Genetic susceptibility to coronary artery disease: from promise to progress

Hugh Watkins and Martin Farrall

Abstract | Family history is an important independent risk factor for coronary artery disease (CAD), and identification of susceptibility genes for this common, complex disease is a vital goal. Although there has been considerable success in identifying genetic variants that influence well-known risk factors, such as cholesterol levels, progress in unearthing novel CAD genes has been slow. However, advances are now being made through the application of large-scale, systematic, genome-wide approaches. Recent findings particularly highlight the link between CAD and inflammation and immunity, and highlight the biological insights to be gained from a genetic understanding of the world's biggest killer.

Penetrance
The conditional probability that carriers of a specific genotype are affected by an inherited disease. If the penetrance is 100%, all carriers will be affected.

Low density lipoprotein (LDL) cholesterol
Circulating complexes of apolipoprotein B and cholesterol constitute LDL particles. Raised LDL cholesterol level is predictive of coronary artery disease.

Department of Cardiovascular Medicine and Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 9DU, UK.
e-mail: hugh.watkins@cardiov.ox.ac.uk
doi:10.1038/nrg1805
Published online
7 February 2006

Coronary artery disease (CAD) is the most common cause of death in industrialized countries and is rapidly increasing in prevalence in developing countries. Since 1990 more people have died worldwide from CAD than any other disease[1]. The age-adjusted mortality that is due to CAD in North America and Western Europe is falling each year, but as these populations age the prevalence of CAD is actually climbing, and this trend is expected to accelerate in the future owing to the burgeoning epidemic of obesity. In developing countries, the reduction in childhood mortality is expected to change demographic structures, with the result that more adults will be at risk from CAD.

CAD has long been known to 'run in families'. A small proportion of cases can be attributed to rare, highly penetrant, monogenic effects, but most are multifactorial in aetiology, involving numerous environmental and heritable risk factors. The identification of specific susceptibility genes will add to our knowledge of the molecular pathophysiology of CAD, refine the identification of high-risk individuals and suggest areas of research for drug discovery.

Because modifiable lifestyle risk factors such as smoking make an important contribution in CAD, pharmacological intervention is only one part of any preventive strategy. Nevertheless, the demand for effective drugs is considerable, and some important advances in preventive treatment have already been made. Most notable has been the advent of statins — drugs that lower low density lipoprotein (LDL) cholesterol levels. It is important to acknowledge the contribution of human genetics to this discovery: the study of a Mendelian disorder, familial hypercholesterolaemia, implicated the LDL receptor (LDLR) in CAD, which advanced our understanding of the cholesterol metabolic pathways that have a key role in CAD pathogenesis.

Despite this success, new drugs — and therefore new drug targets — are needed. In addition, available markers are not adequate for predicting individual risk of developing CAD, which places an even greater importance on increasing our understanding of CAD pathophysiology. Currently, there is no convincing animal model of CAD in a genetically tractable species. Genetically manipulated mouse models and inbred strains are informative for aspects of atherosclerosis (a key component in the development of CAD) and for lipid intermediate phenotypes, but not for important aspects of human myocardial infarction, the main clinical outcome of CAD (see below). Therefore, new pathophysiological insights into this key area are likely to depend to a large extent on human genetics.

Molecular genetic studies of rare Mendelian diseases have identified a few mutations that cause premature CAD. Genetic studies of intermediate phenotypes that are based on these findings are unravelling the heritable basis of several atherosclerosis risk factors, which are presumed to explain a fraction of an individuals' inherited risk for common, multifactorial CAD (see REF. 2 for reviews). However, there is still a large gap in our knowledge of the genes that are involved in susceptibility for most patients with multifactorial CAD.

With this in mind, we review recent advances in genetic approaches to understanding CAD. These

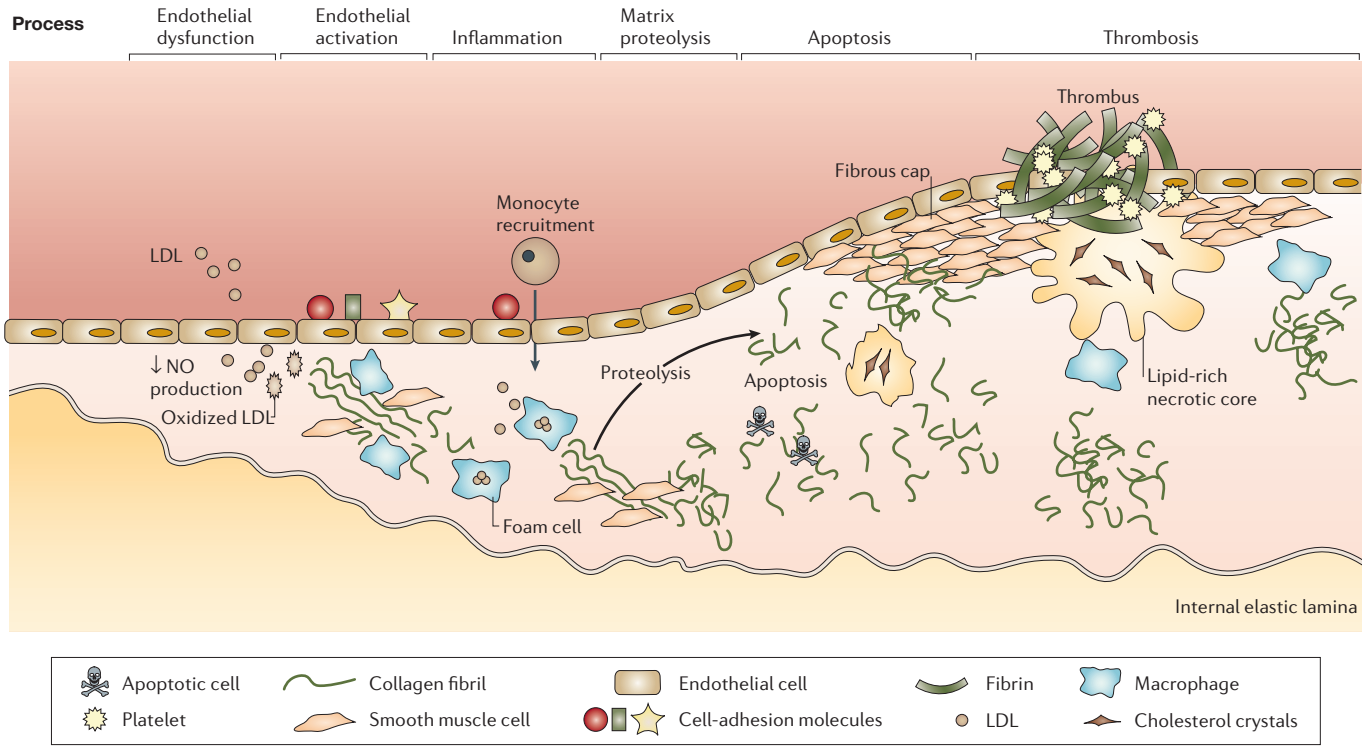

Figure 1 | **Cellular processes in the development of an atherosclerotic plaque.** Low density lipoprotein (LDL) cholesterol enters dysfunctional endothelium (which is damaged by smoking or diabetes, for example, and this is reflected by decreased nitric oxide (NO) production) and is oxidized by macrophage and smooth muscle cells. Release of growth factors and cytokines, and upregulation of adhesion molecules, attracts further monocytes. Foam cells (arising from lipid-laden macrophages) accumulate and smooth muscle cells proliferate, which results in the growth of the plaque. Inflammatory cell infiltrate, smooth muscle cell death through apoptosis, and matrix degradation through proteolysis (by matrix metalloproteinases — MMPs) generate a vulnerable plaque with a thin fibrous cap and a lipid-rich necrotic core. Plaque rupture can cause thrombosis which might be sufficient to cause vessel occlusion. Modified with permission from *Nature Reviews Drug Discovery* REF. 84 © (2004) Macmillan Publishers Ltd.

involve genome-wide linkage mapping and large-scale gene-association studies as core human genetics strategies, which are starting to produce exciting results. These have been complemented by comparative genomics studies and mapping of quantitative traits in the mouse. The genes identified through these combined approaches have diverse functions: they include transcription factors that are implicated in vasculogenesis, signalling molecules that are involved in inflammation, innate and adaptive immunity, and novel apolipoproteins, as well as genes for which the physiological roles are as yet unclear. Even with the benefit of hindsight, few of these genes are obvious candidates for CAD. The study of the genetic basis of CAD therefore emphasizes the power of genetics to generate new insights into the molecular basis of complex diseases.

