Peripheral arterial disease (PAD) is characterized by reduced blood flow to the limbs, usually as a consequence of atherosclerosis, and affects ≈12 million Americans. It is a common cause of cardiovascular morbidity and an independent predictor of cardiovascular mortality. Similar to other atherosclerotic diseases, such as coronary artery disease, PAD is the result of the complex interplay between injurious environmental stimuli and genetic predisposing factors of the host. Genetic susceptibility to PAD is likely contributed by sequence variants in multiple genes, each with modest effects. Although many of these variants probably alter susceptibility both to PAD and to coronary artery disease, it is likely that there exists a set of variants specifically to alter susceptibility to PAD. Despite the prevalence of PAD and its high societal burden, relatively little is known about such genetic variants. This review summarizes our limited present knowledge and gives an overview of recent, more powerful approaches to elucidating the genetic basis of PAD. We discuss the advantages and limitations of genetic studies and highlight the need for collaborative networks of PAD investigators for shedding light on this dark corner of vascular biology. (Arterioscler Thromb Vasc Biol. 2007; 27:000-000.)

**Key Words:** peripheral vascular disease ■ genetics ■ epidemiology ■ atherosclerosis

**Epidemiology, Diagnosis, and Established Risk Factors of PAD**

Peripheral arterial disease (PAD) is a disease characterized by reduced blood flow to the lower extremities, most often because of atherosclerosis. The prevalence of PAD varies substantially with age and certain risk factors. In population-based studies of adults, estimates of the prevalence of PAD range from 3% to 19%.1–9 In high-risk groups, such as subjects over the age of 50 years with a history of diabetes or smoking or subjects over the age of 70 years, the prevalence of PAD is >25%.10 In North America and Europe, PAD affects an estimated 27 million adults.8,11

PAD is a leading cause of morbidity through functional decline, intermittent claudication, and critical leg ischemia.10,12,13 PAD is also a significant predictor of cardiovascular mortality. Patients with PAD are twice as likely to have prevalent CAD as those without disease.1,2 Both asymptomatic and symptomatic PAD are associated with increased risk of stroke, myocardial infarction, and death.5,6,14–16 In particular, the risk of cardiovascular death over a 10-year period is increased 6-fold in patients with PAD.14

Office-based diagnosis of PAD is made by medical history, physical examination, and the measurement of ankle blood pressures using a hand-held Doppler ultrasound device. The ankle brachial index (ABI) is defined as the ratio of ankle pressure to the brachial systolic pressure. The reference range of ABI is 0.91 to 1.3 (systolic pressure is normally slightly higher at the ankle than the arm because of pulse-wave reflections along the length of the limb arteries).17 An ABI <0.9 is consistent with a diagnosis of PAD, and the lower the ABI, the more hemodynamically severe the disease (ie, 0.90 is mild, 0.70 is moderate, and ≤0.5 is severe PAD). Individuals with an ABI >1.30 likely have noncompliant vessels secondary to pathological processes involving vascular fibrosis and calcification; these individuals tend to evidence greater coronary artery calcium.18 Some have suggested a tighter range for normal ABI, because individuals with an ABI between 0.90 and 1.00 manifest more ischemic leg pain,19 subclinical atherosclerosis,18,20,21 and are at greater risk for major adverse cardiovascular events.22,23 Although the ABI is an excellent screening test for the presence of PAD, it correlates only modestly with functional limitation.24,25 For example, a severe reduction in ABI (eg, an ABI of <0.5) is typically observed in a patient with resting pain or tissue loss. However, a similar ABI might be discovered in an individual with modest symptomatology (eg, calf discomfort after walking 2 blocks). These differences may be because of heterogeneity in collateral formation or skeletal muscle adaptation. Because most patients are asymptomatic or have atypical symptoms12 and because ABI measurements are often not performed or reimbursed, PAD frequently goes undiagnosed and untreated.3,8–10,16 The 2005 American Col-
Atherosclerosis is a result of various injurious exposures that lead to endothelial dysfunction followed by chronic inflammatory infiltration in the arterial wall. Not surprisingly, there is substantial overlap between the pathogenesis of PAD and other forms of systemic atherosclerotic diseases, like coronary artery disease (CAD) and carotid disease. PAD and CAD share several risk factors, including diabetes, smoking, hypertension, insulin resistance, low high-density lipoprotein levels, elevated low-density lipoprotein levels, and advanced age. Approximately 70% of the cases of PAD (as defined by ABI) can be attributed to these risk factors, somewhat higher than the corresponding estimates for CAD. Convincing evidence exists that some of these risk factors have proportionately greater effects on the development of PAD than CAD. For instance, diabetes mellitus (DM) and smoking are particularly strong risk factors for PAD and CAD share several risk factors, including diabetes, smoking, hypertension, insulin resistance, low high-density lipoprotein levels, elevated low-density lipoprotein levels, and advanced age. Approximately 70% of the cases of PAD (as defined by ABI) can be attributed to these risk factors, somewhat higher than the corresponding estimates for CAD. Convincing evidence exists that some of these risk factors have proportionately greater effects on the development of PAD than CAD. For instance, diabetes mellitus (DM) and smoking are particularly strong risk factors for PAD.

