The recent publication of the human genome sequence is widely thought to offer the opportunity for a radical change in our understanding of a variety of common human diseases. One particular hope is that the new information available from the genome sequencing effort will facilitate the conduct of population genetic studies, which will discover the genetic variants responsible for ‘complex’ or ‘polygenic’ (i.e. resulting from the action of more than one gene) diseases. In this review I attempt to put this aspiration into perspective for coronary heart disease (CHD). Firstly, I consider what is known thus far regarding the ‘genetic architecture’ of CHD. Secondly, I discuss the implications of this ‘genetic architecture’, and of certain population genetic issues, for study design. Thirdly, I consider the reasons why the results of genetic-epidemiological studies of CHD to date have tended to be discrepant, and explore strategies for increasing the reliability of such studies in the future. Finally, I consider how genetic-epidemiological studies of adequate size may be uniquely well placed to resolve particularly important and difficult controversies regarding the causative nature of hypothesized novel risk factors for CHD.

There are a number of reasons why genetic studies of complex diseases such as CHD have moved to a central position in the thinking of both geneticists and epidemiologists. Firstly, genetic studies offer the opportunity to identify determinants of disease that are very likely to be causative. This is because genotypes are unchanged throughout life, and therefore not susceptible to confounding via mechanisms related either to the presence of disease itself or the body’s response to disease (in the way that measurement of hypothesized risk factors in plasma, for example, could be). Secondly, the current pharmacological armamentarium for complex diseases consists of drugs which act on a very small number (several hundred) of the 30 000 or so genes in the human genome. Identification of novel genes that contribute to CHD risk could therefore lead to new therapeutic targets with strong molecular underpinning. Thirdly, identification of risk genotypes might enable more accurate determination of an individual’s risk of CHD than is currently possible from measurement of known risk factors. This is partly because genotypes could measure the activity of novel or otherwise unmeasurable biological pathways, and partly because genotypes would not be susceptible to short-term fluctuations and measurement error, in contrast to measurements of plasma and other quantitative phenotypic risk factors.

How large is the genetic contribution to CHD risk?

A variety of study designs have been used to attempt to quantify the size of the genetic component to CHD susceptibility. Classic twin studies compare concordance for disease status between members of monozygotic and dizygotic twin pairs—a higher relative risk of disease to the monozygotic co-twins of cases compared with the dizygotic co-twins of cases is evidence for a genetic contribution to disease aetiology. These studies generally assume that the degree of similarity in environmental exposures between members of a twin pair is the same for both monozygotic and dizygotic pairs, which is almost certainly an over-simplification; such an assumption tends to inflate the estimate of the effect of genes. Thus, estimates from twin studies should probably be regarded as upper extremes of the true genetic contribution to disease risk. The large (21 004 individuals, of whom 2810 had fatal CHD) Swedish Twin Registry study showed that the relative risk of fatal CHD in the second male twin among pairs where the first twin had died of CHD before the age of 55 was 8.1 (95% CI: 2.7–24.5) for monozygotic (MZ) twins and 3.8 (95% CI: 1.4–10.5) for dizygotic (DZ) twins. The relative risk of fatal CHD in the second female twin among pairs where the first twin had died of CHD before the age of 65 was 15.0 (95% CI: 7.1–31.9) for MZ twins and 2.6 (95% CI: 1.0–7.1) for DZ twins. These results indicate a significant genetic contribution to the risk of CHD death. However, as the age at which a co-twin died of CHD increased, the differences between MZ and DZ twins became smaller, with no significant difference between MZ and DZ twins in pairs where the first twin died after the age of 75. Thus, in common with certain environmental risk factors, the risk of CHD attributable to genes appears larger at younger ages, with the genetic risks becoming non-significant in older individuals. In this context it is worth noting that the majority of CHD deaths occur in older age groups. Together with data from other types of investigation, such as the study of adoptees and of nuclear families containing multiple affected members, these findings indicate that CHD,
in common with the majority of late-onset diseases causing the bulk of population morbidity in industrialized countries, is a disease of relatively low, although undoubtedly significant, heritability.