The pathobiology of CAD — an overview

The clinical manifestations of CAD are the main complications of atherosclerosis — the underlying process that is also responsible for ischaemic stroke and peripheral vascular disease. Atherosclerosis is a degenerative disease that is characterized by the progressive deposition of lipids and fibrous matrix in the arterial wall (FIG. 1). The underlying pathology reflects a complex series of events, and much has been discovered about the molecular and cellular processes that are involved. The key steps are the loss of the normal barrier function of the endothelium, lipoprotein abnormalities that favour lipid entry (including high levels of LDL and/or low levels of high density lipoprotein (HDL) cholesterol), recruitment of monocytes and lymphocytes to the artery wall, and proliferation of smooth muscle cells (see REF. 3 for a review).

Although this process reflects global changes in endothelial function, the clinically important lesions are focal and tend to occur at specific sites (FIG. 2). Early lesions appear as fatty streaks, which are characterized by lipid-containing foam cells. These later develop into atherosclerotic plaques, which consist of a lipid core and an overlying fibrous cap. Several variables govern the processes that precipitate clinical events and ultimately influence disease outcome: plaques can remain silent, can progressively narrow the lumen, which restricts flow (producing angina), or can precipitately occlude vessels through acute thrombosis, which leads to myocardial infarction. The clinical entity of CAD therefore results from a particularly complex disease process, with scope for a large degree of heterogeneity among patients who have apparently similar clinical presentations and, as a corollary, extensive genetic heterogeneity.

Atherosclerosis
Also known as arteriosclerosis, this is a 'narrowing' or 'hardening' of the arteries, and is a pathological process that involves the progressive expansion of lesions (plaques) that reduce blood flow in arteries.

Intermediate phenotypes
Biochemical or physiological traits that show quantitative genetic variation. These traits are thought to represent fundamental disturbances of normal cellular processes, thereby conferring risk for complex disease.

Myocardial infarction
Commonly known as heart attack. This is caused by late-stage atherosclerosis in coronary arteries, involving plaque rupture, coronary thrombosis and heart muscle necrosis.

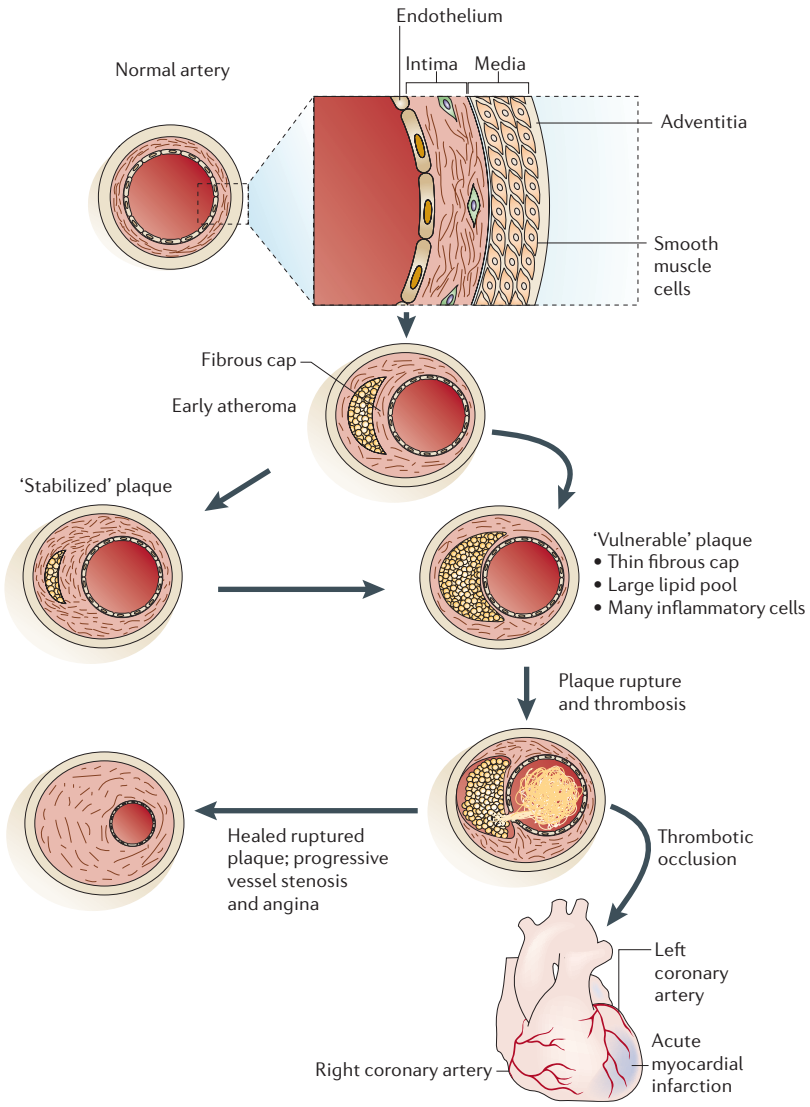

Figure 2 | **Stages in atherosclerosis and the clinical outcomes.** Atheromatous plaques progress by sequential episodes of plaque rupture and subsequent healing. As a plaque encroaches on the luminal area, blood flow becomes limiting and flow cannot increase on demand; in the heart this manifests as chest pain on exertion, which is labelled as chronic stable angina. Alternatively, rupture or erosion of a vulnerable plaque (one that has a thin fibrous cap) can precipitate an acute occlusion owing to the formation of a thrombus or blood clot, which manifests clinically as an acute coronary syndrome. Partial or transient occlusion, often accompanied by downstream embolization, results in a partial-thickness myocardial infarction, which can be confirmed by the release of troponin into the blood. Complete occlusion results in a full-thickness myocardial infarction with subsequent scar formation. The equivalent event after a plaque rupture in the carotid artery is an ischaemic stroke. Modified with permission from *Nature* REF. 3 © (2002) Macmillan Publishers Ltd.

Genetic basis of CAD as a complex trait

Molecular genetic studies of rare, Mendelian forms of the disease have identified several mutations that cause premature CAD (TABLE 1), many of which affect levels of LDL and HDL cholesterol. However, common, multifactorial CAD has also been shown to have a heritable basis. Data collated over 36 years in a cohort of 20,966 Swedish twins have shown that the heritability of fatal CAD events is 57% and 38% for men and women, respectively[4]. The influence of genes is evident across the age-range 36–86 years, declining slightly in older men (but not women), possibly owing to the increasing variance in environmental factors with age[5]. This has implications for the selection of affected individuals for genetic studies (BOX 1).

The extent to which the familial clustering of CAD can be explained by heritable quantitative variation in classical risk factors — such as LDL and HDL cholesterol and blood pressure (BOX 2) — is contentious[6,7]. In the Framingham Heart Study, a large community-based prospective study, parental history of premature coronary, cerebrovascular or peripheral vascular disease (after adjustment for age and smoking) was a strong risk factor for CAD (with odds ratios (OR) of 2.4 for men and 2.2 for women). After adjustment for several classical risk factors, family history continued to predict an excess risk (OR = 2.0 for men and 1.7 for women)[8]. Similar conclusions were drawn from the PROCAM study, in which a family history of myocardial infarction was found to be an independent predictor of CAD risk[9]. The INTERHEART study[10] confirmed that family history of CAD (after adjustment for age, sex, smoking and geographical region) is a risk factor (OR = 1.55); after adjustment for nine classical risk factors, the risk was reduced only slightly (OR = 1.45). This indicates that other genetic factors have a role, as well as those that influence known risk factors such as cholesterol levels and blood pressure.

These findings strengthen the arguments for genome-wide linkage and gene-association studies to identify CAD-susceptibility genes. These approaches do not depend on previous information about biochemical pathways or pathophysiology. They therefore complement candidate-gene studies, which concentrate on well-known risk factors and biological pathways, such as cholesterol levels, metabolism and transport.

Genetic architecture of CAD susceptibility

Although the total genetic contribution to CAD risk can be quantified, the determination of the size and number of contributing effects is impossible without identifying all CAD-susceptibility genes. Nevertheless, the genetic complexity of CAD susceptibility is estimated to be considerable. First, the list of known risk factors is long and many of these traits have their own complex genetic basis. Second, CAD is rare before the age of 50 years, and is therefore confined to aged populations. Consequently, it can be considered to be an entirely modern disease that will not have been subject to direct selective pressure as it is unlikely to have an effect on reproductive success. Variants that confer either susceptibility or protection for CAD might therefore have evolved neutrally in the past, and so could be present at a wide range of frequencies[11]. Only if these variants have pleiotropic effects that affect susceptibility to more 'ancient' diseases (for example, infectious disease) would alleles with large effects on susceptibility to CAD be expected to be present at a detectable frequency.

Within this anticipated complexity, both common and rare variants are likely to have a part in CAD

Linkage mapping
A method for localizing genes that is based on the co-inheritance of genetic markers and phenotypes in families over several generations.