Despite the prevalence and the high societal cost attributed to PAD, genetic susceptibility to PAD remains a "dark corner" in vascular biology. Our understanding of the genetic basis of PAD is limited and, so far, based on studies that examine a small number of genes in a small number of samples. However, the completion of the human genome draft sequence, the cataloging of common human DNA variation by the International HapMap Project, and the technology for large-scale data collection have provided new opportunities to carry out more powerful studies that will hopefully more definitively identify genetic determinants of PAD. Currently, there are no genetic tests available that reliably identify the subset of subjects carrying inherited risk factors for developing PAD. Such tests could potentially be developed if genetic determinants of PAD are uncovered.

In this review, we first summarize and critically appraise the results of studies that support the role of genetics in the pathogenesis of atherosclerotic PAD. We do not discuss syndromic forms of PAD, such as those seen in Buerger’s disease, Ehlers-Danlos syndrome, or systemic vasculitides, which are pathologically distinct from “garden-variety” atherosclerotic PAD and are likely to have different genetic contributions. We then summarize the results of existing studies of novel genetic determinants of PAD, highlighting the basic principles, strengths, and weaknesses of various study designs. Thereafter, we conclude by describing novel, more comprehensive, and hopefully more successful approaches to the discovery of genetic determinants of PAD.

The Role of Genetics in PAD
An important first step in the field of complex trait genetics is to estimate the contribution of inherited factors to the development of the disease. This estimate can be obtained through various types of familial aggregation studies.

Three family studies to date have reported estimates of the heritability of PAD. In all 3 of the studies, the ABI was used as a surrogate for the presence and severity of PAD. In the first report from the National Heart, Lung, and Blood Institute twin study, ~48% of the overall variability in the ABI could be attributed to additive genetic effects after adjusting for risk factors of atherosclerosis. However, inconsistent with this observation was the lack of a statistically significant higher concordance rate for an ABI <0.9 in monozygotic twins (33%) compared with dizygotic twins (31%). Of note, the authors of this study acknowledged that the estimation of genetic effects may have been biased by selective mortality and loss to follow-up of participants at high risk for PAD. Furthermore, twin studies are also known to overestimate genetic effects because of the presence of a higher degree of shared environmental factors compare with nontwin studies. A second study of hypertensive nontwin sibs participating in the Genetic Epidemiology Network of Arteriopathy Study reported an overall lower estimate of the heritability of ABI. In non-Hispanic whites, this estimate was 35.7% (P<0.001) before adjustment for atherosclerosis risk factors and 21.2% (P=0.006) after adjustment of risk factors. Estimates of heritability before and after adjustment for covariates were similar in African Americans (35.7%, P<0.001 and 19.5%, P=0.002) participating in the Genetic Epidemiology Network of Arteriopathy Study. In the third and final family study, investigators from the Framingham Offspring Study estimated a heritability of ABI similar to the one reported in the Genetic Epidemiology Network of Arteriopathy cohort. Using variance component analysis, this estimate was 27% before adjustment for covariates and 21% after adjustment for covariates (both P values <0.0001).

A significant limitation of the heritability studies mentioned above is that a large majority of subjects had ABIs in the borderline or reference range (>0.9). In this range, there is either no correlation of the ABI with the degree of lower extremity atherosclerosis or it is very poor. Thus, heritability studies of ABI to date may reflect more the degree of genetic influence of ABI in the reference range, which may differ significantly from the degree of the genetic influence on PAD. Consistent with this possibility is the fact that established strong predictors of clinically significant PAD explain a relatively small proportion of the total variance of ABI in the 2 family studies that reported these estimates.

