypertension, dyslipidemia, obesity, type 2 diabetes, cigarette smoking, and physical inactivity) and to develop treatments, such as coronary care units, percutaneous coronary interventions, and beta-blockers. These efforts have resulted in an age-adjusted decline in cardiovascular mortality. In addition, it is now possible to detect subclinical disease by means of blood biomarker testing or imaging measurements years before the onset of symptoms or other clinical manifestations. Despite this progress, mechanisms that underlie individual differences in the presentation and pathophysiological features of cardiovascular disease are poorly understood. In this article, we review genetic and genomic studies in cardiovascular medicine that have helped to elucidate some of these mechanisms during the past decade (Fig. 1, and interactive timeline, available with the full text of this article at NEJM.org).

Mendelian and Candidate-Gene Studies

Ten years ago, the draft sequence of the human genome, which was produced by scientists working on the Human Genome Project and others, was first described, leading to an expansion in the understanding of genetic contributions to cardiovascular disease. Before the Human Genome Project, many genes associated with mendelian cardiovascular disease had been identified. These forms of cardiovascular disease are rare and constitute a minority of clinical cardiovascular diseases. Genetically, they are simple in that a mutation in a single gene is sufficient to cause disease, so mendelian disease is said to be monogenic. Examples include forms of premature myocardial infarction, dilated and hypertrophic cardiomyopathy, heart failure, arrhythmogenic right ventricular dysplasia, the long-QT syndrome, and aortic aneurysms. Recessive mutations underlie familial forms of cardiovascular risk factors, such as hypertension, hypercholesterolemia, and type 2 diabetes. Knowledge obtained through the identification of genes associated with mendelian disease has led to breakthrough discoveries in mechanisms of cardiovascular disease and its treatment. A compelling illustration is the Nobel-Prize–winning discovery that mutations affecting the low-density lipoprotein (LDL) receptor cause hypercholesterolemia and early-onset myocardial infarction, which led to LDL cholesterol–lowering therapies that reduce the risk of cardiovascular events.

Genomics of Cardiovascular Risk and Disease

Role of Genomewide Association Studies

The large majority of cardiovascular diseases, however, are polygenic, with both heritable and environmental contributions. Moreover, there are heritable, polygenic com-
Figure 1. Timeline of Genetic and Genomic Research in Cardiovascular Medicine.

- 1973–1975: Mutations affecting the LDL receptor reported to cause familial hypercholesterolemia
- 1975: DNA sequencing described by Sanger et al. and by Maxam and Gilbert
- 1977: Discovery of introns
- 1980: Development of electrophoretic methods for DNA sequencing
- 1981–1982: First reports of transgenic mice and flies
- 1983: First disease gene mapped, for Huntington’s disease
- 1986: First disease gene (DMD, mutation of which causes Duchenne’s muscular dystrophy) discovered by positional cloning
- 1989: Microsatellites used as genetic markers
- 1990: Launch of Human Genome Project
- 1991: Expression of sequence tags (ESTs) used to describe fragments of genes
- 1993: Mutations in factor V ("Leiden") gene associated with resistance to protein C
- 1994: Mutant dystrophin implicated in dilated cardiomyopathy
- 1995: Discovery of several canonical Romano–Ward susceptibility genes for the long-QT syndrome (LQTS)
- 1996: Mouse genetic map completed
- 1997: Discovery of genetic variants causing susceptibility to Jervell and Lange-Nielsen LQTS
- 1999: Exon sequencing identifies mutations in the ANGPTL3 gene associated with familial hypertriglyceridemia
- 2000: Sequence of the human genome described
- 2002: Development of in situ sequencing
- 2003: Mutations in genes affecting glycosgen-storage disease genes shown to cause hypertrophic cardiomyopathy
- 2004: Detailed map of single-nucleotide variants reported by the International HapMap Project
- 2005: First report of a gene associated with a cardiovascular trait using GWAS: NOS1AP associated with prolonged QT interval
- 2006: Variants in genes encoding sarcomere proteins and affecting metabolism implicated in cardiac hypertrophy
- 2007: Mutations in PCSK9 implicated in lower levels of LDL and lower risk of coronary heart disease
- 2008: GWAS uncover variants in FTO associated with body-mass index and obesity
- 2009: Two reports of SNPs at 9p21 associated with myocardial infarction and related forms of coronary artery disease (revealed by GWAS)
- 2010: Wellcome Trust Case Control Consortium reports associations using GWAS of gene variants with seven conditions including type 1 and type 2 diabetes mellitus and coronary artery disease
- 2011: GWAS reveals association between gene variants on chromosome 4q25 and atrial fibrillation.