Table 1 | **Mendelian diseases that involve premature coronary artery disease**

Disease	Genes	Effect of mutations	Prevalence	OMIM number
Familial hypercholesterolaemia	*LDLR*	Defective binding of LDL by receptor	1/500	#143890
Familial defective APOB	*APOB*	Reduced binding affinity of APOB to LDLR	1/3,250	#144010
Sitosterolaemia	*ABCG5, ABCG8*	Increased absorption of plant sterols	Rare	#210250
Autosomal recessive hypercholesterolaemia	*ARH*	Defective endocytosis of LDLR	Rare	#603813
APOA1 deficiency	*APOA1*	Deletion or loss-of-function mutation that leads to very low HDL	Rare	#107680
Tangier disease	*ABCA1*	Impaired cholesterol efflux in macrophages (foam cells)	Rare	#205400
Homocystinuria	*CBS*	Homocysteine increases thrombotic tendency	Rare	#236200

ABCG, ATP-binding cassette, subfamily G, member 5; APO, apolipoprotein; *ARH*, autosomal recessive hypercholesterolaemia protein; *CBS*, cystathionine β-synthase; HDL, high density lipoprotein; *LDLR*, LDL receptor.

Gene-association study
An experimental design to test whether genetic markers predict the risk of developing a disease.

Quantitative trait
A biological trait that shows continuous variation (such as height) rather than falling into distinct categories (such as diabetic or healthy). The genetic basis of these traits generally involves the effects of multiple genes and gene–environment interactions. Examples of quantitative traits are body mass index, blood pressure and blood lipid levels.

Apolipoproteins
Proteins that when complexed with lipids form lipoprotein particles that are involved in the transport and metabolism of cholesterol and triglycerides.

Ischaemic stroke
A type of cerebrovascular accident that results from occlusion of cerebral arteries owing to atherosclerosis and leading to a neurological deficit or death.

High density lipoprotein (HDL) cholesterol
Cholesterol that is contained in HDL particles; it is involved in reverse cholesterol transport and is protective against coronary artery disease.

susceptibility. However, as with any complex trait, there is some uncertainty as to which will be more important. A recent candidate-gene study to identify variants that are associated with low plasma levels of HDL cholesterol indicates that multiple rare alleles, each of which is associated with a large phenotypic effect, jointly make a substantial contribution to the population distribution of this quantitative trait[12]. This is at odds with the common disease/common variant (CDCV) model for the genetic architecture of common disease. In addition, a population study of consanguinity and the prevalence of complex diseases inferred an important role for deleterious recessive alleles in CAD, a conclusion that is also in conflict with the CDCV model[13]. These findings raise potential concerns about the ability of association studies (which have low power to detect rare variants) to be effective for CAD. However, in contrast to this, a review of 25 reported associations supports the importance of common alleles in common diseases, including CAD[14]. Furthermore, several CAD genes have been discovered in genome-wide studies that are consistent with the CDCV model, as highlighted in the following sections.

Positional cloning: linkage-based family studies

***A role for* MEF2A?** Families with Mendelian CAD without an associated dyslipidaemia syndrome are not commonly described. Nevertheless, Wang and colleagues[15] reported the positional cloning of an autosomal dominant mutation with a role in CAD on the basis of a genome-wide linkage screen of a single multiplex family. Fine mapping of the locus was limited, as only 14 informative meioses (assuming complete penetrance) were available for study, and a linked region containing at least 93 genes was identified. The authors focused their attention on the myocyte enhancer factor 2A (*MEF2A*) gene, which encodes a transcription factor, as its embryonic expression pattern indicated a role in vasculogenesis. A 21-bp in-frame deletion co-segregated with disease in this family, and was absent in 119 individuals who had normal angiograms. *In vitro* studies indicated that the deletion affects gene function. In a subsequent mutational survey of unrelated cases and normal controls, three rare variants that encode non-synonymous mutations of *MEF2A* were associated with CAD[16]. One of these variants, Pro279Leu, has also been associated with CAD risk in a Spanish cohort[17].

However, the link between *MEF2A* and CAD remains controversial. Weng and colleagues[18] carried out a resequencing survey, focusing on exonic variation in this gene among premature CAD cases. This identified five non-synonymous variants of which only one, Ser360Leu, was not found in asymptomatic control samples. Furthermore, the 21-bp deletion identified by Wang and colleagues[15] was identified in the control population and failed to co-segregate in a CAD family with a limited number of informative meioses[18]. These data do not support the *MEF2A* deletion as a high-penetrance CAD allele, but might still be compatible with a role for *MEF2A* variants in susceptibility for CAD. Alternatively, as other completely linked sequence variants must exist within the originally reported interval, an as yet unidentified gene could explain the original linkage finding.

A success for linkage: evidence for the role of leukotrienes. Genetic linkage studies have been heavily criticized as a first step in the positional cloning of susceptibility genes for common disease, as their results are notoriously difficult to replicate for low-penetrance mutations with small effects. Progress from 'linkage-to-locus' also remains a formidable technical challenge, despite the available genomic resources, and there are relatively few examples of complex disease loci that have been cloned in this way. However, one example of the successful use of such a strategy was the identification of *ALOX5AP* (which encodes 5-lipoxygenase activation protein (FLAP)) as a susceptibility gene for myocardial infarction and stroke[19].

In a linkage study of Icelandic families, moderate LOD scores (1.5–2.9) were obtained for a locus that was identified for myocardial infarction and stroke when the data were divided on the basis of sex and age-of-onset. However, none of these statistics fulfilled the stringent criteria for genome-wide significance (which is not unusual for linkage studies of complex diseases). Going against the current fashion of using SNPs, 120 microsatellite markers were then used to scan the 8-Mb linked region for evidence of a genetic association with myocardial infarction, leading to the identification of *ALOX5AP*.

The FLAP protein activates arachidonate 5-lipoxygenase (ALOX5), which catalyses the biosynthesis of the leukotriene LTA4 — a pathway that has been independently implicated in the pathogenesis of atherosclerosis in humans[20] (FIG. 3). Furthermore, mouse alleles of *Alox5* that are derived from certain inbred strains provide a high degree of protection from atherosclerosis on a high

Box 1 | **Design considerations in genetic studies of coronary artery disease**

Genetic enrichment

The multifactorial liability threshold model predicts that individuals who are affected by coronary artery disease (CAD) who have low risk profiles in terms of lifestyle factors carry greater genetic loads and therefore will be more informative in genetic studies. Some researchers have therefore biased their recruitment strategy towards younger patients. This 'best practice' is supported by a recent analysis of the Framingham Heart Study using a prospective study design to assess parents and offspring[8]. Parental premature coronary, cerebrovascular or peripheral vascular disease (in fathers aged <55 years and mothers <65 years) was found to be a stronger risk factor than non-premature disease (odds ratio (OR) = 2.0 versus 1.5 for younger versus older fathers and OR = 1.7 versus 1.1 for mothers). However, the magnitude of the genetic enrichment in young patients seems modest (particularly in men), so investigators should perhaps be wary of linkages and gene associations that seem to be confined to cohorts of young patients. Genetic enrichment can also be achieved in case–control studies by enrolling individuals with first-degree relatives who are affected as cases — for example, members of an affected sibling pair.

Clinical collections for association studies

Case–control and family-based studies provide alternative experimental designs for detecting gene associations; the latter design is usually discounted for late-onset diseases, as parental DNA samples are expected to be impractical to collect in sufficient numbers. But case–control studies for late-onset diseases such as CAD are not without their technical problems. For reasons of genetic enrichment as described above, young cases are likely to be preferentially collected. As atherosclerosis is endemic to the Western world, all controls — even in the absence of clinical signs or symptoms — will carry some lesions. Screening of healthy controls using invasive diagnostic tests (for example, coronary angiography) is currently infeasible, so it might be tempting to collect aged controls who have no symptoms and will hopefully carry relatively low genetic loads. But comparison of young cases (for example, those who are <55 years old) with aged (for example, >85 years old) controls introduces potential confounding sources. Examples of these are differential survival that is due to genes that are unrelated to CAD, genetic drift and mismatching for potential covariates. Furthermore, even low levels of population stratification might be problematic when interpreting the results of case–control studies, despite strategies to reduce confounding[74]. It therefore seems prudent to re-examine the prospects for collecting families who are suitable for gene-association studies[50] to circumvent the problems of control misspecification and admixture.

Monocytes
Circulating phagocytic white blood cells that develop into macrophages when they enter tissues.

Lymphocytes
White blood cells that are responsible for humoral and cellular immunity.

Foam cells
Macrophages that are laden with lipid (predominantly cholesterol), and that have differentiated from circulating monocytes.

Chronic stable angina
Chest pain that is experienced on exertion and results from coronary artery stenosis.

Prospective study design
An experimental design that is used in epidemiology studies. A cohort of healthy subjects are followed up over a prolonged period to determine the association of risk factors with disease.

Odds ratio
A statistical measure of the strength of a risk factor.

Pleiotropy
The phenomenon in which a single gene is responsible for several phenotypic effects.

Common disease/common variant model
A theory of causation of common, complex disease in which 'ancient' common genetic variants, some of which might have been selected for, are associated with modest disease risks.