Evidence quite suggestive for a genetic basis of PAD comes from a study of Valentine et al where they showed that both premature, asymptomatic CAD and occult lower-extremity vascular disease are more common in families of probands with premature PAD, indicating a strong familial aggregation of vascular disease. However, this study was not able to quantify the degree of familial aggregation that was independent of that related to smoking and other established risk factors of atherosclerosis.

A report from the National Heart, Lung, and Blood Institute Multi-Ethnic Study of Atherosclerosis on the determinants of PAD provides some indirect evidence for the role of genes in the development of PAD. In this report,
self-reported race/ethnic group predicted the presence of PAD (defined as an ABI <0.9) independent of all other “established” and “novel” cardiovascular disease risk factors of atherosclerosis with the lowest risk in Hispanics and Chinese and the highest risk in African Americans. These results suggest the presence of determinants of PAD that correlate with race, some of which could be genetic in nature. Several other studies have reported greater prevalence of PAD in African American individuals.50–52

Collectively, family studies to date do suggest that PAD is heritable. However, what remains to be quantified accurately is the degree of genetic influence on the development of PAD independent of established risk factors of atherosclerosis that also clusters in families, such as smoking, diabetes, hyperlipidemia, and hypertension. What is clearer is that the genetic susceptibility to diseases that result from atherosclerosis is significant and cannot be attributed to a single gene.28,53–55 Rather, susceptibility is likely a result of a combination of alleles in multiple genes, each with a modest effect on risk.

Existing Work From Linkage and Case-Control Association Studies of PAD

Traditional linkage studies examine the relationship between the sharing of DNA markers (microsatellite markers or single nucleotide polymorphisms [SNPs]) identity-by-descent and the sharing of phenotype between relatives within families. The conventional likelihood ratio–based test statistic is referred to as the log-odds score. Genome-wide linkage analysis using microsatellite markers often scans markers at a density of 3 to 10 centimorgans. A log-odds score >3 is considered statistically significant.56 Because of the limited number of recombination events within a pedigree, the resolution of linkage study may be quite low.57,58 Once linkage with a marker is detected, other more closely spaced markers in the same region are genotyped in a process referred to as fine mapping. The marker with the strongest signal is usually near 1 or more genes that may be causal for the phenotype under study. Traditionally, linkage studies have been very successful in identifying susceptibility genes with very large effects most commonly seen in diseases that follow Mendelian patterns of inheritance.59,60 However, they have proven to be more difficult to apply to the study of common and complex traits such as PAD.61–64

Two linkage studies relating to PAD have been reported to date. In the first study, investigators from deCODE genotyped ≈1000 microsatellite markers in 272 Icelandic patients and 612 relatives (from 116 families)65 and identified a linkage peak on chromosome 1p. During fine-scale mapping of this region, a peak log-odds score of 3.9 was obtained at the marker D1S2895. However, the same linkage signal has not yet been reported in an independent sample. If confirmed, further fine mapping of this region will be needed to identify the causal genetic variant. The second study used ABI as the phenotype and only 350 microsatellites.46 In this study of 1310 African Americans and 796 non-Hispanic whites, no convincing evidence of linkage was reported (highest multipoint log-odds score of 2). However, as mentioned above, a limitation of this study was the high proportion of ABIs in the reference range.

In contrast to linkage analysis, association studies explore differences in allele frequencies between unrelated individuals with and without a particular phenotype/disease. Carefully designed association studies may achieve greater power and finer mapping resolution than family based linkage studies in detecting susceptibility loci with modest effects.30,64,66,67 For many common diseases, association studies are also much more practical to conduct than family based studies. The 2 basic association study designs are the cohort design and the case-control design, with the latter design having been more widely adopted because of its efficiency. To date, there have only been a small number of association studies of PAD examining a total of <20 genes. These studies are summarized in Table 1. All of these studies have used the case-control design with sample sizes ranging from ≈100 to 1300 subjects (median number of case subjects: <200; median number of control subjects: <450). In most of these studies, only 1 polymorphism per gene was genotyped. Collectively, these studies have not convincingly uncovered any novel genetic determinants of PAD.30,36–38,57,66–71

Limitations of Studies to Date in Search of Novel Genetic Determinants of PAD

Many genetic studies to date, including those focusing on the genetic determinants of PAD, suffer from a number of weaknesses. These weaknesses include, but are not limited to, a lack of statistical power and confounding because of population stratification.64,72 Below, we briefly discuss the various reasons behind these design weaknesses, as well as reiterate solutions proposed by experts in the field of genetic epidemiology of complex traits both in general terms and as they relate to the studies of PAD published to date.73 We also make some of our own recommendations for studies of PAD.

