Summary
Clinical vascular disease occurs as the result of the chronic development of atherosclerosis, often with acute occlusive thrombus formation as the final event. Both atherosclerotic and thrombotic disorders are complex processes resulting from genetic and environmental interactions. We review genes currently implicated as risk factors and the approaches for identifying novel genetic risk factors.

Keywords: coronary artery disease, acute myocardial infarction, genetic risk determinants

Genes and the development of vascular disease

Carole A Foy, Peter J Grant

The pathophysiology of vascular disease

Vascular disease is the main cause of death in western countries. The major clinical manifestations are coronary artery disease (CAD), which follows the chronic development of coronary artery atheroma, and acute myocardial infarction (AMI), which is the final pathological process of plaque rupture and coronary thrombosis.

Fatty streaks in the coronary artery walls start to occur in late childhood with atherosclerotic plaques consisting of macrophage and smooth muscle cells, lipids and fibrous deposits developing in adulthood. Increases in plaque size leads to narrowing of the arterial lumen which may impede the flow of blood, resulting in CAD which can be clinically silent or may produce symptoms of chest pain or exertional dyspnoea. The subsequent rupture of unstable plaques and the formation of a thrombotic occlusion leads to the clinical syndrome of AMI. The pathogenesis of both processes are multifactorial, involving interactions between an unknown number of genetic and environmental factors. The emphasis of this review will be on the involvement and identification of the genetic components of CAD and AMI.

Evidence for a genetic contribution to vascular disease

The familial clustering of CAD and AMI has been recognised for some time. Many studies have demonstrated that having a first degree relative with premature CAD (before 60 years of age) can increase up to 10-fold the risk of the proband developing CAD, the risk being greatest the younger the patient.1-5 This increase in risk could be explained by several factors:

(a) Familial clustering of CAD risk factors

The major risk factors for CAD including total cholesterol, lipoproteins, hypertension, obesity, smoking and diabetes mellitus, often demonstrate familial clustering. The control of many of these factors has been shown to be genetically influenced. In a study of 1377 first degree relatives from 575 families it was estimated that the genetic heritability and cultural determinants of total cholesterol were 0.46 and 0.05, respectively (box 1), with the figures for high-density lipoprotein cholesterol, triglycerides, systolic and diastolic blood pressures being (0.42 and 0.1), (0.21 and 0.07), (0.48 and 0.04) and (0.35 and 0.05), respectively.6 The familial clustering of CAD could therefore be due to the effects of shared environment and inheritance of factors such as hyperlipidaemia and hypertension which predispose to CAD.

(b) Inheritance of independent genetic risk factors

Many studies have demonstrated that a family history of CAD is a strong predictor of CAD, independent of the familial clustering of risk factors.2-4 Freidlander et al examined a random sample of 1044 men aged between 40-70 years and demonstrated a significant relationship between CAD and the presence of AMI in a first-degree relative before the age of 60, independent of the effects of age, cholesterol, hypertension and high-density lipoproteins.2 Shea et al examined 223 CAD patients and 57 patients without demonstrable CAD and assigned a CAD risk score for each patient based on age, sex, blood pressure, cholesterol, cigarette smoking, diabetes and left ventricular hypertrophy.3 The presence of vascular disease was ascertained in 1319 first-degree relatives and an increased frequency of CAD was observed in the relatives of patients with CAD. There was a significant increase in the frequency and an earlier age of onset of CAD in the relatives of those patients with CAD at the lowest risk indicating that a risk associated with family history is independent of the familial accumulation of other known risk factors.

Box 1

Heritability

Heritability ($h^2$) is the proportion of the variance of a trait due to genetic causes, according to the equation $h^2 = V_g / V_p$ where $V_g$ is the variance which is genetic and $V_p$ is the total variance of the phenotype
(c) Inheritance of susceptibility

The familial clustering of CAD may also be caused by an increased familial susceptibility to the effects of cardiovascular risk factors such as cigarette smoking. Cigarette smoke is believed to increase platelet adhesiveness, thereby increasing the risk of thrombotic events. Any inherited abnormalities in coagulation may further increase the risk for CAD caused by smoking due to accentuation of a previously pro-thrombotic state. Conroy et al studied 792 males under 60 who had survived a first episode of unstable angina or AMI. No significant difference in association with a positive family history (under 60 years) or in the distribution of cigarette smoking, hypertension and hypercholesterolaemia was observed between those with and without a family history. However, patients with a positive family history were significantly younger than those with a negative history, which may in itself have been due to a greater familial susceptibility to risk factors.