Dyslipidaemia
Abnormality in, or abnormal levels of, lipids and lipoproteins in the blood.

Multiplex family
A family in which multiple members are affected by an inherited disease.

Angiogram
A radiological investigation that is used to clinically evaluate the severity of coronary atherosclerosis.

LOD score
The base 10 logarithm of a likelihood ratio statistic; this is a preferred statistic for linkage analysis that indicates the strength of linkage.

cholesterol (*Ldlr*$^{-/-}$ knockout) background; similar findings were obtained when heterozygous *Alox5* knockout mice were crossed with *Ldlr*$^{-/-}$ animals[21].

Identifying the causal variant in *ALOX5AP* has been more difficult. In the Icelandic study[19], a survey of SNPs in the affected families revealed many *ALOX5AP* variants that were in strong linkage disequilibrium with the disease phenotypes, but none of these had obvious potential effects on function (for example, none caused non-synonymous substitutions). In the Icelandic cohorts, a 4-SNP haplotype (HapA) showed strong association with myocardial infarction, but this was not confirmed in a cohort of British patients with myocardial infarction; however, further analysis implicated a different 4-SNP risk haplotype (HapB) in the British cohort. These different risk haplotypes in populations with differing genealogies might indicate allelic heterogeneity for CAD susceptibility.

The Icelandic HapA risk haplotype was associated with increased neutrophil release of leukotriene B4 (LTB4), which is synthesized from LTA4, indicating that the *ALOX5AP* variants might be pro-inflammatory. LTB4 activates monocytes, which migrate through arterial endothelium and differentiate into macrophages that take up lipid to become foam cells, a key early event in atheroma development (FIG. 1). Blockade of the LTB4 receptor has been shown to reduce the amount of lipid accumulation, monocyte infiltration and lesion size in *Ldlr*$^{-/-}$ or *ApoE*$^{-/-}$ (apolipoprotein E) knockout mouse models of atherosclerosis (which are susceptible to atherosclerosis owing to dyslipidaemia)[22]. Therefore, the initial linkage studies that indicated a role for *ALOX5AP* in CAD susceptibility provided the first genetic evidence for the involvement of the leukotriene pathway in this disease, a finding that has more recently been supported by candidate-gene association studies[23] (see below).

The outlook for further linkage studies. Several genome-wide linkage screens of varying sizes have been published for CAD, mostly using families that are chosen on the basis that they include survivors of early myocardial infarction (to enrich for a genetic component). Apart from the study that led to the identification of ALOX5AP, these screens — including the larger ones — have produced loci but have not yet identified genes[27–30]. Putative loci have not been replicated, but this should not be seen as surprising (or indeed worrying) as none of the studies has had the statistical power to replicate a specific locus. Rather, the signal-to-noise ratio is such that these studies have low power to detect individual susceptibility genes, and so each will tend by chance to identify a different selection. For the same reason, little is likely to be gained by pooling or meta-analysis[31] of available studies unless larger samples are studied compared with those that initially identified specific loci. In the future, putative loci that are implicated from genome-wide linkage scans are likely to inform selection of regions for systematic large-scale gene-association studies.

Candidate-gene association studies

Candidate-gene studies build on existing knowledge by investigating variation in genes that are already implicated in the pathophysiology of a disease. Studies are motivated by various types of evidence (which can also be of varying quality). For example, studies of apolipoproteins, which have well-understood roles in

Microsatellites
Short (2–5 nucleotides) direct sequence repeats, which are often highly polymorphic and are therefore useful markers in linkage analysis.

Leukotrienes
Regulators of allergic and inflammatory reactions that are produced from arachidonic acid by the lipoxygenase pathway in white blood cells.

Genetic drift
The (stochastic) change in frequency of genetic markers across generations.

atherosclerosis, have identified QTLs that are associated with CAD (for example, lipoprotein A (REF. 2)). However, human geneticists are often intrigued by less obvious connections, hoping that these will open up new opportunities for therapy — a half-way house between classical candidate-gene studies of validated biological pathways and genome-wide positional cloning projects that have the potential to identify new genes and pathways.

An overlap in inflammatory responses that was observed in chronic infection and atherosclerosis prompted investigation into the possible role of infectious agents that might trigger plaque formation[32]. This led to interest in candidate genes for CAD that are involved in innate immunity. Among these, CD14, the membrane-bound glycoprotein receptor for bacterial lipopolysaccharide (LPS), was the first gene to be implicated in CAD susceptibility. A promoter polymorphism (Cys260Thr) that has been proposed to be transcriptionally active[33] has been repeatedly studied in case–control studies[34], although no consistent association has emerged.

Box 2 | **Intermediate phenotypes for coronary artery disease**

Classical risk factors for coronary artery disease

Many of the known risk factors for development of coronary artery disease (CAD) are themselves genetically determined quantitative traits. For some, a causal pathological role can be considered proven: for example, both total and low density lipoprotein (LDL) cholesterol levels. High density lipoprotein (HDL) cholesterol levels are also strongly heritable, but confer protection from CAD. Other genetically influenced traits, which are themselves disease entities, also increase the risk of CAD, including high blood pressure, type II diabetes and metabolic syndrome. Measuring these risk factor phenotypes in family-based and case–control studies is helpful in understanding the genetic basis of these traits and exploring gene–environment interactions. These parameters are also important as covariants in analysis of risk. Furthermore, in hypothesis-free systematic linkage studies, measuring these phenotypes can help the prioritization of candidate genes within a linkage interval.

'New' plasma markers and markers of inflammation

In addition to LDL and HDL cholesterol levels, a growing number of other plasma factors have been identified that are associated with CAD risk. However, there is often uncertainty whether these are causal links or associations that are attributable to unknown confounders. Many of these traits are also strongly genetically influenced: for example, lipoprotein Lp(a) levels are up to 90% heritable, and homocysteine and fibrinogen levels are also strongly genetically influenced. More recently, molecules that are involved in inflammation have been similarly implicated, notably C-reactive protein and circulating adhesion molecules. Combined genetic and intermediate phenotype analysis within single, large study groups provides a means of comparing the risk that is attributable to genetically determined life-long variations in the measured intermediate phenotype with the expected increment in CAD risk that is based on observational epidemiological studies. This approach, sometimes referred to as 'Mendelian randomization', could potentially discriminate between causal relationships and those that are due to confounding or 'reverse causality'[75].

Vascular phenotypes

Component phenotypes of atherosclerosis have been measured in an effort to define subsets of disease mechanisms. These can be expected to reflect the influence of a smaller number of genes that should therefore have a larger effect size and be easier to identify. These phenotypes also provide surrogate end-points in longitudinal studies and allow the classification of patients into more homogenous groups. Well-validated quantitative vascular intermediate phenotypes include carotid intima-medial thickness, which shows clear heritability. More recently proposed phenotypes aim to measure earlier components of the atherosclerotic process, and these traits also seem to have a substantial genetic component. Examples of these traits are non-invasive measures of endothelial function or arterial stiffness[76–77].

Another component of innate immunity signalling, toll-like receptor 4 (TLR4) — which is an essential co-receptor for CD14 — has also been implicated in CAD susceptibility[34,35]. TLR4 recognizes exogenous pathogen-associated molecular patterns (PAMPS) (for example, LPS), as well as endogenous ligands, and activates inflammatory cells through the NF-κB pathway. A common non-synonymous polymorphism (Asp299Gly) in TLR4 has been associated with lower levels of several biomarkers of inflammation and a reduction in carotid-artery intima-media thickness (IMT)[36]. The reduction of CAD risk that is associated with this polymorphism[37] was also evident in studies of acute coronary syndrome and in a subgroup of men who were treated with statins[38]. The hypothesis that TLR4 provides a link between innate immunity, inflammation and atherosclerosis was further supported by a study of *Tlr4*/*ApoE* double-knockout mice[39]. These mice showed a modest reduction in the size of atherosclerotic plaques when compared with *ApoE* single-knockout animals, which are susceptible to atherosclerosis owing to dyslipidaemia.

However, a large association study of this polymorphism showed no difference in inflammatory markers in carriers of the TLR4 299Gly variant, and even showed a modest increased risk of myocardial infarction[40]. In addition, a study of coronary artery stenosis showed no evidence of association with this variant[41]. Such inconsistent results might be explained by differential pathogenic mechanisms in myocardial infarction (where plaque rupture leads to coronary occlusion) and other CAD phenotypes in which plaques can be large but relatively stable[40]. Alternatively, it might simply reflect the familiar life cycle of gene-association studies for complex diseases, in which small genetic effects at the limits of detection are difficult to replicate[14,42].