Statistical Power

The most important root cause, by far, for the lack of reproducibility of most genetic association studies to date of complex traits, including PAD, is a lack of statistical power to detect the positive association described in either the original report or in subsequent replication studies. A lack of power in the setting of a very low pretest probability of an association between a given genetic variant and a phenotype, as well as a P value that is only marginally significant by “traditional” standards (ie, ≈0.05 or ≈0.01), virtually guarantees that the original report is a false positive.74 False-negative findings may also occur because of low power.

Critical factors influencing power include sample size, the prevalence of the exposure in the general population, and the odds ratio that one wishes to detect. It is now clear that, for complex diseases like PAD, sample size needs to be large enough to provide adequate power to detect a genotypic odds ratio as low as 1.2. Typically, this would require genotyping thousands of subjects.66

Linkage Disequilibrium

Linkage disequilibrium (LD) refers to the correlation between alleles at 1 polymorphic site with those at another polymorphic site nearby, and results from a polymorphism arising in a genomic region characterized by a unique combination of
alleles in other polymorphic sites. This unique combination of closely linked polymorphisms is frequently referred to as a haplotype.\textsuperscript{75} LD between SNPs in a region is gradually reduced over time by chromosomal recombinations generally proportional to the distance in base pairs separating 2 polymorphic sites and is also related to the demographic history of the population.\textsuperscript{57,76,77} Through large-scale characterization of genetic variants in multiple race/ethnic groups, patterns of LD have been significantly refined.\textsuperscript{40,78,79} These efforts have confirmed the presence of considerably different patterns of LD between major racial/ethnic groups, with the lowest LD between SNPs present in populations of African descent.\textsuperscript{77}

As a consequence of LD, the allele at 1 locus of a given region can often predict the allele of neighboring loci.\textsuperscript{80} Many of the positive genetic association studies of PAD to date have genotyped only 1 SNP and have not explored the strength of the association in other SNPs in high LD with the positive SNP. Doing so could reveal even stronger and more convincing associations. On the other hand, it is possible that some of the negative reports to date are falsely negative.

