Analysis of gene-environment interaction in coronary artery disease

Steve E. Humphries, Maria Benedetta Donati*

*Centre for Cardiovascular Genetics, British Heart Foundation Laboratories, Royal Free and University College London Medical School, London, UK, *Department of Vascular Medicine and Pharmacology, Consorzio Mario Negri Sud, S. Maria Imbaro (CH), Italy*

(Ital Heart J 2002; 3 (1): 3-5)

Address:

Prof. Steve E. Humphries

Centre for Cardiovascular Genetics University College London Rayne Building 5 University Street London WC1E 6JJ UK
E-mail: rmhaseh@ucl.ac.uk

Research goals - Understanding pathology and genetic testing

In common with all genetic analyses of human disease, genetic research into coronary artery disease (CAD) has been based on the premises that the identification of disease-causing mutations would both help in understanding the underlying disease pathology and would lead to the development of DNA-based tests to identify those at risk. These issues will be dealt with in turn.

Understanding pathology - Mutations identify rate-limiting steps in biochemical pathways

The detection of association between a mutation in a particular enzyme and plasma levels of an "intermediate phenotype" for CAD risk (e.g. plasma levels of lipids, clotting factors or homocysteine) identifies the enzyme (and the pathway that the enzyme is involved in) as a pharmacological target. The model for how association studies allow rate-limiting steps in a biochemical pathway to be explored is shown in figure 1. In this simplified pathway, a dietary component is absorbed from the gut and is metabolised, for example in the liver, by a series of enzymes coded for by genes A-C. The end product is secreted from the liver and can be measured in plasma, and high levels are noted in epidemiological studies of CAD to be associated with increased risk. From a therapeutic point of view it may be most efficient to target the rate-limiting step in the pathway for lowering drug therapy. The question is which of the enzymes is rate-limiting in the pathway?

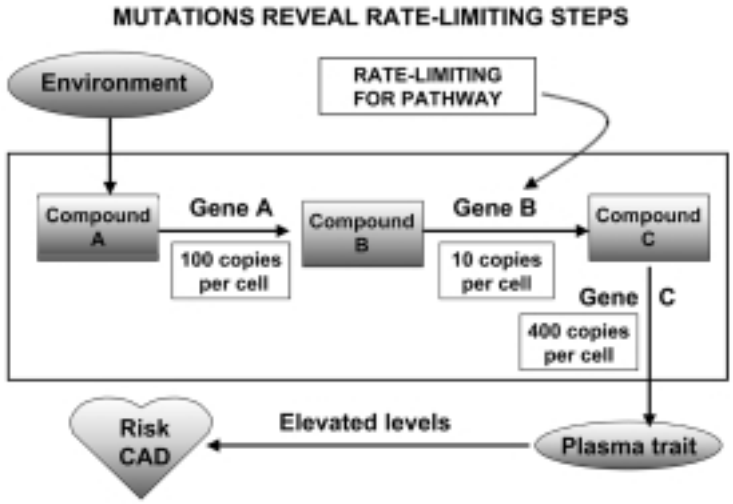

Figure 1. *Model showing how genetic variation studies can help identify rate-limiting steps in biochemical pathways. Variants in gene B that result in about for example 30% difference in levels are likely to affect plasma trait levels and therefore risk of coronary artery disease (CAD), while similar effects in genes A or C will not have significant effects on plasma trait level since they are not rate-limiting.*

Common variants in all of the genes have been identified using standard molecular screening techniques, and some of these may be directly of functional consequence while others could be in allelic association with functional variants that have yet to be found. An initial frequency comparison in subjects with high versus low plasma levels of the trait, or better an association study determining trait levels in population-based subjects with different genotypes at all loci, would reveal largest differences with the gene coding for the rate-limiting step – in this case gene B whose level in the cell is the lowest. Complete knock-outs of gene A or C would also have major effects on trait level, but would be rare (probably pediatric) metabolic defects and would not be likely to be common in the general population (because of negative selection pressure). As well as identifying gene B as a pharmacological target, it may also be the basis for a CAD predisposition genetic test, especially (and maybe only) in the presence of an environment containing high levels of the dietary component.