GWAS also implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.

GWAS implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.

GWAS also yield variants in MYH6 associated with sick sinus syndrome and variants associated with dilated cardiomyopathy and carotid intima–media thickness.

Large GWAS yields >30 gene loci associated with clinical coronary artery disease.

Large GWAS reports >30 gene loci associated with blood pressure and hypertension.

GWAS also yield variants in MYH6 associated with sick sinus syndrome and variants associated with dilated cardiomyopathy and carotid intima–media thickness.

NHLBI begins whole-exome and whole-genome sequencing of the Framingham Heart Study and other cohorts for studying cardiovascular disease.

GWAS implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.

Exome sequencing identifies mutations in SLCE6A3 for renal salt wasting in Bartter’s syndrome.

NHLBI begins whole-exome and whole-genome sequencing of the Framingham Heart Study and other cohorts for studying cardiovascular disease.

GWAS implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.

Exome sequencing identifies mutations in SLCE6A3 for renal salt wasting in Bartter’s syndrome.

Large GWAS yields >30 gene loci associated with clinical coronary artery disease.

Large GWAS reports >30 gene loci associated with blood pressure and hypertension.

GWAS also yield variants in MYH6 associated with sick sinus syndrome and variants associated with dilated cardiomyopathy and carotid intima–media thickness.

NHLBI begins whole-exome and whole-genome sequencing of the Framingham Heart Study and other cohorts for studying cardiovascular disease.

GWAS implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.

Exome sequencing identifies mutations in SLCE6A3 for renal salt wasting in Bartter’s syndrome.

Large GWAS yields >30 gene loci associated with clinical coronary artery disease.

Large GWAS reports >30 gene loci associated with blood pressure and hypertension.

GWAS also yield variants in MYH6 associated with sick sinus syndrome and variants associated with dilated cardiomyopathy and carotid intima–media thickness.

NHLBI begins whole-exome and whole-genome sequencing of the Framingham Heart Study and other cohorts for studying cardiovascular disease.

GWAS implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.

Exome sequencing identifies mutations in SLCE6A3 for renal salt wasting in Bartter’s syndrome.

Large GWAS yields >30 gene loci associated with clinical coronary artery disease.

Large GWAS reports >30 gene loci associated with blood pressure and hypertension.

GWAS also yield variants in MYH6 associated with sick sinus syndrome and variants associated with dilated cardiomyopathy and carotid intima–media thickness.

NHLBI begins whole-exome and whole-genome sequencing of the Framingham Heart Study and other cohorts for studying cardiovascular disease.

GWAS implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.

Exome sequencing identifies mutations in SLCE6A3 for renal salt wasting in Bartter’s syndrome.

Large GWAS yields >30 gene loci associated with clinical coronary artery disease.

Large GWAS reports >30 gene loci associated with blood pressure and hypertension.

GWAS also yield variants in MYH6 associated with sick sinus syndrome and variants associated with dilated cardiomyopathy and carotid intima–media thickness.

NHLBI begins whole-exome and whole-genome sequencing of the Framingham Heart Study and other cohorts for studying cardiovascular disease.

GWAS implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.

Exome sequencing identifies mutations in SLCE6A3 for renal salt wasting in Bartter’s syndrome.

Large GWAS yields >30 gene loci associated with clinical coronary artery disease.

Large GWAS reports >30 gene loci associated with blood pressure and hypertension.

GWAS also yield variants in MYH6 associated with sick sinus syndrome and variants associated with dilated cardiomyopathy and carotid intima–media thickness.

NHLBI begins whole-exome and whole-genome sequencing of the Framingham Heart Study and other cohorts for studying cardiovascular disease.

