



Editorial

Molecular genetics of myocardial infarction: many genes, more questions than answers

Giuseppe Vassalli^{a,*}, Bernhard R. Winkelmann^{b,*}

^a Department of Cardiology, BH-10, University of Lausanne, Medical Hospital (CHUV), Rue du Bugnon 21, CH-1011 Lausanne, Switzerland

^b Cooperation Unit Pharmacogenomics/Applied Genomics, University of Heidelberg, Im Neuenheimer Feld 221, D-69120 Heidelberg, Germany

This editorial refers to “Genotypes and haplotypes predisposing to myocardial infarction: a multilocus case-control study”.[†] by Martin D. Tobin et al. on page 459

In his 1999 Schattuck lecture to the Massachusetts Medical Society, Francis Collins provided an exciting vision of the development of genetically based, individualised preventive medicine by 2010.¹ According to that vision, within a decade or two it should be possible to sequence anyone’s entire genome for a laboratory cost of less than \$1000. This would allow clinicians to provide personalised advice on environmental risk and to adapt prevention and treatment to an individual’s genotype. Four years after that lecture, however, the vision of genomic medicine is not likely to materialise anytime soon.

Coronary artery disease (CAD) and myocardial infarction (MI) are complex diseases that result from a life-long interplay between genetic and environmental factors. To assess an individual’s risk, clinicians currently use score matrices based on traditional clinical and biochemical risk factors, such as the US Framingham score² and the European SCORE.³ Genotypic information of interest includes common gene variants or polymorphisms (= gene variants with a prevalence $\geq 1\%$ of the rare allele in the population) ranging from single nucleotide substitutions (single nucleotide polymorphisms or SNPs) to insertions or deletions in the DNA sequence, a variable number of repeats of two or more nucleotides (VNTRs), and haplotypes. The latter incorporate the pattern of allelic variation obtained by looking at more than one locus of genetic variation in the DNA sequence of a gene (or the entire chromosome) of a given individual.

Although many candidate genes that code for proteins known or suspected to be involved in the pathophysiology of atherosclerosis have been studied, early reports of positive associations between genotype and phenotype (generally observed in small studies) were mostly not replicated in larger cohorts. Although there can be valid reasons for non-replication, including multiple causative genes and disease mechanisms, gene–environment interactions, and population heterogeneity, conflicting reports are common and perplexing, casting doubts on this type of studies. Moreover, only genetic variants that impart a relative risk of two or more (which is comparable to that of cigarette smoking in middle-aged men) are likely to be useful in risk algorithms. It should be emphasised, however, that minor risk predictors in the overall population may still impart high risk in specific subsets, for example, due to the interaction of a gene with other genes and the environment (e.g., cigarette smoking).⁴

Myocardial infarction is the most ominous complication of CAD. Although established risk factors of CAD predict MI as well, not all patients afflicted with CAD will suffer from MI, suggesting that additional factors that promote progression from a stable to an unstable atherosclerotic plaque with plaque rupture and thrombosis are involved. Most genetic association studies of CAD and MI have focused on a single gene, thus ignoring their multifactorial pathophysiology by excluding the remainder of the genome, as well as environmental factors. However, recent advances in high-throughput genomic technology made it possible to study multiple gene polymorphisms. The GeneQuest study analysed 72 SNPs within 62 genes in 352 cases with familial premature MI and in 418 controls in a population of white Americans, identifying a potential association with three variants of the thrombospondin gene family.⁵ Another 112 polymorphisms in 71 candidate genes were analysed in a Japanese study of 2819 cases with MI and 2242 controls, identifying a potential association with SNPs in the

* Correspondence to: Tel.: +41-213140076; fax: +41-213140013.

E-mail addresses: giuseppe.vassalli@chuv.hospvd.ch (G. Vassalli); Bernhard.Winkelmann@med.uni-heidelberg.de (B. R. Winkelmann).

[†] doi:10.1016/j.ehj.2003.11.014.

connexin 37 and the plasminogen-activator inhibitor type 1 (PAI-1) genes.⁶ Finally, in this month's *Journal*, Tobin and coworkers genotyped 58 SNPs in 35 genes and estimated haplotypes at 14 loci encompassing 16 genes in 549 MI patients and 505 control subjects.⁷ Two positive findings for SNPs (the α -adducin 460trp variant and the CETP -629A variant) and two for the haplotypes (in the paraoxonase and apolipoprotein C genes) were reported. Correction for multiple testing was not performed in this study. While we wait for such findings to be replicated in future studies, important methodological questions can be addressed now.

First, despite large epidemiological studies, the true prevalence of MI remains unknown. Recently, the European and American Societies of Cardiology redefined the laboratory definition for MI, which now includes plasma troponin levels. Although this definition will pick up smaller MIs, clinically silent MIs are frequent in advanced CAD and it is only the tip of the iceberg that we realise clinically. Thus, in case-control studies of MI, we should keep in mind that cases refer to a phenotype (i.e., clinically diagnosed MI) that only reflects a subset of all MIs. Consequently, identification of appropriate controls remains a critical issue.⁸ Tobin and coworkers identified visitors without a self-reported history of CAD in the hospitals where the cases had been recruited as an approach towards a control population. Obviously, this approach cannot rule out silent CAD in controls.