Genetic tests and gene-environment interaction

To date there has been little progress in developing a "battery" of genetic tests for CAD, but a detailed understanding of the way common genetic variations of modest effect interact with environmental factors, suggests that this goal was always going to be difficult and technically challenging to achieve. Since we predict that the impact of single common mutations on CAD development will be modest, increasing relative risk (RR) by 20-40% at most (e.g. the well known variants of apolipoprotein E[1]), the main issue of clinical relevance is whether the conferred risk of such a mutation is very much higher in some population subgroups. To be clinically useful in any risk algorithm, we might require for any factor to have a RR of 2 or greater; which is, for example, the RR estimated to be associated with smoking habit in middle-aged men[2]. Such subgroups might be those carrying a second important mutation in another gene (i.e. those with a gene-gene interaction), and such individuals might be identified using conventional genetic strategies. Alternatively, one might identify individuals exposed to a given environment which amplifies the risk associated with that gene (i.e. gene-environment interaction). This miniseries focuses on common genetic variants (at polymorphic frequency i.e. > 1% carrier frequency) which are associated with significant excess risk only when the individual is exposed to a "high-risk" environment. The examples given show the way forward in developing potentially useful DNA tests over the next few years.

Environmental stratification and "stress-the-genotype" approaches

The classical way to search for gene:environment interaction is to carry out an association study with stratification by the presence of an environmental risk factor (e.g. smoking, diet, obesity, non-insulin-dependent diabetes mellitus, hypertension, or CAD itself), and with analysis comparing the impact on the trait of interest in carriers and non-carriers of a gene variant. The review by Talmud et al. discusses this approach and gives examples of where it has been successfully applied. A complementary and potentially more powerful approach is by a "genotype-stress" study, which has already been shown by us and other groups to be an excellent way of magnifying modest genotype effects on traits. The reviews by Montgomery et al. and by Kluft discuss this approach and give examples. Finally the review by Vischetti et al. summarizes both approaches discussing the role of environmental factors in modulating cardiovascular risk associated with variants in the promoter region of the beta-fibrinogen gene.

Analytical problems for gene-environment interaction studies

Although the failure of a second study to reproduce an association found may cast doubt on the validity of the first report, it may also reflect the presence of a potentially interesting gene:environment interaction that requires further exploration. If, as with the example of the lipoprotein lipase gene variants, smoking is the key environmental stress that amplifies or diminishes the genotype effect[3], then the different proportion of smokers in the samples being studied (and possibly the number of cigarettes being smoked), will have a major effect on the power of the sample to detect the association. Similarly for the ACE gene, if hypertension or extent of exercise is the key environmental stress that amplifies or diminishes the genotype effect[4], then the different proportion of hypertensives or athletes in the samples studied (and probably medication use and frequency of exercise) will be critical. Power calculations must be carried out knowing the prevalence of the particular environmental factor in the sample under study, and the size of the sample to be studied can be increased accordingly. From a practical point of view this means that large samples are required for interactions to be statistically significant and that the characteristics of the sample being studied must be carefully recorded and compared between samples.

The second problem in this field is that some reported associations (e.g. between a genotype and plasma levels of some CAD risk factors) may have occurred by chance alone (i.e. a type I error). The probability of an association having been observed by chance is greatly increased by the "data-dredging" approach used by some workers in the field, whereby many different genotypes and many different traits are examined; for a study using 10 polymorphisms and 10 traits (i.e. 100 contrasts), 5 would be expected to show $p < 0.05$ by chance alone, and if 8 contrasts are observed significant

at $p > 0.05$ there is no way of knowing which reflect a biological reality and which a statistical artifact. There are several ways of reducing the probability of reporting such spurious associations. One is to use the "Bonferroni" correction where essentially each p value is multiplied by the number of contrasts used, and only those still $p < 0.05$ are reported (in the example used here $p < 0.0005/100$ would be reported as significant but $p > 0.0006$ would not). For many studies this is probably too conservative approach, in part because polymorphisms in the same gene may not be wholly independent (i.e. if they show allelic association) and many CAD traits also are correlated and thus cannot be viewed as statistically independent (e.g. total and LDL cholesterol levels). A better strategy is to set up a few *a priori* hypotheses (based on biochemical pathways where the coded protein is centrally involved) that will be tested first, and to distinguish clearly these from secondary, hypothesis-generating analyses[5]. For example, a polymorphism in the gene coding for apolipoprotein B (the major protein component of LDL particles) is most likely to affect plasma apolipoprotein B or LDL cholesterol levels, and only after reporting these contrasts (uncorrected for multiple testing) are the results of contrasts reported between apolipoprotein B genotype and traits such as weight, or fat mass, or plasma levels of HDL cholesterol, triglycerides, VLDL or free fatty acids.