It is probable that all three mechanisms are important in accounting for the familial associations of CAD and AMI and the identification of the major genetic influences involved will increase understanding of the disease and lead to future areas for therapeutic intervention and disease prevention.

The identification of genetic factors

As for all multifactorial diseases, the identification of the genes involved in atherosclerosis is going to be a difficult task. Some of the difficulties and possible solutions are listed in box 2.

Because of the complexities of the issue, it is often easier to study the effect of a gene with respect to intermediate quantitative phenotypes (such as lipid levels) which have been implicated as CAD risk factors than with respect to the disease itself. Alternatively, it may be useful to study genes associated with an intermediate clinical phenotype such as hypertension, which is a known risk

<table>
<thead>
<tr>
<th>Problems associated with studying multifactorial diseases and possible solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem</strong></td>
</tr>
<tr>
<td>• The multitude of genes involved means that in isolation each may make only a small contribution to overall risk</td>
</tr>
<tr>
<td>• Many individuals will have subclinical atherosclerosis due to the high prevalence of CAD, this will lessen the power to detect genetic determinants</td>
</tr>
<tr>
<td>• Gene/gene and gene/environment interactions (epistasis) may make individual contributions difficult to evaluate. The penetrance of many genes will be affected by genetic and environmental background, possibly leading to differing phenotypic expressions</td>
</tr>
<tr>
<td>• Certain genes determine the mean level of quantitative traits, different genes may control the variability around this mean in response to environmental and genetic factors. These genes are known as 'level' and 'variability' genes*</td>
</tr>
<tr>
<td>• The end-points of CAD and AMI may be reached by differing genes in different families and populations (genetic heterogeneity). Ethnic background may also lead to differences in the effects of genetic factors. Pima Indians from Arizona, US, have a very low prevalence of heart disease.* This is surprising as Pimas have a high prevalence of type 2 diabetes mellitus (&gt;60% of population), a potent risk factor for CAD, indicating that different factors may be important in this population</td>
</tr>
</tbody>
</table>

Box 2
factor for CAD and under a degree of genetic control. There are two major approaches for identifying genetic determinants of disease risk: the candidate gene approach and the linked marker approach.

THE CANDIDATE GENE APPROACH

**General principles**

Genes are selected for study because they code for a protein, either demonstrated or deduced from first principles, which might be involved in atherosclerosis or thrombosis. Independent genetic risk factors for atherosclerosis may include genes responsible for vessel tone and the response to inflammation and vessel wall damage. Risk factors for AMI could include genes controlling haemostasis which may predict thrombotic events. Box 3 lists some of the processes which might be involved in the development of CAD and AMI and possible candidate genes. Prospective studies of the haemostatic system and vascular risk have identified elevated circulating factor VII, fibrinogen and suppression of fibrinolysis as indicators of future coronary events. The fibrinolytic inhibitor, plasminogen activator inhibitor-1, has been found to be elevated in young survivors of AMI who developed a recurrent event. Initial levels of fibrinogen, von Willebrand factor and tissue plasminogen activator are also independent predictors of future acute coronary syndromes. These findings indicate that abnormalities in coagulation and fibrinolysis predict vascular events in apparently normal subjects and in subjects at increased risk by virtue of the presence of coronary atheroma. Many of the proteins involved in coagulation and fibrinolysis have been shown to be under some degree of genetic control and the genes involved may, therefore, be considered possible risk determinants. Lipid synthesis and processing are also under genetic control with the strongest evidence of a link between a genetic defect and CAD provided by the autosomal dominant condition of familial hypercholesterolaemia which is caused by a mutation in the low-density lipoprotein receptor gene.

If a gene is involved in the disease pathogenesis, some functional variation in the sequence should be present which is responsible for the effect. Polymorphisms occurring in the gene of interest, either functional or in strong linkage disequilibrium (box 4) to the functional site or in closely linked areas, are examined to determine associations between specific alleles and a disease or intermediate phenotypic trait. Box 5 lists several candidate gene polymorphisms which have been associated with either disease or an intermediate phenotype itself associated with disease.