How many genes?

The next question is how many genes, with what sizes of effects, are responsible for this moderate heritability. Clearly, if the genes are few, with large effects, it would be expected that they will be easier to find; conversely, if the genes are many, with small individual effects, discovery will be more difficult. A number of ‘classical’ risk factors for CHD have been validated by studies carried out from the level of population to molecule (such as levels of blood pressure and plasma lipids). For most of these risk factors, some degree of familial aggregation can be demonstrated, and genetic epidemiological studies have suggested that the factors themselves are each under a similarly moderate degree of genetic control as is CHD. The population distributions of such quantitative risk factors as blood pressure and plasma lipids are continuous, and a number of investigations (dating back, in the case of blood pressure, to ‘Platt versus Pickering’) have been unable to find evidence for single ‘major genes’ (that is, genes with appreciable population frequency and large effect) leading to discontinuities in these distributions. Some Mendelian conditions have been discovered which have very substantial effects on the risk of CHD due to their effect on one or other of these risk factors, but they are rare and contribute relatively little to the population burden of disease: familial hypercholesterolaemia, the commonest example, which is due to inactivating mutations in the LDL receptor gene, affects only 1 in 500 individuals in most populations.

It could thus be argued that CHD is a singularly unpromising phenotype for genetic investigation, since susceptibility to disease is very significantly influenced by a single environmental exposure (cigarette smoking), and since the other recognized major risk factors are themselves probably under the control of multiple genes and the environment. This leads to the expectation that genes affecting CHD risk will be many, and of small individual effect. Whilst this expectation is the single most important issue in the discussion of CHD genetics, there are a number of reasons why studies of CHD endpoints rather than the study of CHD.

How many alleles?

Most genetic-epidemiological studies of common disease subscribe, explicitly or implicitly, to the ‘common disease—common variant’ model. Assuming that a given gene contributes a certain amount to the susceptibility to a common disease, this model posits that the variants within that gene conferring susceptibility will be common within the population (that is, have allele frequencies of greater than about 5%; there are probably about 7 million such common variants distributed throughout the human genome). Under the alternative ‘multiple-rare-variant’ model, a large number of variants, each with low allele frequency within the population, account for the contribution of a gene to disease risk (although the contribution of all the variants of the gene considered together may be of the same size as under the ‘common variant’ model). Simulation studies show that under the ‘multiple-rare-variant’ model, detection of causative alleles is far more difficult, and only those complex diseases in which a particular gene very substantially affected risk could be successfully investigated using realistic sample sizes. With currently available technology and population resources, it is not possible to investigate the ‘multiple-rare-variant’ model systematically. If that model truly represented the genetic architecture of CHD, then although the detection of novel causative pathways and drug targets could still be possible in very large studies, the incorporation of genetic data in risk assessment for public health purposes would be impossible, since no individual variant would have a population frequency sufficiently high to contribute materially to the population’s risk of disease.

Available data shows that either model may apply in different diseases. A good example of a common allele of large effect, which consequently contributes significantly to the population risk of disease, is the Factor V Leiden mutation which causes activated protein C resistance (APCR). APCR is present in 20–50% of patients with venous thromboembolism, and is in most cases due to a single mutation (a G to A transversion at nucleotide 1691) in the gene for coagulation factor V. This mutation is common (with an allele frequency of about 3–5%, yielding a carrier frequency of 5–10% in most Caucasian populations), and confers a relative risk of venous thrombosis to carriers of about fivefold. Conversely, in the case of Crohn’s disease, recent data have demonstrated that a number of rare variants clustered within the NOD2 gene underlie susceptibility in a particular subset of patients. So far, no allele of similar population significance to Factor V Leiden has been found in the study of CHD.