Two successes of candidate-gene association studies in the identification of CAD susceptibility have provided further evidence for a role of components of the leukotriene pathway, following on from the linkage studies described above that identified the role of *ALOX5AP*. In the first of these, promoter variants in the *ALOX5* gene were associated with increased carotid-artery IMT, a preclinical marker of atherosclerosis[23]. In this study, higher dietary levels of arachidonic acid (an omega-6 polyunsaturated fatty acid) increased carotid IMT, whereas fish-derived omega-3 fatty acids had the opposite effect, indicating a gene–environment interaction. Interestingly, C-reactive protein (CRP) levels were also associated with the *ALOX5* variation in this study[23]. CRP is a non-specific biomarker of inflammation and a moderately strong predictor of CAD risk[24] (BOX 2), and is thought to have a direct role in plaque development[25]. A phase II clinical trial of a compound that inhibits FLAP is underway to assess its effects on biomarkers of inflammation in cardiovascular disease[26].

Other components of the leukotriene pathway might now be considered as plausible candidate genes for CAD. The *LTA4H* gene codes for the leukotriene A4 hydrolase protein, the metallopeptidase of which catalyses the rate-limiting step in LTB4 biosynthesis. Genetic variation in

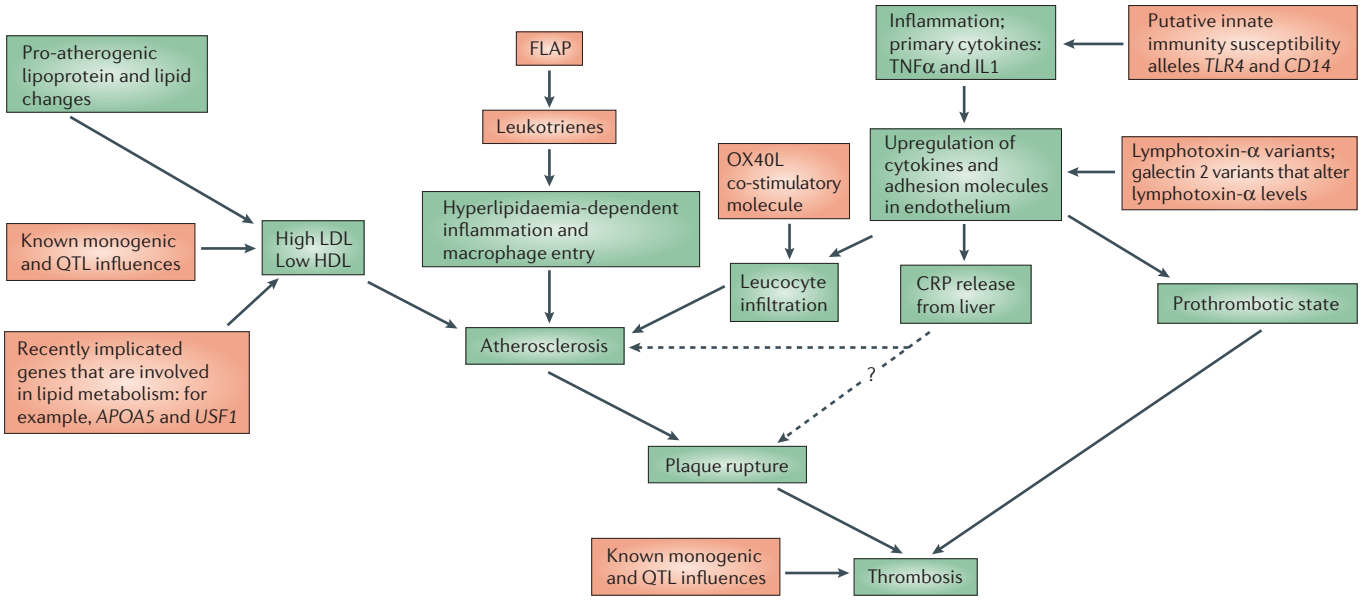

Figure 3 | **Lipid metabolism, inflammatory and thrombotic pathways that are involved in coronary artery disease.** Possible points of action of recently implicated susceptibility genes and the proteins they encode are indicated (green indicates pathophysiological changes; orange indicates genetic influences). *APOA5*, apolipoprotein A5; *CD14*, the membrane-bound glycoprotein receptor for bacterial lipopolysaccharide; CRP, C-reactive protein; FLAP, 5-lipoxygenase activation protein; HDL, high density lipoprotein; IL1, interleukin 1; LDL, low density lipoprotein; OX40L, OX40 ligand; *TLR4*, toll-like receptor 4; TNFα, tumour necrosis factor α; *USF1*, upstream transcription factor 1.

LTA4H has recently been shown to be associated with CAD (REF. 83). The strength of the association was modest in patients of European descent, but much more striking in African-American patients, which indicates that factors that are specific to ancestry groups interact to confer *LTA4H* susceptibility. These results confirm the importance of genetic variation in inflammatory pathway genes for CAD pathogenesis.

Altogether, despite some successes with replicated findings[14], single-gene association studies remain problematic, with most being poorly reproducible. Because only a small proportion of variants will have significant effect sizes, these studies have a low *a priori* likelihood of identifying the important variants. However, as small-scale studies are easy to carry out they have become numerous; some, by chance, yield a false-positive association and there is evidence of publication bias in the literature that might enrich for false positives[43]. For these reasons the field is moving towards systematic testing of large numbers of candidate genes or anonymous variants.

Large-scale, systematic, gene-association studies

Although a handful of positional cloning projects have successfully used linkage in first-pass mapping surveys of complex human diseases (for example, the study that identified *ALOX5AP* (REF. 19)), the 'smart money' is now being invested in large-scale gene-association studies[44–45]. Unlike in candidate-gene studies, where small numbers of variants that have been implicated in disease are tested for association, these studies look for association with an unbiased selection of variants that are located throughout the genome. This shift acknowledges the limitations of genetic linkage to reliably map genes that have only a modest effect on susceptibility[46]. Large-scale gene-association studies provide much greater power to identify these variants and, if enough markers can be typed at a sufficient density, this power can theoretically be harnessed to search systematically for susceptibility genes. These studies are then free from the biological assumptions that limit candidate-gene association studies. The formidable practical problems of undertaking such genome-wide association studies are being eased by The SNP Consortium and The International HapMap Project, which aim to reduce the number of variants that need to be genotyped to cover the whole genome, and by advances in genotyping technologies. Both case–control and family-based studies can be exploited to search for association; case–control series are easier to assemble, but there are issues — both general to complex traits and specific to CAD — that confer some advantages to family-based studies (BOX 1).

In a pioneering study of nearly 93,000 predominantly gene-based SNPs in a Japanese case–control cohort, polymorphisms in the lymphotoxin-α (*LTA*) gene, which encodes a member of the TNF ligand family, were found to be associated with susceptibility to myocardial infarction[47]. This association has been replicated in one of two other Japanese cohorts[48,49], as well as a study of survivors of myocardial infarction that are of white European ancestry[50]. A non-synonymous SNP (Thr26Asn) in the LTA haplotype block that was associated with CAD risk in these studies increases the transcriptional induction of the adhesion molecules VCAM1 (vascular-cell adhesion

Population stratification
Populations that are composed of subgroups with different ancestries. If the frequency of disease varies across the subpopulations then spurious (false-positive) gene associations can be detected.

Linkage disequilibrium
The non-random association of tightly linked genetic markers, or of genetic markers with disease.

Neutrophils
White blood cells that are involved in chemotaxis and phagocytosis.

Macrophages
Mononuclear phagocytic cells that are found in tissues; they are derived from circulating monocytes.

Lipopolysaccharide
(Also known as endotoxin.) A molecule with lipid and complex sugar moieties that is used by gram-negative bacteria to evade host immune defences.

Box 3 | **Combined genetic and genomic approaches**

Much excitement surrounds genomic tools that provide systematic surveys of an intact biological system, such as expression profiling, proteomics or metabolomics. Nonetheless, these tools are simply descriptors — albeit efficient and comprehensive ones — without the power of genetics to determine causality. However, recent data from crosses in model organisms show that genetic and genomic approaches can be combined — for example in studies of the genetics of gene expression. In this approach, measurements from genomic analyses — such as transcript levels — are treated as quantitative traits to be analysed genetically. So, when evidence is found for segregation with anonymous genetic markers, heritability can be inferred and the genetic factors that determine the genomic trait in question can be identified[78,79]. Therefore, instead of merely describing changes in expression that might be independent or reactive, causal associations between gene expression and disease are defined. In addition, genetic linkages that are associated with levels of particular transcripts ('expression QTLs') are probably closer to the underlying metabolic and regulatory pathways than would be any conventional quantitative trait (which might reflect the summation of several of these pathways).

Significant recent progress has been achieved in rodent models — for example, in exploring previously mapped conventional QTLs such as a known locus for cholesterol level[81]. Similarly, variation in 5-lipoxygenase (the gene implicated in human CAD by positional cloning[19]) has recently been shown to also underlie previously identified mouse QTLs for body fat, lipid levels and bone density[82]. The fundamental molecular insights that have been obtained have also been enlightening; for example, the expected modulation of single genes by adjacent *cis*-acting variation has been shown to be overshadowed by more extensive *trans*-acting loci[80]. Many of the latter modulate many genes and, surprisingly, reflect many diverse classes of genes, rather than transcription factors alone. Moreover, these *trans*-acting expression QTLs seem to cluster in 'hubs' at given chromosomal loci.