**Subjects**

The frequency of individual alleles in a patient population with a clearly defined clinical phenotype and in an age-, sex- and race-matched control population are determined to detect allele frequency differences. Statistical analysis can identify alleles which are represented significantly more frequently than expected in the patient group.

In the absence of a clear association of a genotype with a disease phenotype, the association of allelic variants with the intermediate phenotypic trait for which the candidate gene codes for can also be examined. This approach will identify allelic variants which are associated with significantly differing levels of the trait under investigation which may be CAD risk predictors; the direct association between disease and gene may be masked due to other factors or too small a study group to detect an association.

---

**Processes and genes which may be involved in the development of CAD and AMI**

<table>
<thead>
<tr>
<th>Possible candidate processes</th>
<th>Proposed candidate genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• lipid synthesis and control</td>
<td>• lipo- and apolipoprotein genes and receptor genes</td>
</tr>
<tr>
<td>• blood pressure maintenance</td>
<td>• angiotensinogen and angiotensin-converting enzyme</td>
</tr>
<tr>
<td>• coagulation</td>
<td>• fibrinogen, von Willebrand factor, factor VII</td>
</tr>
<tr>
<td>• fibrinolysis</td>
<td>• plasminogen activator inhibitor-1, tissue plasminogen activator</td>
</tr>
<tr>
<td>• insulin resistance</td>
<td>• insulin receptor substrates</td>
</tr>
<tr>
<td>• cell proliferation</td>
<td>• angiotensin II receptor</td>
</tr>
<tr>
<td>• macrophage function</td>
<td>• scavenger receptor genes</td>
</tr>
</tbody>
</table>

---

**Linkage disequilibrium**

Linkage disequilibrium describes the non-random association of alleles at linked loci and is most commonly due to the shared common ancestry of populations.

---

Box 3

**Box 4**

Linkage disequilibrium describes the non-random association of alleles at linked loci and is most commonly due to the shared common ancestry of populations.
Box 5

The candidate gene approach can also be used to study affected families. The co-inheritance of disease with specific alleles can be tracked through suitable families. Linkage programs are available to determine the maximum likelihood of a gene being in linkage disequilibrium with, or the same as a disease locus.

Genetic methods
Polymorphisms can be identified by a number of methods and may vary from single point mutations through to deletions and variations in numbers of short repeated sequences. Several methods for detecting polymorphisms are listed in box 6. Most methods rely on the polymerase chain reaction to amplify small, specific regions of DNA containing polymorphic sites. In general, polymorphisms occurring within the gene of interest are likely to yield more information than polymorphisms occurring in nearby regions, due to the problems associated with recombination events. If strong associations, confirmed in independent large studies, can establish a link between an allelic form of a gene with an intermediate phenotype and a disease phenotype then the gene in question can be designated a ‘risk factor’, although this association may not hold across all populations and in all circumstances. Functional studies of any identified linked genes can be performed to determine the precise effect of specific alleles.

Discussion
The above studies can provide powerful evidence for the suspected associations of genes with disease. However, caution must be attached to any results from population-based studies as multiple comparisons may lead to positive findings by chance alone and ethnic differences in the strength of linkage disequilibrium, gene function and environmental background may all interfere with findings. Associations detected by population studies should be confirmed in family studies. One such family-based test to confirm association study findings is the transmission disequilibrium test.14 This test examines families consisting of a parent heterozygous for the marker locus identified by association studies and at least one affected offspring. The test compares the frequency of the putative disease-associated allele being transmitted to the affected child with the transmittal of the non-disease associated allele. The candidate gene approach is also dependent on the fact that the genes chosen for study must have already been identified, characterised and mapped.