Thus, considering the available evidence, it is likely that the effect of genotype at any individual polymorphism on the risk of CHD will be small—this has been borne out in perhaps the only robust genetic association with CHD thus far described, that of the apolipoprotein E ε4 allele, which confers a relative risk of myocardial infarction of 1.2–1.3 to carriers that has been
confirmed in several thousand cases of disease.9 This contrasts with a relative risk to smokers of about 5.0 compared to non-smokers in similar age groups.10

Case-control genetic association studies of CHD

There are two broad categories of study design to address the genetics of complex traits. These are linkage studies and association studies. Linkage studies examine the co-inheritance of chromosomal segments with disease in families, whereas association studies examine correlation between the presence of specific alleles and disease in populations consisting of cases of disease and controls. Provided sufficiently large genetic effects exist, linkage studies are a highly efficient method of finding genes, and have been the mainstay of the extremely successful efforts to map and clone the genes responsible for Mendelian diseases in the modern era. However, over the last 5 years it has been increasingly accepted that, to identify loci for complex diseases of relatively low heritability such as CHD, large-scale association studies will be necessary, as such studies have greater power to resolve small effects than do linkage studies. Such association studies may be carried out in unrelated cases and controls in the same way as in classical environmental epidemiology, or in cases and members of their families (where the ‘control’ genotypes with which the case genotypes are compared are constructed from those alleles that heterozygous parents did not pass to a case). The mathematical basis for the above assertions has been elegantly discussed by Risch.11

Overwhelmingly the most frequently used association study design is that involving unrelated cases of disease and unrelated controls, principally because such studies are far easier to collect than are family-based studies. Allele frequencies at the candidate polymorphisms of interest are compared between cases and controls. There are well-recognized caveats regarding the conduct and interpretation of such studies in classical epidemiology: studies involving large numbers of cases produce more precise estimates of any association of a particular factor with risk, and in the case of small effects, very large studies may be needed; there is potential for confounding if cases and controls are not well matched; false positive results may occur if excessive subgroup analysis is carried out, particularly if the number of events in the study is small; and replication in an independent cohort provides strong evidence in favour of the correctness of the conclusions.

A caveat specific to genetic association studies of this type relates to the potential for confounding due to subtle, undetected ethnic stratification between cases and controls. If there is such an ethnic difference between cases and controls then that will be reflected by differences in allele frequency at a large number of genetic markers, few if any of which will be indicating ‘true’ association with disease arising from chromosomal proximity with a disease-causing locus. Although this issue is frequently cited as a potential cause of false positive results, it has only rarely been convincingly implicated.12,13 Even in these studies it could be convincingly argued that recent ethnic admixture of differing degrees in cases and controls was quite clear from available demographic data, and should have been detected by vigilant investigators. Recent evidence would suggest that in outbred populations selected on the basis of self-reported ethnicity, and in whom reasonable safeguards are applied in the ascertainment process to avoid clearly admixed populations, the risk of unsuspected ethnic stratification sufficient to cause false positive results is rather low.14 Additionally, recently developed mathematical techniques involving comparison of allele frequencies at a large number of randomly selected polymorphisms have the potential to discover whether significant stratification is indeed present between cases and controls in such a study.15,16

Linkage disequilibrium: islands in a stream of variation?

Association studies depend on a phenomenon known as ‘linkage disequilibrium’, which describes the degree of non-random association between a specific allele at one locus (for example, the D allele at the ACE insertion/deletion polymorphism) and a specific allele at another nearby locus. In an extreme case, an allele found at one locus is a perfect predictor of the allele that will be found at a neighbouring locus. If two such loci in ‘complete linkage disequilibrium’ are being used as markers, one of the loci is effectively redundant for mapping purposes; if one such locus is the marker polymorphism under test and the second the unknown causative polymorphism, then complete linkage disequilibrium ensures maximum power to detect the effect of the causative polymorphism using the marker. The power of a marker to detect association with disease falls rapidly as linkage disequilibrium between the marker and any neighbouring causative variant decreases. In general, the further away two loci are, the more likely that over time any linkage disequilibrium that may have existed between alleles at the two loci (as a result of mutation, admixture or a population bottleneck in the population’s history) will have been broken down by recombination between the loci. This will tend to redress the balance towards a joint genotype frequency at the two loci dependent only upon the allele frequencies at each locus. The genetic distance over which linkage disequilibrium operates thus governs how many markers one would have to type in a particular region of the genome, or in a particular candidate gene, to be certain that one had ‘captured’ all the significant variability present in the population studied, which would be necessary to definitively implicate or exclude a gene or region. Consequently, this topic has been one of the most interesting subjects to human geneticists, and the number of markers that would have to be typed in order to cover the entire genome has been variously estimated at between 50 000 and 500 000 based largely on simulated data.11,17