### TABLE 1. Examples of Existing Case-Control Studies to Identify PAD Susceptibility Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>HUGO Name</th>
<th>Polymorphism (as Described)</th>
<th>Case Subjects, No.</th>
<th>Control Subjects, No.</th>
<th>Ethnicity (as Described)</th>
<th>Odds Ratio</th>
<th>P</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiotensin-converting enzyme</td>
<td>ACE</td>
<td>Insertion/deletion</td>
<td>98</td>
<td>240</td>
<td>German</td>
<td></td>
<td></td>
<td>140</td>
</tr>
<tr>
<td>Angiotensin-converting enzyme</td>
<td>ACE</td>
<td>Insertion/deletion</td>
<td>45</td>
<td>316</td>
<td>Chinese</td>
<td></td>
<td></td>
<td>141</td>
</tr>
<tr>
<td>Angiotensin-converting enzyme</td>
<td>ACE</td>
<td>Insertion/deletion</td>
<td>522</td>
<td>522</td>
<td>White, Austrian</td>
<td></td>
<td></td>
<td>142</td>
</tr>
<tr>
<td>Chemokine, CC motif, receptor 5</td>
<td>CCR5</td>
<td>Δ32</td>
<td>76</td>
<td>172</td>
<td>Italian</td>
<td></td>
<td></td>
<td>143</td>
</tr>
<tr>
<td>Chemokine, CX3C motif, receptor 1</td>
<td>CX3CR1</td>
<td>V249I, T280 mol/L</td>
<td>522</td>
<td>522</td>
<td>White, Austrian</td>
<td></td>
<td></td>
<td>144</td>
</tr>
<tr>
<td>Cytochrome B α subunit (p22 phox)</td>
<td>CYBA</td>
<td>C242T</td>
<td>324</td>
<td>295</td>
<td>Austrian</td>
<td></td>
<td></td>
<td>145</td>
</tr>
<tr>
<td>Factor II</td>
<td>F2</td>
<td>G20210A</td>
<td>336</td>
<td>300</td>
<td>Austrian</td>
<td></td>
<td></td>
<td>146</td>
</tr>
<tr>
<td>Factor II</td>
<td>F2</td>
<td>G20210A</td>
<td>184</td>
<td>330</td>
<td>White, French</td>
<td>4.3</td>
<td>0.02</td>
<td>147</td>
</tr>
<tr>
<td>Factor II</td>
<td>F2</td>
<td>G20210A</td>
<td>433</td>
<td>433</td>
<td>Austrian</td>
<td></td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>Factor V</td>
<td>F5</td>
<td>G1691A (Leiden)</td>
<td>336</td>
<td>300</td>
<td>Austrian</td>
<td></td>
<td></td>
<td>146</td>
</tr>
<tr>
<td>Factor V</td>
<td>F5</td>
<td>G1691A (Leiden)</td>
<td>184</td>
<td>330</td>
<td>White, French</td>
<td></td>
<td></td>
<td>147</td>
</tr>
<tr>
<td>Factor V</td>
<td>F5</td>
<td>G1691A (Leiden)</td>
<td>433</td>
<td>433</td>
<td>Austrian</td>
<td></td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>Factor VII</td>
<td>F7</td>
<td>R333Q</td>
<td>88</td>
<td>423</td>
<td>Scottish</td>
<td></td>
<td></td>
<td>149</td>
</tr>
<tr>
<td>Factor XIII, A subunit</td>
<td>F13A1</td>
<td>V34L</td>
<td>873</td>
<td>523</td>
<td>Austrian</td>
<td></td>
<td></td>
<td>150</td>
</tr>
<tr>
<td>Fibrinogen</td>
<td>FGB</td>
<td>G-455A</td>
<td>88</td>
<td>423</td>
<td>Scottish</td>
<td></td>
<td></td>
<td>149</td>
</tr>
<tr>
<td>Fibrinogen</td>
<td>FGB</td>
<td>T1689G</td>
<td>104</td>
<td>663</td>
<td>Scottish</td>
<td></td>
<td></td>
<td>151</td>
</tr>
<tr>
<td>Fibrinogen</td>
<td>FGB</td>
<td>Bcl I digestion</td>
<td>121</td>
<td>126</td>
<td>Scottish</td>
<td>7.6</td>
<td>0.03</td>
<td>152</td>
</tr>
<tr>
<td>Haptoglobin</td>
<td>Hp</td>
<td>Hp1, Hp2</td>
<td>141</td>
<td>1000</td>
<td>White, Belgian</td>
<td>1.82</td>
<td>&lt;0.001</td>
<td>153</td>
</tr>
<tr>
<td>Hemochromatosis</td>
<td>HFE</td>
<td>C282Y, H63D</td>
<td>522</td>
<td>522</td>
<td>White, Austrian</td>
<td></td>
<td></td>
<td>154</td>
</tr>
<tr>
<td>Hepatic lipase</td>
<td>LPC</td>
<td>G-250A</td>
<td>241</td>
<td>241</td>
<td>White, Austrian</td>
<td>1.69</td>
<td>0.02</td>
<td>155</td>
</tr>
<tr>
<td>Intracellular adhesion molecule 1</td>
<td>ICAM1</td>
<td>K469E</td>
<td>75</td>
<td>227</td>
<td>White, Italian</td>
<td>3.5</td>
<td>0.004</td>
<td>156</td>
</tr>
<tr>
<td>Integrin β3 (platelet glycoprotein Illa)</td>
<td>ITGB3</td>
<td>PLA1/A2</td>
<td>815</td>
<td>518</td>
<td>Austrian</td>
<td></td>
<td></td>
<td>157</td>
</tr>
<tr>
<td>Integrin β3 (platelet glycoprotein Illa)</td>
<td>ITGB3</td>
<td>P1α</td>
<td>104</td>
<td>663</td>
<td>Scottish</td>
<td>0.49</td>
<td>&lt;0.05</td>
<td>151</td>
</tr>
<tr>
<td>Interleukin 6</td>
<td>IL6</td>
<td>G-174C</td>
<td>84</td>
<td>183</td>
<td>White, Italian</td>
<td>4.6</td>
<td>0.001</td>
<td>158</td>
</tr>
<tr>
<td>Methylene-tetrahydrofolate reductase</td>
<td>MTHFR</td>
<td>C677T</td>
<td>51</td>
<td>51</td>
<td>Swiss</td>
<td></td>
<td></td>
<td>159</td>
</tr>
<tr>
<td>Methylene-tetrahydrofolate reductase</td>
<td>MTHFR</td>
<td>C677T</td>
<td>433</td>
<td>433</td>
<td>Austrian</td>
<td></td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>Plasminogen activator inhibitor 1</td>
<td>PAI1</td>
<td>HindIII</td>
<td>88</td>
<td>423</td>
<td>Scottish</td>
<td></td>
<td></td>
<td>149</td>
</tr>
<tr>
<td>Purinergic receptor P2Y</td>
<td>P2RY12</td>
<td>H2 haplotype</td>
<td>184</td>
<td>330</td>
<td>White, French</td>
<td>2.3</td>
<td>0.002</td>
<td>160</td>
</tr>
</tbody>
</table>