GWAS implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.

Exome sequencing identifies mutations in SLCE6A3 for renal salt wasting in Bartter’s syndrome.

Large GWAS yields >30 gene loci associated with clinical coronary artery disease.

Large GWAS reports >30 gene loci associated with blood pressure and hypertension.

GWAS also yield variants in MYH6 associated with sick sinus syndrome and variants associated with dilated cardiomyopathy and carotid intima–media thickness.

NHLBI begins whole-exome and whole-genome sequencing of the Framingham Heart Study and other cohorts for studying cardiovascular disease.

GWAS implicate genetic variants in early-onset myocardial infarction, QT-interval prolongation, blood pressure and hypertension, echocardiographic measures of cardiac structure, and blood-cell traits.
ponents of cardiovascular risk factors and subclinical disease, including coronary artery disease. Approaches to identifying the genetic causes of polygenic cardiovascular diseases (and other polygenic diseases) before completion of the draft sequence of the human genome were largely unsuccessful. A decade later, hundreds of loci associated with many cardiovascular diseases and traits have been identified. This genetic bounty is the yield of genomewide association studies, which involve testing of a large set of genetic variants in case and control subjects from a population to determine which variants are associated with the disease in question (see the Glossary). The common disease–common variant hypothesis proposes that common variants, which are defined as variants with a prevalence of at least 5% in the population, have a role in the cause and

<table>
<thead>
<tr>
<th>Glossary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Allele:</strong> One of two or more copies of a genetic sequence at a chromosomal location. Alleles can be considered according to their frequencies in the human population, ranging from common variants (minor allele frequency, &gt;5%) to low-frequency variants (minor allele frequency, 0.5 to 5%) to rare or private alleles in one or a limited number of families (minor allele frequency, &lt;0.5%). A null allele is not functional.</td>
</tr>
<tr>
<td><strong>Candidate-gene study:</strong> An approach used in genetics research that focuses on suspected, or candidate, genes that have been selected because of a perceived match between their known or presumed function and biologic function in the disease under investigation.</td>
</tr>
<tr>
<td><strong>Exome:</strong> All known protein-coding sequences, or exons, in the human genome, constituting approximately 1 to 2% of the 3.2 billion nucleotide base pairs in the human genome.</td>
</tr>
<tr>
<td><strong>Gene enhancer:</strong> A short region of DNA that can be bound with proteins, such as trans-acting factors (i.e., factors acting from a different molecule), to enhance transcription levels of genes in a gene cluster. Although enhancers are usually cis-acting (i.e., acting from the same molecule), an enhancer does not need to be close to the genes it acts on and may not be located on the same chromosome.</td>
</tr>
<tr>
<td><strong>Genomewide association study:</strong> An approach used in genetics research to look for associations between many (typically hundreds of thousands) specific genetic variations (most commonly single-nucleotide polymorphisms) and particular diseases.</td>
</tr>
<tr>
<td><strong>Genotyping array:</strong> A technique used to study many genes at once. Thousands of gene sequences are placed in known locations on a glass slide. A sample containing DNA or RNA is deposited on the slide, or gene chip. The binding of complementary base pairs from the sample and the gene sequences on the chip can be measured with the use of fluorescence to detect the presence and determine the amount of specific sequences in the sample.</td>
</tr>
<tr>
<td><strong>Induced pluripotent stem cell:</strong> A type of pluripotent stem cell derived from a nonpluripotent cell, typically an adult somatic cell such as a fibroblast, by transfection of a stem-cell–associated gene.</td>
</tr>
<tr>
<td><strong>Haplotype maps:</strong> Maps of a set of DNA variations, or polymorphisms, that tend to be inherited together. Haplotype maps are used to aid gene-discovery efforts.</td>
</tr>
<tr>
<td><strong>Haplotype structure:</strong> Patterns of haplotypes that define the structure of regions of the genome. Haplotype structure is used to compare genome regions across populations or species.</td>
</tr>
<tr>
<td><strong>Locus:</strong> The specific chromosomal location of a gene or other DNA sequence of interest.</td>
</tr>
<tr>
<td><strong>Mendelian inheritance:</strong> Patterns of inheritance characteristic of organisms that reproduce sexually, as described by Austrian monk Gregor Mendel in the mid-19th century.</td>
</tr>
<tr>
<td><strong>Mendelian randomization:</strong> A method that uses measured variation in genes of known function to examine the causal effect of a modifiable exposure on a disease. It provides unbiased estimates of the effects of a putative causal variable without the need for a traditional randomized trial.</td>
</tr>
<tr>
<td><strong>Next-generation sequencing:</strong> DNA sequencing that harnesses advances in miniaturization technology to simultaneously sequence multiple areas of the genome rapidly and at low cost.</td>
</tr>
<tr>
<td><strong>Noncoding DNA sequence:</strong> A DNA sequence that does not encode proteins. Noncoding DNA sequences, once referred to as “junk DNA,” account for the majority of genome sequences and are now known to harbor regions that regulate gene expression.</td>
</tr>
<tr>
<td><strong>Polygenic:</strong> Produced by two or more genes.</td>
</tr>
<tr>
<td><strong>Single-nucleotide polymorphism:</strong> A single-nucleotide variation in a genetic sequence; a common form of variation in the human genome.</td>
</tr>
<tr>
<td><strong>Splice variants:</strong> Abnormal variations in RNA splicing that are implicated in a disease. Many genetic disorders result from splicing variants.</td>
</tr>
<tr>
<td><strong>Whole-exome sequencing:</strong> Sequencing of the coding regions, or exons, of an entire genome.</td>
</tr>
</tbody>
</table>
pathophysiology of common diseases. It is on this premise that the genomewide association study is based, since it comprises tests of association between disease and common variants spread throughout the genome. On the basis of this approach, researchers have assembled catalogues of cardiovascular variants, using genotyping arrays, haplotype maps, and statistical methods. Changes in data-sharing policies have led to the creation of publicly available genome databases. International collaborations combining study cohorts, often including tens of thousands of research participants, have been formed.