Recent molecular studies, however, show that linkage disequilibrium (LD) throughout the human genome is structured in blocks, within which there is very substantial disequilibrium between polymorphisms, separated by recombination hotspots within which there is little disequilibrium.5,18,19 The length of these ‘blocks’ of disequilibrium appears to be variable, but in some cases they may extend over tens of kilobases within which just one or two single nucleotide polymorphisms (SNP—polymorphisms which result from a simple transversion of one nucleotide to another, and are thus biallelic) would capture the majority of the variation in certain populations. This is extremely good news for genetic epidemiologists contemplating association studies, since the amount of genotyping required

...
to achieve coverage suddenly looks much less than before. However, much remains to be done in defining the limits of the blocks of disequilibrium, and views differ regarding whether effort should be put into a genome-wide characterization of LD blocks or whether regions in which specific candidate genes for particular conditions are located should be prioritized. Another important caveat is that patterns of LD are dependent not only on the genetic distance between polymorphisms and the presence of hotspots, but the population history. Most populations of African origin seem to have fairly extensive blocks of LD; however, a number of studies have shown that there is far less LD in African populations. Thus, long-distance mapping using LD will be far more difficult in African origin populations, with potentially a far larger number of loci needing to be typed in order to achieve coverage of a specific region or gene. It is also clear that, within a block of LD, it will be difficult to identify the causative polymorphism(s) since all the polymorphisms within a block, causative and non-causative, will be in LD with each other and therefore hard to distinguish. These differences between populations may be turned to geneticists’ advantage, however: one possible approach, successfully used by some groups, is trans-ethnic fine mapping, wherein the original ‘coarse’ localization is made in a population with extensive LD, and then fine localization made in an African origin population. This approach does, however, assume that the variants causing disease will be the same, or at least in the same part of a gene, in both populations, which may not be the case in some diseases. In general, since association studies depend not only on the relationship between specific genes and disease but also on the history of the population studied, results may not be replicable between populations unless causative polymorphisms can be tested. Whether the majority of the effort is first focussed on candidate genes or genome-wide, the recent findings regarding LD structure in a few regions have illustrated an urgent need for an internationally co-ordinated effort to describe this structure in multiple populations.

Candidate genes or ‘large hypothesis’ approach?