Examples of the candidate gene approach

Angiotensin-converting enzyme (ACE) gene  ACE is a component of the renin–angiotensin system which acts to convert angiotensin I to angiotensin II. Angiotensin II is a powerful vasoconstrictor and smooth muscle growth factor and these processes are thought to be involved in the regulation of blood pressure. A polymorphic region of the ACE gene consisting of a 287 bp insertion/deletion sequence which is related to plasma ACE levels has been identified.15 Individuals homozygous for the deletion (DD) allele had higher levels of circulating ACE than those homozygous for the insertion (II) allele,
with heterozygous individuals (ID) having intermediate levels. An excess of the DD genotype and the associated higher ACE levels was observed in patients who had had a heart attack and in those individuals whose parents had suffered a premature MI compared to controls. These associations were greatest in those patients otherwise considered at low risk by such factors as smoking and apolipoprotein concentrations. There was evidence from population studies to suggest that the I/D polymorphism was in strong linkage disequilibrium with a major-gene effect elsewhere, which controls up to 44% of the variability in ACE levels. The association of the ACE I/D genotype with disease has been confirmed by other groups. However, several investigators have failed to find an association with disease and concluded that the DD genotype did not confer an increased risk for CAD or AMI. These discrepancies could be due to study design or to different linkage disequilibrium of the polymorphism to the functional site in different populations, and serve to highlight the difficulties in identifying genetic risk factors.

Work from our unit using single strand conformation polymorphism analysis and sequencing identified a common polymorphism consisting of a T to C substitution 93 bases from the transcription start (figure 1). In both cases (patients characterised by coronary angiography) and control groups, ACE levels were strongly associated with the polymorphism (p < 0.001), those homozygous for the mutant C allele having the highest levels (C Foy, unpublished data). Statistical analyses revealed that the polymorphism identified had some effect on ACE levels independent of the I/D polymorphism but that neither could account for all the variation in ACE levels such as to support previous suggestions of a functional polymorphic site elsewhere in the gene. The search for the functional site controlling ACE levels is continuing.

Plasminogen activator inhibitor-1 (PAI-1) gene Early studies of the gene for PAI-1 which is an inhibitor of tissue plasminogen activator identified a Hind III restriction fragment polymorphism and a dinucleotide (CA) repeat sequence that indicated a degree of regulation by both insulin and triglyceride. Subsequently, a 4G/5G polymorphism 675 bp 5' of the start site of transcription in the PAI-1 promoter was identified that related to circulating PAI-1 concentrations (figure 2). This site had functional and informative characteristics. The functional nature indicated that the 5G allele binds both a transcription and a repressor factor, whilst the 4G allele only binds the transcription factor. This has been shown to lead to higher PAI-1 secretion in cells in vitro and higher PAI-1 concentrations in population studies in the presence of the homozygous 4G state. In a study of 94 young Swedish men with a first infarct, 4G allele frequencies of 0.63 occurred in cases compared to 0.53 in controls (odds ratio 2.15, 95% CI 1.17–3.96). Patients characterised for the extent of atheroma by coronary angiography and assessed for a history of myocardial infarction demonstrated a significant association between possession of the 4G/4G genotype and a history of MI (p < 0.03, odds ratio 1.98 CI 1.1–3.7). This association was stronger in males (p < 0.007) and the mutation did not relate to the extent of atheroma. However, another study reported no association between possession of the 4G allele and risk of myocardial infarction compared to controls, although the highest levels of PAI-1 were found in the subjects with the 4G/4G genotype and the lowest in the homozygous 5G carriers.

![Figure 1](image_url)
THE LINKED MARKER APPROACH

General principles
It may not be possible to identify all of the genetic determinants of atherosclerosis using the candidate gene approach. Many genes may appear unrelated to the processes involved in atherogenesis, either because they code for products which are not easily detectable (ie, cell-associated proteins) or they make a small contribution to the disease pathogenesis. The candidate gene approach also requires that the genes have been previously identified. The linked marker approach does not rely on the presence of identifiable genes as it utilises polymorphic markers evenly spaced throughout the genome to identify chromosomal regions that contain putative atherogenic genes which cosegregate with the marker. This approach has successfully been used to identify regions involved in the pathogenesis of type 1 diabetes mellitus.29

Subjects
The identification of chromosomal regions containing possible disease genes can be accomplished in various ways. A common approach is to screen disease-affected siblings. Sibling pairs should receive the same allele at any given genotype locus from a parent in approximately 50% of cases. If a locus contains a disease gene, or is closely linked to a locus containing such a gene, then affected siblings would be expected to share the same alleles for that locus more frequently. Statistical packages are available to determine the likelihood of any particular locus being linked to a disease. This method does not require any assumptions about the mode of inheritance of disease or estimates of gene penetrance. A similar approach can be used with families containing affected members and is known as the affected-pedigree-member method. The transmission of a locus within a pedigree with a disease phenotype can be determined and marker similarity between affected members compared.