The astonishingly large number of new loci associated with myocardial infarction and heart failure (Fig. 2) provide insights into the biologic pathways that underlie disease. The application of such findings to the prediction of risk and to the prevention and treatment of disease is premature and awaits considerable research.

**Coronary Artery Disease and Myocardial Infarction**

Genomewide association studies have identified about 30 loci associated with myocardial infarction and coronary artery disease (Table 1). A meta-analysis of 14 such studies that involved 22,233 case subjects with coronary artery disease and 64,762 control subjects of European descent and that were followed by replication studies involving 56,682 case and control subjects identified 13 new loci associated with coronary artery disease, in addition to confirming 10 of 12 previously reported loci. ABO and ADAMTS7 were found to be associated with angiographically confirmed coronary atherosclerosis, CLOCK with high blood pressure, and the APOA5 gene cluster with elevated levels of triglycerides and cholesterol subfractions. The majority of loci associated with myocardial infarction reside in genomic regions that have not previously been implicated in coronary artery disease; only a minority of loci mediate an
effect through known risk factors. A second meta-analysis of genomewide association studies involving more than 30,000 case and control subjects showed an additional four new loci associated with coronary artery disease in multiple ethnic groups.15

Studies of early-onset myocardial infarction have identified more than 10 risk loci.13 The most strongly associated locus is 9p21.36,37 This region harbors genes (CDKN2A and CDKN2B) that are implicated in cell cycling and cancer, although the single-nucleotide polymorphisms (SNPs) associated with myocardial infarction are located not near these or other protein-coding genes but rather within a noncoding RNA molecule called ANRIL (for antisense noncoding RNA in the INK4 locus). The expression of ANRIL splice variants, but not CDKN2B or other nearby genes, is associated with atherosclerosis.38 SNPs at the 9p21 locus are also associated with other cardiovascular diseases, including stroke39 and aortic aneurysm.40

Loci associated with coronary artery disease harbor genes known to be important in lipid variation, including SORT1, PCSK9, HNF1A, MRAS, and LPA. The position of other SNPs implicates inflam-
matory processes in conferring a risk of coronary atherosclerosis.