Second, the authors argue that there have been few reports analysing a sample size of 500 cases and 500 controls or more, and that their study has a power of 90% to detect a 1.5-fold increase in risk of MI associated with the angiotensin-converting enzyme insertion/deletion gene polymorphism. But will that sample size be sufficient to detect a genotypic risk associated with a SNP or haplotype in a complex disease like MI where multiple testing is done?

Third, high-throughput genotyping and multiple hypothesis testing increase the risk that associations will be observed just by chance. According to J. Ott,⁹ statistical correction for multiple testing could be performed with the following formula: given that α_1 is the probability of a false positive result when m tests are done, each at the α_1 level, the probability that at least one of them leads to a false positive result is $\alpha_m = 1 - (1 - \alpha_1)^m$. For instance, to keep α_m at a level of 0.05, one must carry out each individual test at the α_1 level, which can be calculated as $\alpha_1 = 1 - (1 - \alpha_m)^{(1/m)}$. To meet an overall significance level of $\alpha_m = 0.05$, the individual significance level α_1 needs to be set at $p = 0.001$ if 50 independent tests are carried out, or at $p = 0.0005$ if 100 tests are carried out, in order to call any of them positive. None of these significance levels is met by the data of Tobin and coworkers, leaving us wondering whether we are dealing with false positive results. A recent association study of 207 SNPs in 110 genes in a subset of 305 unrelated individuals with premature MI from the GeneQuest study identified 8 genes associated with the metabolic syndrome (p -values ranging from 0.008 to 0.047), but similarly failed to correct for multiple testing.¹⁰ The general excuse of omitting this correction is the intention to do

an exploratory hypothesis-generating study. If standards were set too high, important candidate genes might be missed. Without correction for multiple testing, however, any such results are to be regarded with extreme caution before they are replicated in independent studies. Especially in the case of low allelic frequencies of the mutant allele (i.e. <10%), minor changes in prevalence may lead to significant or non-significant results (see Table 3 in Bugert et al.¹¹). Thus, validation will require large databases of carefully defined clinical phenotypes and accurate genotyping data, along with long-term follow-up for key clinical endpoints in order to assess the clinical relevance of such genetic variation.

The study by Tobin and coworkers seems to provide more questions than answers, following the general pattern of studies on genetic associations of common diseases. For most polymorphisms, therapeutic implications in complex diseases are a long way off. Thus, validated genetic approaches to the prevention and treatment of common diseases are unlikely to be available by 2010. Instead, we currently are witnessing a proliferation of pseudoscientific approaches involving direct-to-consumer marketing of genetic tests and genetic testing sites on the Internet. In the United Kingdom, for instance, a chain of health stores began to sell a series of nine genetic tests in combination with a lifestyle questionnaire, which allegedly offered "genetically tuned" advice. Not only are these tests of little usefulness, if any, but they can also cause anxiety or false reassurance.

We are entering a new era of genetic association analyses for complex diseases. High-throughput genotyping makes it possible to test as many as 1 million genetic variants in a given individual using recently developed microchips.¹² Future studies need to address evolutionary relations among haplotypes and tackle the issue of gene-gene and gene-environment interaction in order to bring us closer to genomic medicine. However, these approaches will come with new challenges. Therefore, at the moment, studies like the one reported by Tobin et al.⁷ are very important and useful, as long as we keep in mind their limitations.

References

- Collins FS. Shattuck lecture - Medical and societal consequences of the human genome project. *N Engl J Med* 1999;341:27-8.
- Wilson PWF, D'Agostino RB, Levy D et al. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837-47.
- Conroy RM, Pyörälä K, Fitzgerald AP et al. On behalf of the SCORE project group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;24:987-1003.
- Stephens JW, Humphries SE. The molecular genetics of cardiovascular disease: Clinical implications. *J Int Med* 2003;253:120-7.
- Topol EJ, McCarthy J, Gabriel S et al. For the GeneQuest investigators and collaborators. Single nucleotide polymorphisms in multiple novel thrombospondin genes may be associated with familial premature myocardial infarction. *Circulation* 2001;104:2641-4.
- Yamada Y, Izawa H, Ichihara S et al. Prediction of the risk of myocardial infarction from polymorphisms in candidate genes. *N Engl J Med* 2002;347:1916-23.

7. Tobin MD, Braund PS, Burton PR et al. Genotypes and haplotypes predisposing to myocardial infarction: a multilocus case-control study. *Eur Heart J* 2004;**25**:459–67.
8. Peters RJG, Boekholdt SM. Gene polymorphisms and the risk of myocardial infarction – an emerging relation. *N Engl Med* 2002;**347**:1624–5.
9. Ott J. Multiple tests. In Documentation to LINKAGE UTILITY programs. Free download at <http://linkage.rockefeller.edu/soft/>; (selection: <http://linkage.rockefeller.edu/ott/linkutil.htm#MULTIPLE>), 2003.
10. McCarthy J, Meyer J, Moliterno DJ et al. Evidence for substantial effect modification by gender in a large-scale genetic association study of the metabolic syndrome among coronary heart disease patients. *Hum Genet* 2003;**114**:87–98.
11. Bugert P, Hoffmann M, Winkelmann BR et al. The variable number of tandem repeat polymorphism in the P-selectin glycoprotein ligand-1 gene is not associated with coronary heart disease. *J Mol Med* 2003;**81**:495–501.
12. <http://www.perlegen.com>.