Combined genetic and genomic approaches, particularly mapping of expression QTLs, show great promise for studies of complex traits. However, these approaches have yet to be validated in natural outbred populations. If genetical genomics can similarly be exploited in human populations, then a surge of new information about quantitative heritable variation should accelerate research into the genetic basis of complex traits and diseases. But at present this remains uncertain. These approaches also tend to be limited to genomic analyses carried out on readily accessible samples that are available in large numbers of individuals; success stories so far relate to studies on blood or blood cells. For this reason also, it is not yet clear how informative these approaches will be for understanding atherosclerosis in humans.

molecule 1) and SELE (soluble selectin E), which have both been implicated in atherosclerosis and are believed to facilitate neutrophil recruitment (FIG. 1). A second, anonymous SNP in intron 1 (252A>G) was associated with increased *in vitro* transcriptional activation of LTA[47], which is consistent with other expression studies[51,52]. Both alleles are present on the haplotype that was associated with susceptibility to myocardial infarction, which might be mediated through a pro-inflammatory mechanism that leads to plaque rupture. LTA was shown to be expressed in fatty streak lesions in the aortic sinus of mice that were fed an atherogenic diet (a model of early atheromatous plaques) and *LTA*$^{-/-}$ knockout animals showed smaller lesions[53].

Building on this success, Ozaki and colleagues[54] used the *Escherichia coli* two-hybrid system and phage display to identify proteins that interact with LTA. This identified galectin 2 (LGALS2), a β-galactosidase-binding lectin that has no known physiological function. *LGALS2* SNPs that were identified through resequencing were tested for disease association in a case–control study of myocardial infarction, and an intronic SNP (3279C) showed a convincing association with myocardial infarction. *In vitro* transcription assays indicated that this allele is roughly twice as active as the 3279T allele. LGALS2 was shown to be co-expressed with LTA in smooth muscle cells and macrophages in atherosclerotic plaques, and also to colocalize with α- and β-tubulin in a pattern that is consistent with a role in intracellular protein trafficking, and so potentially relating this allele to altered LTA secretion. These intriguing findings of a possible role of LGALS2 in CAD susceptibility show how large-scale association studies and functional genomics approaches can be used together to indicate entirely unanticipated molecular mechanisms for complex disease.

Mouse genetic studies

As discussed above, although dietary and transgenic manipulation can produce a mouse model of accelerated atherosclerosis, there are not yet any convincing models of the fulminating event of plaque rupture with thrombotic occlusion[55]. However, researchers have gained insights by studying mouse models of atherosclerosis susceptibility, or by applying comparative genomic strategies that depend on the conservation of physiological and biochemical pathways in mammalian species.

A further link to immunological pathways. Atherosclerosis 1 (*Ath1*) is a QTL on mouse chromosome 1 that renders C57BL/6 mice susceptible to diet-induced atherosclerosis, whereas it has the opposite effect in C3H/He mice[56]. One of the 11 known genes in the linked region — *Tnfsf4* (tumour necrosis factor (ligand) superfamily, member 4), which encodes the ligand of the OX40 receptor — was further implicated by knockout of this gene, which resulted in smaller atherosclerotic lesions than those seen in control mice. The OX40 ligand is expressed on activated B cells, endothelial cells and macrophages. It generates co-stimulatory signals in these cells by interacting with OX40 on T lymphocytes to enhance their proliferation and differentiation, and might therefore contribute to the accumulation of these cells in atherosclerotic lesions. In two independent, moderately sized patient populations, the less common allele of one *TNFSF4* SNP was associated with an increased risk of myocardial infarction in women but not in men, which indicates a gender-specific effect[56]. Although the human data are still preliminary, these findings are notable because they indicate a candidate gene for human disease that might be particularly amenable as a drug target through receptor blockade.

New insights into dyslipidaemia. A high level of circulating triglycerides is recognized as a moderate risk factor for CAD[57] and has long been associated with genetic variation in the *APOA1–APOC3–APOA4* gene cluster[58]. APOC3 is known to downregulate triglyceride hydrolysis, but no genetic variation that has been observed in *APOC3* has explained the quantitative variation seen[59]. Research in this field was boosted by a comparative sequence analysis of the mouse and human genomes, which revealed an undiscovered member of this cluster, *APOA5* (REF. 60). Knockout and transgenic experiments showed that *APOA5* expression and triglyceride levels were inversely correlated in mice[60].

Metabolic syndrome
The occurrence of hyperinsulinaemia, glucose intolerance, dyslipidaemia, hypertension and obesity in an individual.

Pathogen-associated molecular patterns
Molecules shared by multiple prokaryotic pathogens that are relatively invariant (for example, lipopolysaccharide and flagellin).

NF-κB pathway
The signal transduction pathway that is triggered by the inducible nuclear transcription factor NF-κB. It has a central role in immunological processes.

Carotid artery
The large artery in the neck that carries blood from the heart to the brain.

These findings prompted a series of human studies that examined how genetic variation in *APOA5* — probably in combination with *APOC3* — influences triglyceride levels[61]. Data that directly map CAD susceptibility to the cluster come from the study of familial combined hyperlipidaemia (FCHL), an inherited dyslipidaemia that shows complex patterns of hypercholesterolaemia and hypertriglyceridaemia[62]. A prospective study showed that 20-year cardiovascular mortality is roughly doubled in families that show FCHL[63]. The first hint of a connection with the *APOA1–APOC3–APOA4* gene cluster came from a combined linkage and association analysis for FCHL, which implicated *APOC3* variants in susceptibility[64], although this finding was refractory to replication. However, several recent association studies support the link between the *APOA1–APOC3–APOA4–APOA5* cluster and FCHL, probably through a mechanism that involves alleles in the *APOA5* and *APOC3* loci[61].

The information that is inferred from comparative genomics is not, however, infallible. The spontaneous mouse mutant *Hyplip1* shows several metabolic attributes of FCHL, including raised levels of triglyceride, cholesterol and APOB. The mutation maps to mouse chromosome 3 (REF. 65), and a nonsense mutation in the gene that encodes thioredoxin interacting protein (*Txnip*) has been identified by positional cloning[66]. Although several cellular functions had been ascribed to this gene, a putative role in the regulation of lipid metabolism was unforeseen. The human orthologue, *TXNIP*, maps to chromosome 1q21.2 within a linked region identified in Finnish families that show FCHL[67]. However, no association of *TXNIP* with FCHL or triglyceride levels has been detected[67,68].

Instead, another positional candidate gene *USF1* (upstream transcription factor 1) that lies within 15 Mb of *TXNIP* showed evidence of association with several FCHL phenotypes[67]. This gene encodes a transcription factor that is implicated in the regulation of several apolipoproteins and other proteins with roles in lipid and glucose metabolism. No easily interpretable mutation (for example, non-synonymous SNP) was found in *USF1*, and a subsequent screen in British patients with FCHL also failed to detect any coding variant[69]. However, a 60-bp highly conserved element within intron 7 was identified that is thought to influence transcription of the gene *in vitro*[67]; this regulatory function has since been experimentally refined to a 20-bp element[70]. The importance of quantitative variation in sites that are implicated in regulation of gene transcription is increasingly being recognized[71] and involves both *cis* and *trans* variation[72], so any variants in this regulatory region are strong candidates for further analysis. Finally, it is presumably an evolutionary coincidence that *USF1* and *TXNIP* are syntenic (they are both on human chromosome 1), that *Usf1* and *Txnip* are not syntenic (being on mouse chromosomes 3 and 1, respectively), and that mutations in *Txnip* and *USF1* can both lead to FCHL phenotypes.

These studies show that there can be substantial benefits in studying genes that underlie mouse models of atherosclerosis. Sometimes, genetic variation in orthologous genes causes similar pathologies and related phenotypes in humans. Even if the human gene confers little or no susceptibility, information from mouse genetics can still direct drug-discovery programmes by identifying new biological pathways.

Conclusions

CAD genetics is a field that is gathering momentum. Although complex trait analyses for other diseases have longer histories, the public-health importance of CAD has ensured that it is now the focus of some of the largest human genetics programmes (for example, see REF. 47). Genome-wide linkage analyses indicate that although there are no loci for CAD risk that have very large effects, there are mappable loci with effects that are comparable to those in other diseases for which gene identification has been successful. Indeed, the first positionally cloned CAD genes are now emerging[19,47,67]. Progress will accelerate now that genome-wide association studies, using hundreds of thousands of SNPs, are feasible. Similarly, combined genetic and genomic approaches have the potential to reveal new realms of quantitative heritable variation that influence the biological processes underlying CAD (BOX 3). Large studies, involving thousands of cases and controls, will be needed to provide robust evidence of novel susceptibility genes in CAD because effect sizes will be modest and gene–environment interactions will undoubtedly be important. However, even a gene of modest effect can be of enormous significance if it identifies a new therapeutic target or biological pathway.