NS indicates not specified.
because the chosen SNPs are in low LD with nongenotyped causal SNPs in the candidate genes. Based on LD patterns from HapMap data, as well as other publicly available resources (such as the SeattleSNPs Program in Genomic Applications), it is now possible to select a minimal set of SNPs (referred to as "tag" SNPs) that best represents all of the common variations in a region encoding a gene.

Misclassification
Misclassification of the exposure of interest in genetic association studies is a direct consequence of genotyping errors and can be nondifferential or differential. The former is a result of errors that are randomly distributed between case and control subjects, whereas the latter is a result of errors that occur preferentially in either between case and control subjects (eg, when the genotyping conditions are different in case subjects compared with control subjects). Nondifferential misclassification increases the rate of false-negatives by introducing noise and eroding study power but can be overcome by larger sample sizes. However, even an error rate of 1% can have a profound effect on power, especially if the allele in question is rare. Differential misclassification, on the other hand, leads to false-positive associations that are only exaggerated with larger sample sizes. To prevent differential misclassification because of systematic genotyping error, samples should be genotyped in random order. In large-scale studies, it is recommended that researchers routinely perform quality control assays by duplicating assays for some samples (in a blinded fashion) to assess reproducibility, using "no-template controls" to detect DNA contamination, and including "positive control" DNAs with known genotypes. With good laboratory practice and many of the newer technologies, >99% reproducibility and <1% no-call rates can be achieved.

Incomplete characterization of the phenotype may also dilute power, which is usually nondifferential (ie, random with respect to genotype). Nondifferential misclassification is a common consequence of not using a "gold-standard" tool to characterize the phenotype of interest. For example, although the ABI has a high sensitivity and specificity for detecting occlusive lower extremity disease (as defined by at least one 50% stenosis by angiography), its performance is significantly worse in detecting subocclusive disease. Therefore, a considerable number of subjects with subocclusive but significant disease may be misclassified as control subjects. Perhaps the best noninvasive tool to characterize PAD in the context of genetic association studies will turn out to be an imaging procedure, such as an MRI or multislice CT that can accurately quantify the degree of atherosclerosis in the lower extremities even in the absence of calcification. Such imaging protocols are still under development but should be validated in the near future. In the meantime, we recommend using the traditional cutoff value ABI of 0.9 to identify case subjects to be included in genetic association studies of PAD despite the possibility of misclassification. For reasons outlined earlier in this review, we advise against the use of ABI as a continuous measure as a surrogate PAD phenotype for genetic association studies, particularly in populations with a low prevalence of disease.

Population Stratification/Substructure
Study design weaknesses that relate to statistical power can, in general, all be overcome by increasing sample size. In contrast, bias from confounding cannot be overcome by larger sample sizes and must be addressed by either matching case subjects to control subjects on the potential confounder in the recruitment phase of a study or by statistically adjusting the odds ratio for the potential confounder in the analysis phase.

A variable can only be a confounder if it is correlated both with outcome and the exposure of interest. In the field of complex trait genetics, the potential confounder that has deservedly received the most attention is race/ethnic group. Statistical geneticists often refer to this type of confounding as "population stratification" or "population substructure." The potential for population stratification in properly designed association studies has been a subject of debate.