**HEART FAILURE**

Genomewide association studies have identified many possible loci associated with heart failure and death from heart failure, although few of such studies have been replicated. Genomewide association studies for heart failure have been limited by modest numbers of cases of heart failure (relative to the number of such studies for coronary artery disease) and the heterogeneous nature (and thus heterogeneous sets of cases) of heart failure. A recent genomewide association study for idiopathic dilated cardiomyopathy identified variants in HSPB7, encoding a heat-shock protein previously implicated in heart failure, and BAG3; a marked myopathy develops in mice deficient in Bag3.

**ARRHYTHMIAS**

Genomewide association studies have uncovered genetic variants for arrhythmias, including atrial fibrillation, ventricular fibrillation, sudden cardiac death, and the sick sinus syndrome (Table 1). MYH6, a previously unidentified gene associated with susceptibility to the sick sinus syndrome, encodes the alpha heavy-chain subunit of cardiac myosin, suggesting that myosin proteins may regulate cardiac conduction in addition to myocyte function.

**PERIPHERAL AND CEREBRAL VASCULAR DISEASES**

Genomewide association studies have yielded evidence of varying strength for genetic loci associated with ischemic stroke, intracranial aneurysm, peripheral arterial disease, aortic aneurysm, venous thromboembolism, and erythrocyte phenotypes (Table 1). In some cases, these loci are common to coronary artery disease and myocardial infarction, suggesting a common genetic contribution to multiple vascular beds.

**MODIFIABLE RISK FACTORS AND SUBCLINICAL DISEASE**

Genome consortia with sample sizes exceeding 10,000 cases have investigated major modifiable risk factors (e.g., hypertension, dyslipidemia, type 2 diabetes, and cigarette smoking) and obesity, using quantitative measures and clinically relevant extreme end points (Table 2). These studies point to loci associated with hypertension (as defined by the use of antihypertensive therapy or elevations in blood pressure), LDL and high-density lipoprotein (HDL) cholesterol and triglycerides, the number of cigarettes smoked, type 2 diabetes (with the use of quantitative traits of fasting glucose and glycated hemoglobin levels), and obesity (with the use of body-mass index and adiposity-related traits). Studies have also identified multiple loci associated with other presumed quantitative biomarkers of cardiovascular risk and presymptomatic coronary artery disease (e.g., fibrinogen, C-reactive protein, intercellular adhesion molecule 1, plasma homocysteine, and carotid-artery intima–media thickness and plaque).

Newer approaches for establishing evidence of causality use human populations as a model system. Mendelian randomization analysis takes advantage of the lifelong association between a risk allele and a quantitative measure of a biomarker for clinical coronary artery disease to estimate whether there is evidence of a causal association between the biomarker and the disease. Mendelian randomization studies have used genetic variants that have previously been implicated by genomewide association studies to provide evidence for a causal association between LDL cholesterol (variants in LDLR) or lipoprotein(a) (variants in LPA) and coronary artery disease and evidence against a causal association between C-reactive protein (variants in CRP) and coronary artery disease.

Current genomewide association studies focus on large populations in order to strengthen the evidence for association and replication. For example, a genomewide association study of more than 100,000 research participants identified 95 loci associated with at least one of three lipids: LDL and HDL cholesterol and triglycerides. Each individual variant had only a modest effect, the combined effect of the 95 loci explained approximately 25% of the genetic variance in LDL and HDL cholesterol levels. Similarly, large meta-analyses of variants that are associated with type 2 diabetes and myocardial infarction suggest that these variants, in aggregate, account for 25% and 10%, respectively, of the inherited variation in disease outcomes.

We have learned several lessons from genomewide association studies. Although some risk loci that have been discovered through this approach encompass genes that had been previously impli-
cated, most of such loci implicate genes that had not been thought to have a role in conferring disease risk. Many loci show association with cardiovascular disease across groups of differing ancestry, and most cardiovascular traits are influenced by a large number of loci. However, limitations of genomewide association studies have prevented immediate translation of these findings into clinical practice, since each variant has a very small effect and is therefore not useful for prediction. Moreover, the implicated variants are rarely themselves the causal variants; rather, they are linked to the true causal variants, and identification of the latter usually warrants a great deal of additional work.