The majority of genetic association studies published thus far have tended to type only a few polymorphisms of genes that are regarded as candidates because of their importance in a specific biochemical pathway definitely or probably implicated in disease causation. Candidate gene association studies are likely to remain the mainstay of human genetic investigations of complex disease in the next few years, but the information available from the Human Genome Project and the parallel systematic effort to discover SNP in candidate genes and throughout the genome will radically change the scope of these studies. With respect to the number of polymorphisms needing to be typed in any candidate gene to confirm or exclude that gene from involvement in disease aetiology, it is increasingly apparent from the data discussed above regarding LD that the degree to which any particular polymorphism or set of a few polymorphisms describes all the variation in a candidate gene is variable between genes. For example, at the lipoprotein lipase (LPL) gene, a candidate for CHD susceptibility, there is a recombination ‘hotspot’ over a 1.9-kb segment in the middle of the gene, and within this hotspot a number of polymorphisms would have to be typed in order to obtain a full picture of the variability even in this short region. However, outside the ‘hotspot’, the remaining variation in the gene can be described by typing of a relatively few polymorphisms. In the angiotensin-I converting enzyme (ACE) gene, an extensively investigated candidate gene for CHD risk, it has been shown that there are only three common haplotype groups (a haplotype describes the phased array of genotypes along a chromosome) in Caucasian populations, and that one of the haplotype groups arises from an ancestral recombination event between the two commonest haplotypes around exon 7 of the gene. Thus, the majority of the significant population variability at this locus (which extends over 25 kb) in Caucasians could be described by the genotyping of just two of the 78 polymorphisms described within the gene, one on either side of the recombination event in the middle of the gene. The principal change in candidate gene association studies within the next few years will be the explicit consideration of the haplotype structure of the genes under investigation and selective genotyping of ‘haplotype-tagging’ SNP selected based on knowledge of the variation present in the population. This may not be so easy—despite the very large number of SNP (in excess of 2 million) that have been deposited in the SNP Database, a recent study showed that, for several candidate genes, the available SNP did not adequately describe the variation present. Thus, a large amount of preliminary work on SNP definition, haplotype identification and SNP selection in candidate genes is to be expected in the next few years, although several groups have already produced large datasets describing the frequent SNP in exonic and 5’ sequences for a number of candidate genes for CHD.

The increasing facility with which SNP can be typed, the availability of a large number of frequent SNP in databases, and the recent information regarding the structure of LD in a variety of genomic regions, raise the possibility that a ‘large hypothesis’ experiment consisting of a genome-wide SNP survey in CHD may be possible within a few years. An advantage of such an approach would be its unbiased nature and its potential to discover genes participating in novel causative pathways, which might not have been studied in a candidate gene approach. Further advances in genotyping technology, in particular with regard to cost and throughput, would be necessary before such an approach could be successful; also, the interpretation of the vast amounts of data such an experiment would generate would present unique challenges.

Why have results thus far been so unreliable?

The successes of classical epidemiology in the study of CHD over the past 30 years have two foundations. Firstly, the factors that have been successfully investigated for association with disease can be measured fairly accurately; and secondly, the sizes of the effects of the established risk factors on disease are large. How do genetic epidemiological studies of CHD measure up to these benchmarks? Certainly, genotype data is very reliable when compared to the measurement of a plasma risk factor such as fibrinogen level; genotype data is not subject to change over time and thus issues of ‘regression dilution’ do not arise. However, as has been discussed, the size of the effect of any particular polymorphism on risk is likely to be very much smaller
than the effect of the ‘environmental’ risk factors, such as smoking, that have been identified in the classical studies. It is a founding epidemiological principle that the smaller the effect that is being sought, the larger the planned investigation must be in order to produce reliable results; an ill-founded optimism regarding the likely size of genetic effects has resulted in the plethora of under-powered, mutually contradictory studies that characterize the literature on the genetic epidemiology of CHD. In my view, there is an urgent need for a radical change in approach amongst investigators and journal editors to ensure that the field does not fall into total disrepute. Unfortunately, genetic association studies that were far too small to be robust have tended to be accepted for publication if positive results have been obtained. This has been particularly true if a novel aetiological insight has been claimed. A number of meta-analyses have convincingly shown the presence of such publication bias for some particularly intensively investigated polymorphisms, including the I/D polymorphism of the ACE gene.

Perhaps the principal conceptual difficulty in the conduct of these studies is that a far more Bayesian viewpoint than is customary in many biological experiments is necessary. There are estimated to be approximately 5 million single nucleotide polymorphisms with a minor allele frequency of ≥10%, and 11 million SNP with a minor allele frequency of ≥1%. The prior probability that any of these is causally associated with CHD is vanishingly small. Even when polymorphisms are randomly selected in candidate genes known or suspected to lie within a causal pathway for disease, the prior probability of association, while not numerically calculable, must remain low. In this situation, differences in allele frequency between cases and controls significant at the conventionally accepted $P < 0.05$ level are more likely to represent a false than a true positive, because the prior probability of such a difference being causal is so low. This problem tends to be exacerbated by the ease with which genetic polymorphisms can be typed and multiple hypotheses tested with post hoc justification. Although the appropriate threshold for significance in such studies is not, so far, agreed, some authorities have suggested it should be very much more extreme than hitherto customary (for example Risch has suggested $P = 5 \times 10^{-8}$ for randomly selected SNP). In order to provide adequate power to detect small effects even with substantially less stringent significance criteria (for example, $P < 0.001$, which if adopted would still result in the reclassification of most ‘positive’ CHD case-control gene-association studies as negative), far larger studies than has hitherto been usual will be required.