At this stage, one can only speculate on the probable biological roles of the susceptibility genes that will emerge. Many of the genes that have been implicated until now in CAD are involved in innate and adaptive immunity. Inflammation is increasingly recognized as a crucial component in the development of atheromatous plaques and thrombosis[3], so these results are not unexpected. However, the predominance of inflammatory response genes among those that have been identified might also indicate that it will only be in the context of previous selection pressure that CAD variants will be easily discernable in genome-wide studies. If so, this could constrain the types of genetic influence that will be identified; however, it could also yield unexpected synergies across fields. If selection pressures predict an overlap between genes that affect the risk of CAD and the risk of 'ancient' infectious diseases, one would also expect an overlap with susceptibility for other modern diseases that are governed by innate and adaptive immune responses. Of note, a polymorphism in the type III promoter of the MHC class II transactivator (*MHC2TA*; also known as *CIITA*) has recently been associated with increased susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction[73]. As genetic data accumulate across a range of diseases that involve inflammation, we should expect to see more connections of this type being identified and 'added value' being gained from disparate mapping studies. This is another reason to be optimistic that the substantial investment in complex trait analysis in human disease will produce the desired returns.

Acute coronary syndrome
A group of clinical symptoms that is associated with acute myocardial ischaemia following plaque rupture.

Coronary artery stenosis
The chronic narrowing of the coronary artery lumen.

TNF ligand family
Tumour necrosis factors are pro-inflammatory cytokines; the superfamily of TNF ligands have a wide range of functions, including apoptosis, B-cell and T-cell co-stimulation, and bone metabolism.

Aortic sinus
One of the anatomical dilations of the ascending aorta (the main artery that carries blood from the heart to the body), which occurs at the aortic root, just above the aortic valve.

Two-hybrid system
An experimental approach that is used to clone genes for which the protein products interact with another protein of interest.

Phage display
An experimental cloning system in which proteins are displayed on the surface of bacteriophages after fusion with coat proteins to allow the identification of novel protein interactions.

Lectin
A molecule with multiple sugar-binding sites that agglutinates cells.

1. Mackay, J. & Mensah, G. (eds) *The Atlas of Heart Disease and Stroke* (WHO, Geneva, Switzerland, 2004).
2. Lusis, A. J., Mar, R. & Pajukanta, P. Genetics of atherosclerosis. *Annu. Rev. Genomics Hum. Genet.* **5**, 189–218 (2004).
3. Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 868–874 (2002).
4. Zdravkovic, S. *et al.* Heritability of death from coronary heart disease: a 36-year follow-up of 20,966 Swedish twins. *J. Intern. Med.* **252**, 247–254 (2002).
5. Wright, A., Charlesworth, B., Rudan, I., Carothers, A. & Campbell, H. A polygenic basis for late-onset disease. *Trends Genet.* **19**, 97–106 (2003).
6. Myers, R. H., Kiely, D. K., Cupples, L. A. & Kannel, W. B. Parental history is an independent risk factor for coronary artery disease:the Framingham study. *Am. Heart J.* **120**, 963–969 (1990).
7. Wilson, P. W. *et al.* Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837–1847 (1998).
8. Lloyd-Jones, D. M. *et al.* Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA* **291**, 2204–2211 (2004).
9. Assmann, G., Cullen, P. & Schulte, H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation* **105**, 310–315 (2002).
10. Yusuf, S. *et al.* Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case–control study. *Lancet* **364**, 937–952 (2004).
11. Weiss, K. M. & Terwilliger, J. D. How many diseases does it take to map a gene with SNPs? *Nature Genet.* **26**, 151–157 (2000).
12. Cohen, J. C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869–872 (2004).
13. Rudan, I. *et al.* Inbreeding and risk of late onset complex disease. *J. Med. Genet.* **40**, 925–932 (2003).
14. Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. & Hirschhorn, J. N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genet.* **33**, 177–182 (2003).
15. Wang, L., Fan, C., Topol, S. E., Topol, E. J. & Wang, Q. Mutation of MEF2A in an inherited disorder with features of coronary artery disease. *Science* **302**, 1578–1581 (2003).
16. Bhagavatula, M. R. *et al.* Transcription factor MEF2A mutations in patients with coronary artery disease. *Hum. Mol. Genet.* **13**, 3181–3188 (2004).
17. Gonzalez, P. *et al.* The Pro279Leu variant in the transcription factor MEF2A is associated with myocardial infarction. *J. Med. Genet.* 15 Jun 2005 [epub ahead of print].
18. Weng, L. *et al.* Lack of *MEF2A* mutations in coronary artery disease. *J. Clin. Invest.* **115**, 1016–1020 (2005).
19. Helgadottir, A. *et al.* The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nature Genet.* **36**, 233–239 (2004).
A landmark positional cloning study, as it was the first to use linkage, linkage-disequilibrium mapping and functional studies to identify a CAD-susceptibility gene.
20. Spanbroek, R. *et al.* Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. *Proc. Natl Acad. Sci. USA* **100**, 1238–1243 (2003).
21. Mehrabian, M. *et al.* Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. *Circ. Res.* **91**, 120–126 (2002).
22. Aiello, R. J. *et al.* Leukotriene B4 receptor antagonism reduces monocytic foam cells in mice. *Arterioscler. Thromb. Vasc. Biol.* **22**, 443–449 (2001).
23. Dwyer, J. H. *et al.* Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. *N. Engl. J. Med.* **350**, 29–37 (2004).
24. Danesh, J. *et al.* C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N. Engl. J. Med.* **350**, 1387–1397 (2004).
25. Paul, A. *et al.* C-reactive protein accelerates the progression of atherosclerosis in apolipoprotein E-deficient mice. *Circulation* **109**, 647–655 (2004).
26. Fletcher, L. Genomics companies shop around for chemical expertise. *Nature Biotechnol.* **22**, 137–138 (2004).
27. Broeckel, U. *et al.* A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nature Genet.* **30**, 210–214 (2002).
28. Wang, Q. *et al.* Premature myocardial infarction novel susceptibility locus on chromosome 1P34–36 identified by genomewide linkage analysis. *Am. J. Hum. Genet.* **74**, 262–271 (2004).
29. Hauser, E. R. *et al.* A genomewide scan for early-onset coronary artery disease in 438 families: the GENECARD Study. *Am. J. Hum. Genet.* **75**, 436–447 (2004).
30. Harrap, S. B. *et al.* Genome-wide linkage analysis of the acute coronary syndrome suggests a locus on chromosome 2. *Arterioscler. Thromb. Vasc. Biol.* **22**, 874–878 (2002).
31. Chiodini, B. D. & Lewis, C. M. Meta-analysis of 4 coronary heart disease genome-wide linkage studies confirms a susceptibility locus on chromosome 3q. *Arterioscler. Thromb. Vasc. Biol.* **23**, 1863–1868 (2003).
32. Campbell, L. A. & Kuo, C. C. *Chlamydia pneumoniae* — an infectious risk factor for atherosclerosis? *Nature Rev. Microbiol.* **2**, 23–32 (2004).
33. LeVan, T. D. *et al.* A common single nucleotide polymorphism in the CD14 promoter decreases the affinity of Sp protein binding and enhances transcriptional activity. *J. Immunol.* **167**, 5838–5844 (2001).
34. Arroyo-Espliguero, R., Avanzas, P., Jeffery, S. & Kaski, J. C. CD14 and toll-like receptor 4: a link between infection and acute coronary events? *Heart* **90**, 983–988 (2004).
35. Pasterkamp, G. J., van Keulen, K. & de Kleijn, D. P. V. Role of Toll-like receptor 4 in the initiation and progression of atherosclerotic disease. *Eur. J. Clin. Invest.* **34**, 328–334 (2004).
36. Kiechl, S. *et al.* Toll-like receptor 4 polymorphisms and atherogenesis. *N. Engl. J. Med.* **47**, 185–192 (2002).
37. Ameziane, N. *et al.* Association of the Toll-like receptor 4 gene Asp299Gly polymorphism with acute coronary events. *Arterioscler. Thromb. Vasc. Biol.* **23**, e61–e64 (2003).
38. Boekholdt, S. M. *et al.* Variants of toll-like receptor 4 modify the efficacy of statin therapy and the risk of cardiovascular events. *Circulation* **107**, 2416–2132 (2003).
39. Michelsen, K. S. *et al.* Lack of Toll-like receptor 4 or myeloid differentiation factor 88 reduces atherosclerosis and alters plaque phenotype in mice deficient in apolipoprotein E. *Proc. Natl Acad. Sci. USA.* **101**, 10679–10684 (2004).
40. Edfeldt, K. *et al.* Association of hypo-responsive toll-like receptor 4 variants with risk of myocardial infarction. *Eur. Heart J.* **25**, 1447–1453 (2004).
41. Yang, I. A., Holloway, J. W. & Ye, S. TLR4 Asp299Gly polymorphism is not associated with coronary artery stenosis. *Atherosclerosis* **170**, 187–190 (2003).
42. Colhoun, H. M., McKeigue, P. M. & Davey Smith, G. Problems of reporting genetic associations with complex outcomes. *Lancet* **361**, 865–872 (2003).
43. Keavney, B. *et al.* Large-scale test of hypothesised associations between the angiotensin-converting-enzyme insertion/deletion polymorphism and myocardial infarction in about 5000 cases and 6000 controls. International Studies of Infarct Survival (ISIS) Collaborators. *Lancet* **355**, 434–442 (2000).
44. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Rev. Genet.* **6**, 95–108 (2005).
45. Farrall, M. & Morris, A. P. Gearing up for genome-wide gene-association studies. *Hum. Mol. Genet.* **14** (Suppl. 2), R157–R162 (2005).
46. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
47. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genet.* **32**, 650–654 (2002).
A pioneering genome-scale gene-association study for susceptibility to myocardial infarction that identified causal variants in the *LTA* gene.
48. Iwanaga, Y. *et al.* Association analysis between polymorphisms of the lymphotoxin-α gene and myocardial infarction in a Japanese population. *Atherosclerosis* **172**, 197–198 (2004).
49. Yamada, A. *et al.* Lack of association of polymorphisms of the lymphotoxin α gene with myocardial infarction in Japanese. *J. Mol. Med.* **82**, 477–483 (2004).
50. The PROCARDIS Consortium. A trio family study showing association of the lymphotoxin-α N26 (804A) allele with coronary artery disease. *Eur. J. Hum. Genet.* **12**, 770–774 (2004).
51. Knight, J. C., Keating, B. J., Rockett, K. A. & Kwiatkowski, D. P. *In vivo* characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nature Genet.* **33**, 469–475 (2003).
52. Knight, J. C., Keating, B. J. & Kwiatkowski, D. P. Allele-specific repression of lymphotoxin-α by activated B cell factor-1. *Nature Genet.* **36**, 394–399 (2004).
53. Schreyer, S. A., Vick, C. M. & LeBoeuf, R. C. Loss of lymphotoxin-α but not tumor necrosis factor-α reduces atherosclerosis in mice. *J. Biol. Chem.* **277**, 12364–12368 (2002).
54. Ozaki, K. *et al.* Functional variation in *LGALS2* confers risk of myocardial infarction and regulates lymphotoxin-α secretion *in vitro*. *Nature* **429**, 72–75 (2004).
55. Cullen, P. *et al.* Rupture of the atherosclerotic plaque: does a good animal model exist? *Arterioscler. Thromb. Vasc. Biol.* **23**, 535–542 (2003).
56. Wang, X. *et al.* Positional identification of *TNFSF4*, encoding OX40 ligand, as a gene that influences atherosclerosis susceptibility. *Nature Genet.* **37**, 365–372 (2005).
57. Austin, M. A., Hokanson, J. E. & Edwards, K. L. Hypertriglyceridemia as a cardiovascular risk factor. *Am. J. Cardiol.* **81**, 7B–12B (1998).
58. Rees, A., Shoulders, C. C., Stocks, J., Galton, D. J. & Baralle, F. E. DNA polymorphism adjacent to the human apolipoprotein A1 gene: relationship to hypertriglyceridemia. *Lancet* **1**, 444–446 (1983).
59. van Dijk, K. W., Rensen, P. C., Voshol, P. J. & Havekes, L. M. The role and mode of action of apolipoproteins CIII and AV: synergistic actors in triglyceride metabolism? *Curr. Opin. Lipidol.* **15**, 239–246 (2004).
60. Pennacchio, L. A. *et al.* An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* **294**, 169–173 (2001).
An exemplary comparative genomics study that identified and characterized a novel apolipoprotein which was later implicated in human dyslipidaemia.
61. Shoulders, C. C., Jones, E. L. & Naoumova, R. P. Genetics of familial combined hyperlipidemia and risk of coronary heart disease. *Hum. Mol. Genet.* **13**, R149–R160 (2004).
62. Goldstein, J. L., Schrott, H. G., Hazzard, W. R., Bierman, E. L. & Motulsky, A. G. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J. Clin. Invest.* **52**, 1544–1568 (1973).
63. Austin, M. A. *et al.* Cardiovascular disease mortality in familial forms of hypertriglyceridemia: A 20-year prospective study. *Circulation* **101**, 2777–2782 (2000).
64. Wojciechowski, A. P. *et al.* Familial combined hyperlipidaemia linked to the apolipoprotein AI-CII-AIV gene cluster on chromosome 11q23–q24. *Nature* **349**, 161–164 (1991).
65. Castellani, L. W. *et al.* Mapping a gene for combined hyperlipidaemia in a mutant mouse strain. *Nature Genet.* **18**, 374–377 (1998).
66. Bodnar, J. S. *et al.* Positional cloning of the combined hyperlipidemia gene *Hyplip1*. *Nature Genet.* **30**, 110–116 (2002).
67. Pajukanta, P. *et al.* Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). *Nature Genet.* **36**, 371–376 (2004).
Following up on earlier linkage to chromosome 1q, gene-association studies excluded a candidate gene (*TXNIP*) that was implicated in mouse studies and instead identified a novel susceptibility gene that has a conserved intronic regulatory element.
68. Coon, H. *et al. TXNIP* gene not associated with familial combined hyperlipidemia in the NHLBI Family Heart Study. *Atherosclerosis.* **174**, 357–362 (2004).
69. Putt, W. *et al.* Variation in *USF1* shows haplotype effects, gene:gene and gene:environment associations with glucose and lipid parameters in the European Atherosclerosis Research Study II. *Hum. Mol. Genet.* **13**, 1587–1597 (2004).