**TARGETED AND GENOMEWIDE DNA SEQUENCING**
The next generation of genomic approaches is upon us, thanks to the increasing efficiency and decreasing cost of sequencing technology. Deep DNA sequencing is performed with the use of miniatureized technology that simultaneously sequences multiple areas of the genome. This approach is used to sequence candidate regions and specific components of the human genome, such as exons or noncoding DNA. Sequencing studies in community-based cohorts have shown that at least 1 of every 64 persons carries a functional mutation in one of three genes (NCCT, NKCC2, or ROMK) that is associated with clinically significant alterations in blood pressure and that there is a striking excess of nonsynonymous variants in the gene encoding adipokine (ANGPTL4) in persons with low triglyceride levels.

Targeted sequencing in a population with extreme values of LDL cholesterol identified variants in PCSK9 that occur in up to 3% of the population. Such variants are associated with a low LDL cholesterol level and a decreased risk of incident coronary artery disease. A sequencing study of FAAH and MGLL at extremes of the body-mass index uncovered several rare variants in promoter and enhancer regions of these genes, suggesting that sequence variation may regulate extreme obesity.

### Table 2. Representative Large-Scale Genomewide Association Studies of Risk Factors for Cardiovascular Disease.*

<table>
<thead>
<tr>
<th>Study</th>
<th>Risk Factor or Clinical Trait</th>
<th>Sample Size</th>
<th>Major Ethnic Group</th>
<th>Selected Major Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ehret et al., 2011</td>
<td>Systolic and diastolic blood pressure, hypertension</td>
<td>69,395</td>
<td>European</td>
<td>&gt;25 Loci, including CACNB2 and SH2B3</td>
</tr>
<tr>
<td>Teslovich et al., 2010</td>
<td>Total and LDL cholesterol</td>
<td>100,184</td>
<td>European, South Asian, East Asian, African</td>
<td>&gt;35 Loci, including SORT1 and HMGCR</td>
</tr>
<tr>
<td>Teslovich et al., 2010</td>
<td>HDL cholesterol</td>
<td>100,184</td>
<td>European, South Asian, East Asian, African</td>
<td>&gt;35 Loci, including SCARB1 and CETP</td>
</tr>
<tr>
<td>Teslovich et al., 2010</td>
<td>Triglycerides</td>
<td>100,184</td>
<td>European, South Asian, East Asian, African</td>
<td>&gt;20 Loci, including ANGPTL3 and JMJD1C</td>
</tr>
<tr>
<td>Thorgeirsson et al., 2008</td>
<td>Quantity of cigarettes smoked</td>
<td>31,266</td>
<td>European</td>
<td>Top loci: CHRNA3 and 15q25</td>
</tr>
<tr>
<td>TAGC, 2010</td>
<td>Quantity of cigarettes smoked</td>
<td>74,053</td>
<td>European</td>
<td>Top loci: DBH and CYP2A6</td>
</tr>
<tr>
<td>Voight et al., 2010</td>
<td>Type 2 diabetes mellitus</td>
<td>47,117</td>
<td>European</td>
<td>&gt;25 Loci, including TCF7L and IRS1</td>
</tr>
<tr>
<td>Dupuis et al., 2010</td>
<td>Fasting glucose level</td>
<td>46,186</td>
<td>European</td>
<td>&gt;20 Loci, including GCKR and ADRA2A</td>
</tr>
<tr>
<td>Speliotes et al., 2010</td>
<td>Body-mass index</td>
<td>123,865</td>
<td>European</td>
<td>&gt;30 Loci, including FTO and TME1B</td>
</tr>
<tr>
<td>Heard-Costa et al., 2009</td>
<td>Waist circumference</td>
<td>31,373</td>
<td>European</td>
<td>Top loci: NRXN2 and MC4R</td>
</tr>
</tbody>
</table>

* Genomewide association studies were considered to be large in scale if they include a sample of more than 10,000 subjects. HDL denotes high-density lipoprotein, LDL low-density lipoprotein, and TAGC Tobacco and Genetics Consortium.
are associated with cardiovascular disease and its risk factors. Resequeuing of exons in APOA5, GCKR, LPL, and APOB — genes that are implicated in hypertriglyceridemia — revealed twice as many rare missense or nonsense variants in persons with high triglyceride levels than in control subjects, corresponding to a carrier frequency of 28.1% in affected persons and 15.3% in control subjects.64