It may be possible to revise the estimate of an appropriate level of significance based on factors which could, in the case of particular polymorphisms, significantly affect what the prior probability of a causal association is likely to be. In the case of the the ApoE 2/3/4 polymorphism, genotype significantly affects plasma levels of cholesterol and its subfractions, certain genetic defects in ApoE are associated with severe lipid abnormalities and premature CHD, and the different ApoE isoforms have been shown to have biologically significantly different properties in a variety of actions potentially relevant to the development of atherosclerosis. All these factors act to increase the prior probability that an association of ApoE genotype with CHD discovered at a given level of statistical significance is causal. The long history of robust genetic associations at the HLA locus with a variety of autoimmune diseases (most of which were initially discovered and replicated by association studies in unrelated cases and controls, and subsequently confirmed by family studies) provides another example which strongly confirms the view that case-control studies can reliably detect genetic effects if the prior probability of a causal association is sufficiently high.

It is highly likely that individuals with certain genotypes are differentially susceptible to the effects of environmental exposures, with particularly adverse consequences of an environmental exposure in those of particular genotype (‘gene-environment interaction’). A number of studies have focussed on attempts to identify such ‘interactions’ by examining the effects of environmental exposures in subgroups of genotyped cases and controls. In almost every case, such analysis has included very small numbers of cases and controls in the subgroups claimed to show differences, and the caveats of classical epidemiology regarding undue emphasis on the results of such analysis have gone unheeded. Since the likely effect of any particular allele overall is likely to be small, attempts to detect heterogeneity between subgroups of individuals carrying such an allele are likely to be very unreliable unless they are carried out in far larger numbers (many thousands) of cases of disease than has been hitherto usual. In practical terms, Clayton and McKnight have recently pointed out that the public health benefits of targeting interventions towards those individuals of particular genotype that are unusually susceptible to a specific adverse environmental factor are likely to be limited, and that greater benefits are likely to result from interventions directed at the whole population.

This calls into question the emphasis on gene-environment interaction currently influencing the design of several large cohort studies into the genetic epidemiology of complex disease, including a study of 500 000 individuals over 10 years in the UK (for details see http://www.wellcome.ac.uk/en1/bioenpop.html).

Three issues, therefore—the small anticipated size of any genetic effect, the need for high levels of statistical significance to counteract the very low prior probability of association, and the reasonable aim to examine the effect of particular polymorphisms in a limited number of carefully chosen subgroups—all lead to the conclusion that far larger studies than have been hitherto usual (involving many thousands of cases of disease and controls) are necessary in this area for reliable results to be obtained. However, the number of such very large studies will be few—where does this leave those investigators with access to small or medium-sized cohorts who wish to contribute to the field, and how will replication of the results of a very large study take place? In my view, collaborative efforts involving deposition of raw genotype and phenotype data, suitably anonymized, in a central database that can be accessed and analysed by all contributors are needed. Such databases could be set up by particular groups with agreed co-ordinating responsibilities, by national funding agencies, or by journals with a particular interest in resolving whether specific hypothesized associations are indeed real. Science pursued according to the traditional competitive model has not produced impressive results in this uniquely difficult area and large sums of money have probably been wasted in genotyping studies that were individually far too small to yield accurate results.
Genetic associations as a test for causality