Ultimately the best way to reduce the possibility of spurious associations is to replicate the finding in a second independent study, confining the study to only those genotypes, traits and interactions found initially with statistical significance. Although this replication is a criterion which many top journals are now insisting upon[6] this again may be too restrictive for some workers where potentially interesting interactions may be unpublishable for the unavailability of a replication sample, but where publication will stimulate others in the field to repeat the genotyping in stored material and confirm or refute the finding.

Conclusions and future avenues of research

The identification of the relationship between genetic variants and CAD points to the rate-limiting and thus key role in the pathological processes of these proteins. It also reveals potential novel therapeutic possibilities to prevent disease in a molecularly rational manner. Many associations have proved to be robust, and for example the triglyceride-raising effect associated with the lipoprotein lipase variants (see Talmud et al.), the stress-induced greater fibrinogen-raising effect associated with the A-455 allele (see Vischetti et al.) and the left ventricular hypertrophy-raising effect associated with the ACE DD genotype upon exercise (see Montgomery et al.) have now been replicated and therefore are very unlikely to have been observed by chance alone, although the precise molecular mechanism of all of these effects remains to be elucidated. However before contemplating the use of any genetic factor (or gene-environment interaction) in clinical risk assessment[7], it would be reasonable as a minimum to propose that a statistically significant finding should be reproduced in three independent samples (preferably from different laboratories).

Once the effect has been confirmed by replication, and the estimates of risk made more precise and robust in larger samples, such genetic-environment information would be useful for inclusion into a CAD-risk algorithm, such has been prepared by Framingham[8] using essentially non-genetic factors. This will enable us to give genotype-specific life-style advice, or to tailor clinical and therapeutic decisions to an individual's genotype[9]. Finally, once the mechanisms of such gene-environment interactions have been understood at the molecular level it may also point to novel therapeutic possibilities, for example to block inflammatory processes during stress situations, or the novel use of available drugs such as ACE-inhibitors, to prevent disease in a molecularly rational manner. Since there are many widely used and well tolerated drugs which alter ACE activity, this raises the exciting possibility of the use of these drugs in order to maintain human health in hitherto unexplored ways.

Acknowledgements

Steve E. Humphries is supported by grants from the British Heart Foundation (PG2000015) and Ma-ria Benedetta Donati from the Italian Ministry of Health.

References

1. Wilson PW, Schaefer EJ, Larson MG, Ordovas JM. Apolipoprotein E alleles and risk of coronary disease. A meta-analysis. Arterioscler Thromb Vasc Biol 1996; 16: 1250-5.
2. Doll R, Hill AB. Mortality of British doctors in relation to smoking; observation on coronary thrombosis. Natl Cancer Inst Monogr 1966; 19: 205-68.
3. Talmud P, Bujac SR, Hall S, Miller GJ, Humphries SE. Substitution of asparagine for aspartic acid at residue 9 (D9N) of lipoprotein lipase markedly augments risk of ischaemic heart disease in male smokers. Atherosclerosis 2000; 149: 75-81.
4. Myerson SG, Montgomery HE, Whittingham M, et al. Left ventricular hypertrophy with exercise and ACE gene insertion/deletion polymorphism: a randomized controlled trial with losartan. Circulation 2001; 103: 226-30.
5. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990; 1: 43-6.
6. Freely associating. (editorial) Nat Genet 1999; 22: 1-2.
7. Khoury MJ, Wagener DK. Epidemiological evaluation of the use of genetics to improve the predictive value of disease risk factors. Am J Hum Genet 1995; 56: 835-44.
8. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998; 97: 1837-47.
9. Dirckx C, Donati MB, Iacoviello L. Pharmacogenetics: a molecular sophistication or a new clinical tool for cardiologists? Ital Heart J 2000; 1: 662-6.