Population stratification is possible because many phenotypes, including PAD, have significantly different rates across the major ancestral populations, and the reasons for these differences are not strictly genetic in nature. Similarly, allele frequencies for many SNPs and patterns of LD differ substantially across major ancestral populations. Therefore, if the race/ethnic mix is not similar in case and control subjects, false associations may arise as a result of confounding. Under certain circumstances, population stratification can also mask true associations. As a simple surrogate, self-reported race/ethnicity seems to correlate very well with the observed clusters of SNP frequencies specific to major populations. Therefore, in studies of subjects who identify solely with 1 of these major groups matching case and control subjects on self-reported race/ethnicity or performing a stratified analysis probably eliminates a vast majority of the potential confounding. However, the potential for significant residual confounding because of population stratification increases when studying recently admixed complex populations with crude race/ethnicity labels, such as African American or Latino populations. In African American populations, the range of European ancestry varies widely. Similarly, the degree of black ancestry in Latino populations can vary substantially. Because many studies (especially in the United States) involve a significant number of subjects from genetically admixed populations, the degree of admixture within case subjects and within control subjects can vary substantially by chance. Even a modest difference in population substructure between case and control subjects may have profound effects on association outcomes. Fortunately, analytical methods have been developed to effectively deal with residual population stratification using ancestry informative markers. Ancestry informative markers are SNPs that possess large differences (Δ) in allele frequencies among major ancestral populations. Ancestry informative markers can be used to control for population substructure using the "genomic control," "structured association" methods, or methods based on principle components analysis. In general, between 50 and several hundred well-spaced ancestry informative markers are needed to estimate the individual proportions of the major ancestral populations that exist in an admixed individual.
Future Directions

An Approach to PAD Genetic Studies: Candidate Gene Studies

Identification of PAD susceptibility genes could have a significant impact on diagnosis and treatment of this under-recognized and often-untreated disease. Candidate gene studies can play a valuable role, as long as they are designed with certain principles in mind.\(^{107,108}\) Studies should include large sample sets that are accurately phenotyped with respect to PAD status, medications, physical activity,\(^{109–111}\) and medical history. A detailed survey of the presence and duration of all well-established risk factors of atherosclerosis, such as smoking, hyperlipidemia, blood pressure, and diabetes, is crucial to adjust for exposure. Blood should be collected and used for DNA purification and other plasma measurements.

Candidate genes for PAD should be genotyped using platforms with very high accuracy and reproducibility. “Tag SNP” approaches are now standard, including SNPs to adequately cover variation in subjects from all race/ethnic groups in the study.\(^{90}\) It is also reasonable to consider genotyping additional nonsynonymous coding SNPs in candidate genes regardless of their frequency.\(^{67,112}\) Because it is more likely a priori that exonic SNPs will produce functional effects that will be associated with disease status. Ideally, study power should be adequate to detect small effects. Should promising associations arise, associations must be replicated in ≥1 and preferably 2 other independent populations.

Genome-Wide Association Studies

Although a candidate gene approach is a valuable first step in trying to identify PAD susceptibility genes, even with well-informed choices of candidate genes, this approach is limited to a priori hypotheses. A logical next step is to perform a more comprehensive and unbiased scan of variations in all genes or in the entire genome, in what is commonly referred to as a genome-wide association (GWA) study.

The GWA design calls for the use of high throughput genotyping platforms to genotype as many SNPs in the genome as possible irrespective of the location of these SNPs relative to genes. In this respect, the GWA design is similar to linkage studies, because no previous information on gene function is necessary to select SNPs for genotyping. However, the GWA design is unlike linkage because it maintains the ability to detect modest genetic effects expected in complex traits and, furthermore, localizes initial genetic signals to a much smaller genomic region compared with linkage. In the last few years, the GWA design has become analytically feasible primarily as a result of large-scale genetic variation characterization efforts by the International HapMap project and Perlegen in a large number of racially diverse subjects.\(^{79,113}\) These efforts have produced genotypes, frequencies, and assay information on >6 million SNPs of the estimated 11 to 15 million SNPs in the human genome with a minor allele frequency of >1%. Two companies, Affymetrix and Illumina, have developed commercially available high-throughput and highly parallel genotyping technology that allow for 650,000 to >1 million SNPs to be genotyped in a single individual at a fraction of the cost per genotype compared with other reliable platforms. To decrease costs without a substantial loss of power, 2-stage designs\(^{114–116}\) can be used. In this approach, a GWA scan is initially performed in a subset of samples (stage 1), and a portion of the markers, those showing tentative association, are genotyped in the remaining samples. Importantly for PAD, GWA studies can be performed to look for variations associated with dichotomous variables (PAD versus no PAD) or continuous variables (potentially using CT or MRI measurements).