Whole-exome sequencing has uncovered genetic variants in persons with rare forms of diseases that affect blood pressure and circulating blood lipids, including a missense mutation in SLC26A3 (a congenital chloride diarrhea locus) in a patient with a suspected diagnosis of the renal salt-wasting disease Bartter's syndrome,65 a variant in ANGPTL3 in a family with familial combined hypolipidemia,66 and variants in BAG3 in families with familial dilated cardiomyopathy.64 Whole-exome sequencing in large populations may help to identify rare genetic variants in relatively common cardiovascular diseases, such as early myocardial infarction, that confer a major risk of complications and death and for which there is good evidence of heritability; indeed, this is a major focus of whole-exome sequencing at present. Whole-genome sequencing has also confirmed single-gene disease variants in families with rare diseases,65 including mutations in ABCG1 in an infant with severe hypercholesterolemia.66

CARDIOVASCULAR DISEASE PATHWAYS

The best yield of genomewide association studies is the provision of insights into the biologic pathways — often previously unsuspected — that underlie causes of disease. Such insights have led to hypothesis-driven investigations of implicated pathways with the use of molecular, genetic, biochemical, and cellular approaches. The genetic control of myocardial infarction and lipids provides a case in point.

The initial findings that connected genetic variation in the 9p21 region with atherosclerotic progression and myocardial infarction caught cardiovascular scientists by surprise, because this region is devoid of genes that had previously been associated with coronary artery disease. Subsequent research has revealed possible mechanisms by which genes in this region may contribute to atherosclerosis. A recent report implicated the 9p21 risk interval in regulating the expression of cardiac CDKN2A/B expression genes.68 Another report implicated inflammatory pathways through 33 gene enhancers located in the 9p21 region.67 Two SNPs located in one of these enhancers disrupt a binding site for STAT1, a signal transduction protein that regulates inflammation. This enhancer locus physically interacts with the CDKN2A/B locus and an interval downstream of IFNA21, the gene that encodes interferon-γ in human vascular endothelial cells. The activation of interferon-γ affects transcriptional regulation of the 9p21 locus, including STAT1 binding, suggesting a link between genetic susceptibility to coronary artery disease and response to inflammatory signaling in vascular cells. Yet another driver of disease may be the aforementioned noncoding RNA element ANRIL.

Loci that are discovered by genomic approaches are also relevant to lipid biology and potential therapeutic targets. The gene encoding sortilin (SORT1) contains a common variant that creates a binding site for a transcription factor that when bound to this site alters liver expression of LDL cholesterol in humans. Transgenic studies have shown that mice with Sort1 that contains the transcription-factor binding site have altered plasma levels of LDL cholesterol, suggesting a previously unknown regulatory pathway for LDL cholesterol.68 A potentially therapeutic target was implicated by the discovery of a pair of common nonsense variants in PCSK9 in patients of African descent. An estimated 2.6% of persons of African descent carry one of these variants, and each variant results in comparatively low lipid levels and a reduced susceptibility to myocardial infarction.69 Humans who are homozygous for this null allele are healthy and have a reduced risk of myocardial infarction, suggesting that the absence of PCSK9 may be well tolerated and thus rendering PCSK9 an attractive drug target.

SNPs that more recently have been found to be associated with cardiovascular disease and its risk factors implicate components of pathways previously identified with risk factors for or mechanisms of the disease. For example, the 95 loci associated with levels of LDL and HDL cholesterol or triglycerides in more than 100,000 persons of European ancestry implicate nearly all the 18 genes that have previously been shown to be mutated
in rare mendelian lipid disorders. Most of these loci were also associated with cardiovascular disease in persons of African descent and in persons of Asian descent. Some loci housed both common and rare variants; in such cases, the common variants explained more of the heritability.

An exciting approach is the use of induced pluripotent stem cells to model cardiovascular disease and test hypotheses generated from genomic studies. Two lines of induced pluripotent stem cells, which were generated from dermal fibroblasts in patients who had the long-QT syndrome with specific genetic mutations, have been differentiated into cardiomyocytes and are being used to investigate electrophysiological properties, such as action potentials and ion fluxes.