Genetic methods
Subjects are screened with polymorphic markers (usually based on polymorphic dinucleotide repeat sequences) which are spaced throughout the genome at an initial spacing of approximately 10–20 cM. Regions which demonstrate a degree of positive linkage are then ‘saturation’ mapped with more closely spaced markers (0.5–1 cM) to localise the putative gene to as small a region as possible and to exclude any false positive linkage. Once a locus has been identified as containing a putative disease gene then the task is to identify the gene contained by moving inwards from the linked markers to the actual gene. This can be done by various techniques usually based on ‘chromosome walking’ in which fragments of DNA from the region under study are isolated and cloned into suitable vectors such as yeast artificial chromosomes, to form a
Figure 3 The linked marker approach for gene identification. Patient screening is initially performed to identify regions containing putative disease genes, the disease-gene-containing region is then pinpointed using closely spaced markers. The region is then cloned and overlapping regions sequenced until the gene is reached.

CA repeat polymorphism → Linked novel gene

Screen with closer markers

Identify closest marker to the gene

Clone DNA

Probe library with polymorphic sequence

Sequence positive clones and design further probes from ends of sequence

Continue identifying overlapping clones until gene is reached

Library. This library is screened using a probe containing the polymorphic marker. The DNA from any clones which hybridise to this probe is sequenced and further probes designed from sequences at the ends, with which to pick out further clones. Eventually a set of overlapping clones are identified (a contig) one or several of which would contain the gene in question. The basis of this approach is shown in figure 3. A review of the methods used in the physical mapping of the genome is provided by Zabarovskii et al. The gene can be identified by a variety of techniques including sequence comparisons to known genes, identification of common conserved expressed gene motifs such as promoter elements, and identification of unmethylated CpG islands which are often associated with the 5' ends of expressed genes. Once a gene has been identified in this manner it is necessary to characterise it and determine whether it is the putative gene identified by the linkage studies, or whether the gene has no relationship with CAD but is simply in linkage disequilibrium with another, as yet, unidentified gene.

Discussion

The linked marker approach should be capable, in theory, of identifying all regions containing putative atherogenic genes. However, this approach, by necessity, is extremely laborious and requires the specialised facilities and equipment needed to carry out large-scale screening. The statistical difficulties involved with performing the multitude of comparisons needed, also makes the approach prone to both false positive and negative associations.

Future prospects

If the identification of the genetic factors involved in atherogenesis and thrombosis continues at the present rate it may eventually be possible to
determine an individuals' genetic 'risk profile' based on their haplotype for a panel of the most important genes. Identifying those at the greatest risk will allow for early intervention to minimise risk such as diet and exercise programmes and the administration of lipid-lowering drugs. A greater understanding of the pathophysiology of a disease obtained from identifying the genetic factors involved and their interactions with environmental factors may allow new avenues of therapeutic interventions to be explored, and possibly tailor-made interventions, depending on the specific genetic defects underlying the disease in each individual. Ethical considerations will also have to be addressed to prevent discrimination against those deemed to be at the greatest risk of developing the disease.

In the longer term, gene therapy may be possible. Gene therapy consists of introducing specific protein-encoding DNA into cells to correct the effects of a defective gene by coding for the correct form of the gene product or producing a therapeutic protein which stimulates or represses the effects of other disease-associated factors. The methods of introducing such DNA into cells is beyond the scope of this review but advances in this field are occurring at rapid rates.31

A similar approach is the use of antisense oligonucleotides. These oligonucleotides are complementary to mRNA sequences and bind to the mRNA target, preventing translation. This field of research is fairly new and there are still many problems to be overcome,32 but eventually antisense oligonucleotides could allow harmful genes to be switched off.

In summary, advances in technology are allowing the genetic complexities of vascular disease to be determined. Increased understanding of these factors and processes, in turn is presenting opportunities for the prevention and treatment of the disease.

Carole Foy is supported by the National Heart Research Fund.


Genes and the development of vascular disease.

C. A. Foy and P. J. Grant

doi: 10.1136/pgmj.73.859.271

Updated information and services can be found at:
http://pmj.bmj.com/content/73/859/271

These include:
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/