One particularly important contribution that genetic-epidemiological studies could make to the understanding of the pathogenesis of CHD is by overcoming some of the limitations of classical epidemiology and clinical trials through ‘Mendelian randomization’ to test hypotheses about the causality of particular pathways in CHD. A variety of factors that can be measured in plasma (for example, fibrinogen, C-reactive protein, and homocysteine) have been proposed as novel independent risk factors for CHD, however, the causative nature of the associations that have been observed are uncertain. Estimation of whether there is any true association has been difficult, not least because these factors tend to be associated with the presence of established risk factors (such as smoking, and plasma cholesterol levels) with which they may be confounded. Although statistical adjustment can be made in prospective and case-control studies of these factors and risk, it may be imperfect. An alternative approach would be direct pharmacological intervention in a clinical trial to alter the level of the risk factor; however, suitable specific agents are not always available (for example fibrates, which lower plasma fibrinogen level, also affect plasma lipid levels). Could genetic studies clarify this situation? Suppose a polymorphism was available that was reliably associated with differences in the plasma level of one of these hypothesized risk factors. Individuals could be considered to have been randomized at conception, in a Mendelian fashion, to lifelong differences in their plasma level of the risk factor. Provided the polymorphism acted only to change the level of the plasma risk factor, and did not produce a protein with significantly different functional characteristics, then any association between genotype and disease that was present would be expected to be of a size predictable from the ‘component’ associations between genotype and plasma levels of the risk factor, and between the risk factor and disease. If no association between genotype and disease were observed, despite the presence of associations between genotype and risk factor, and between risk factor and disease, it could be inferred that the association between risk factor and disease was falsely inflated, since the genetic result would be independent of measurement error and confounding (except by associated, untyped polymorphisms that affected the functionality of the protein). In this way the causality of the risk factor could be assessed. But, in order to have adequate power to detect whether there were significant differences between the ‘predicted’ and ‘observed’ associations of any polymorphism with disease, very large studies would be needed—for example, a study of over 4000 cases was required to confirm or exclude the predicted relative risk of 1.2 for CHD associated with a polymorphism of the beta-fibrinogen promoter. Polymorphisms have been described which would be appropriate for typing in a similar fashion to assess the causality of plasma levels of homocysteine and C-reactive protein. A polymorphism (C677T) of the MTHFR gene has been associated with elevated homocysteine in many studies, and in some of these the risk of CHD has been investigated for association with genotype. However, the number of events in these studies have been too small for robust conclusions to be reached, and the hypothesis requires testing in a large study. Recently a polymorphism (-174C/T) of the interleukin-6 gene has been shown in one study to be associated with baseline levels of plasma CRP, though this result awaits confirmation. The capacity of genetic-epidemiological studies to discriminate between causal and non-causal associations (in pathways potentially amenable to intervention) may turn out to be the most significant contribution of such studies to patient health.

Conclusions

Genetic-epidemiological studies of CHD have been dogged by an over-optimism regarding the likely size of the genetic effects present which has, thus far, rendered most results very unreliable. Future studies must be much larger in order to provide reliable results. New data regarding the patterns of linkage disequilibrium in the human genome suggest that far more detailed evaluation of candidate genes or indeed of the whole genome in ‘large hypothesis’ studies will shortly be possible. The importance of ‘gene-environment interaction’ has been overstated in studies thus far, and in most cases such claimed interactions are likely to have been misleading, resulting from selective emphasis on findings in small subgroups. Discrimination between causal and non-causal associations of CHD with putative risk factors by ‘Mendelian randomization’ may prove to be a major contribution of such studies to the understanding of CHD aetiology.

KEY MESSAGES

- Genetic epidemiological studies of coronary heart disease (CHD) need to be much larger than has hitherto been usual to produce robust results.
- Recent data regarding the distribution and extent of genetic variation in the human genome suggest that ‘large hypothesis’ studies examining the contribution of many genes to CHD risk will be possible as genotyping technology advances.
- Genetic studies may be particularly useful in providing an alternative to intervention trials in the evaluation of the causality of associations between hypothesized novel risk factors and CHD risk.

References