Reports of successful localization of common genetic variants using a GWA design influencing complex traits, such as macular degeneration,\(^ {117}\) obesity,\(^ {118,119}\) inflammatory bowel disease,\(^ {120}\) rheumatoid arthritis, bipolar disorder, myocardial infarction,\(^ {121,122}\) DM1 and DM2,\(^ {123–128}\) have recently been published. Although the studies differed in size and genotyping platforms used, many of the major findings have been convincingly reproduced. Thus, they collectively represent an important proof of principle: the GWA approach can identify common variants that contribute to common disease.\(^ {64,129}\)

Nevertheless, applying the GWA approach to complex diseases (like PAD) requires careful consideration of study design. The results of GWA studies in complex diseases, like myocardial infarction and DM2, highlight some of the necessary study characteristics, including large sample sizes (often requiring collaboration between multiple smaller studies), due consideration of population structure, and replication in multiple cohorts. The large numbers of association tests that result from GWA scans require that an even lower probability value threshold be used to identify promising associations between variants and disease. Early simulations suggest that a \(P\) value threshold of \(\approx 5 \times 10^{-7}\) to \(5 \times 10^{-8}\) is necessary in the setting of GWA studies.\(^ {96,128}\) Again, sample sizes need to be adequate in GWA studies to maintain power to detect the modest effects expected in complex traits.\(^ {96}\) For instance, in the recent studies in DM2 and CAD, many of the susceptibility variants identified were associated with odds ratios (or hazard ratios) of \(\approx 1.15\) to 1.3.\(^ {122,124,125,127}\) Similarly, a recent GWA study identified a variant responsible for \(\approx 1\%\) of the variance in body mass index.\(^ {119}\) The GWA studies in myocardial infarction and DM2 ultimately required thousands of subjects, although the initial scans can be performed in much fewer subjects.\(^ {122}\) These studies serve as guides when considering this approach in PAD.

This field is still in its infancy, and numerous new approaches toward study design and analysis are being developed. For instance, the analysis of copy number variation\(^ {130,131}\) may prove increasingly important. Furthermore, techniques are being developed to merge data sets from different genotyping platforms, improve power through weighting schemes based on limited previous knowledge,\(^ {132}\) speed processing time of association tests,\(^ {133}\) and analyze sophisticated gene/environment interactions.\(^ {136}\) Already, methods have been developed to increase coverage of the genome through imputation of genomic data based on LD structure\(^ {123}\) or to analyze continuous data at the extremes of phenotype.\(^ {137–139}\) New recommendations have also been published by working groups within the American
Heart Association and National Human Genome Research Institute to guide study design.64,129

Although these studies have been spectacularly successful so far, it is an open question whether we will ultimately be able to identify most of the genetic variation that accounts for common diseases using this approach. It may turn out that rare variants account for much of the genetic susceptibility to common, complex diseases. These variants may not yield to this approach and might have to wait for whole genome sequencing for identification.

Summary and Call for PAD Networks

In summary, PAD is a common and complex trait of which the development is probably modified to some degree by uncharacterized genetic determinants that are independent of traditional risk factors of atherosclerosis. Based on limited studies to date, the heritability of PAD seems to be modest. More family studies of PAD are necessary to better estimate the relative importance of genetic and nongenetic determinants of disease. Carefully designed association studies of well-chosen candidate genes may identify some of these uncharacterized genetic determinants, but GWA studies are expected to be more revealing. Given the complex nature of the disease, future progress also depends crucially on close collaboration within the community of PAD investigators. Coordinated efforts by a large network of investigators are needed to develop standardized diagnostic instruments and to collect accurate phenotypic information in samples of sufficient size to perform adequately powered candidate gene and GWA studies. Within our own case-control association study of genetic determinants of PAD (http://cmedicine.stanford.edu/genePAD/index.html), we have begun to see this spirit of camaraderie play a positive role. This type of interaction will also facilitate independent replication of promising results and the subsequent elucidation of the biological mechanisms responsible for the various associations observed.

Acknowledgments

We thank Dr Hua Tang, Stanford Department of Genetics, for helpful comments.

Sources of Funding

This study was supported by grants from the National Institutes of Health (R01 HL63685 and R01 HL075774) and a National Heart, Lung, and Blood Institute National Research Service Award T32 HL07708, as well as by National Institutes of Health grant M01 RR 00070 (General Clinical Research Center, Stanford University School of Medicine). J.W.K. was supported by postdoctoral fellowship grants from Stanford University and the American Heart Association (0625154Y).

Disclosures

None.

References