70. Naukkarinen, J. *et al.* USF1 and dyslipidemias: converging evidence for a functional intronic variant. *Hum. Mol. Genet.* **14**, 2595–2605 (2005).
71. Farrall, M. Quantitative genetic variation: a post-modern view. *Hum. Mol. Genet.* **13**, R1–R7 (2004).
72. Morley, M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
73. Swanberg, M. *et al. MHC2TA* is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. *Nature Genet.* **37**, 486–494 (2005).
74. Marchini, J., Cardon, L. R., Phillips, M. S. & Donnelly, P. The effects of human population structure on large genetic association studies. *Nature Genet.* **36**, 512–517 (2004).
75. Keavney, B. Commentary: Katan's remarkable foresight: genes and causality 18 years on. *Int. J. Epidemiol.* **33**, 11–14 (2004).
76. Benjamin, E. J. *et al.* Clinical correlates and heritability of flow-mediated dilation in the community: the Framingham Heart Study. *Circulation* **109**, 613–619 (2004).
77. Mitchell, G. F. *et al.* Heritability and a genome-wide linkage scan for arterial stiffness, wave reflection, and mean arterial pressure: the Framingham Heart Study. *Circulation* **112**, 194–199 (2005).
78. Jansen, R. C. & Nap, J. P. Genetical genomics: the added value from segregation. *Trends Genet.* **17**, 388–391 (2001).
79. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genet.* **37**, 710–717 (2005).
80. Jansen, R. C. & Nap, J. P. Regulating gene expression: surprises still in store. *Trends Genet.* **20**, 223–225 (2004).
81. Cervino, A.C. *et al.* Integrating QTL and high-density SNP analyses in mice to identify *Insig2* as a susceptibility gene for plasma cholesterol levels. *Genomics*. **86**, 505–517 (2005).
82. Mehrabian, M. *et al.* Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nature Genet*. **37**, 1224–1233 (2005).
83. Helgadottir, A. *et al.* A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. *Nature Genet*. **38**, 68–74 (2006).
84. Choudhury, R. P. Fuster, V. & Fayad, Z. A. Molecular, cellular and functional imaging of atherothrombosis. *Nature Rev. Drug Discov.* **3**, 913–925 (2004).

Competing interests statement
The authors declare no competing financial interests.

DATABASES

The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
ALOX5 | *ALOX5AP* | LGALS2 | *LTA* | *LTA4H* | *MEF2A* | *MHC2TA* | SELE | TLR4 | *Tnfsf4* | *Txnip* | VCAM1
OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
familial combined hyperlipidaemia | familial hypercholesterolaemia

FURTHER INFORMATION

The International HapMap Project: http://www.hapmap.org
The SNP Consortium: http://snp.cshl.org
WHO atlas of heart disease and stroke: http://www.who.int/cardiovascular_diseases/resources/atlas/en
Access to this interactive links box is free online.