**Prediction, Prevention, and Treatment**

The improved understanding of cardiovascular pathophysiology that has been achieved through genetic discovery provides new opportunities for prediction, prevention, and treatment (Fig. 3). Genetic risk prediction is at an early stage, and insufficient evidence exists at present to warrant the use of a genetic risk score on the basis of SNPs identified through genomic approaches. Additional research is needed to prospectively assess the utility of genetic risk scores in the prediction of cardiovascular disease, such as myocardial infarction and coronary artery disease, before clinical use. Some observers have suggested that 150 genes with odds ratios of 1.5 or 250 genes with odds ratios of 1.25 will be needed. Future studies of genetic risk scores will probably require hundreds of associated SNPs (identified through genomewide association studies), combined with low-frequency risk alleles discovered through whole-genome sequencing, to provide evidence-based prediction.

Studies are under way to investigate the use of next-generation sequencing for the screening of rare forms of cardiovascular disease and to annotate all mutations within an individual’s genome. Although the costs of accurate whole-genome sequencing are dropping dramatically, the costs of data analysis and storage remain high and are barriers to clinical implementation.

Loci that are uncovered by genomewide association studies, despite their modest effects, may
have therapeutic implications. For example, a common variant at the HMGCR locus (with a prevalence of 40% among persons of European ancestry) is associated with a very modest elevation in the LDL cholesterol level (2.8 mg per deciliter [0.07 mmol per liter]).43 There are no known rare mutations at this locus, presumably because they would be lethal. Yet the encoded protein, 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, is the target of statin drugs, which reduce LDL cholesterol levels and the risk of myocardial infarction.

Pharmacogenomics has potential near-term application. Discoveries that are provided by genomewide association studies have strengthened the evidence for pharmacogenetic interactions in a number of commonly used cardiovascular drugs. Variants in CYP2C9 and VKORC1 explain up to 40% of the variation in the adjusted dose of warfarin, and the Food and Drug Administration (FDA) recently revised the label on warfarin to allow for genotype-specific dose ranges.77 Variations in a cytochrome P-450 enzyme, CYP2C19, are associated with decreased antiplatelet efficacy and an increased risk of cardiovascular disease among patients taking the antiplatelet drug clopidogrel, which prompted an FDA warning and recommendations for a dose adjustment or use of alternative drugs in patients with this variant.77 Evidence of other pharmacogenetic interactions that may be clinically important is accumulating. Variation in the β2-adrenergic–receptor gene, ADRB1, is associated with altered responsiveness to beta-blockade in heart failure.78 A variant in SLC01B1 has been implicated in statin-related myopathy.79 Deep sequencing of variants and genes related to drug absorption, distribution, metabolism, and excretion may identify specific variants that contribute to the heterogeneity of responsiveness to cardiovascular drugs.

**FUTURE DIRECTIONS**

The field of cardiovascular genomics has two distinct goals: understanding biologic mechanisms and applying that knowledge to personalized medicine. Knowledge of molecular pathways can lead to improved therapeutics on a broad basis (regardless of the individual genotype) or at an individualized level (targeted specifically to the genotype). During the past 5 years, the discovery of hundreds of cardiovascular loci is a start. In the years to come, we will require studies of tens of thousands of patients with cardiovascular disease that combine tests of genomewide association (involving missense, rare, and common variants) with sequencing.80 Functional studies are essential to characterize molecular and cellular pathways and to develop appropriately targeted therapies. Genomics is permeating biomedical research, and these genomic advances must proceed through basic discoveries, functional characterization, preclinical proof-of-principle studies, first-in-human studies, clinical trials, and regulatory approval. Cardiovascular science and medicine have made enormous strides over the past century, beginning with the brilliant elucidation of cardiovascular physiology and leading to molecular and cellular studies, with concurrent epidemiologic determinations of risk factors. The cardiovascular field is now primed for genomic medicine to make equal, if not greater, contributions.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

REFERENCES


Copyright © 2011 Massachusetts Medical Society.