Thrombosis Research 129 (2012) 320–324

Contents lists available at SciVerse ScienceDirect

Thrombosis Research

journal homepage: www.elsevier.com/locate/thromres

Review Article

From candidate gene to genome-wide association studies in cardiovascular disease ☆

Francesco Gianfagna, Daniela Cugino, Iolanda Santimone, Licia Iacoviello *

Laboratory of Genetic and Environmental Epidemiology. Research Laboratories, Fondazione di Ricerca e Cura "Giovanni Paolo II", Università Cattolica del Sacro Cuore, Campobasso, Italy

ARTICLE INFO

Article history:
Received 16 September 2011
Received in revised form 7 November 2011
Accepted 8 November 2011
Available online 9 December 2011

Keywords:
cardiovascular diseases/genetics
genetic epidemiology study design
genetic techniques
ANRIL

ABSTRACT

Continuous updating of the genotyping technology has led to improvement of genetic study design. The recent advances in technology coupled with the advances in our understanding of the molecular mechanisms have allowed a more comprehensive examination of the role of genetics, environment and their interaction in determining the individual risk of cardiovascular disease (CVD).
Initial candidate gene studies identified a limited number of polymorphisms associated with disease, explaining only a minor part of trait variance. Furthermore, results were not often concordant, with meta-analyses not reaching the statistical power to confirm an association in many cases. The advent of the genome-wide design furnished an enormous quantity of information and decreased time of genotyping, while increased complexity of analyses and costs. Their results were more concordant, even when they suggested associations between CVD and polymorphisms distant from codifying regions or in genes involved in previously unsuspected pathways. Future results from genome-wide studies coupled with results from functional studies and investigation on gene-environment interactions will allow improvement of cardiovascular risk assessment and discovery of new targets for therapy and prevention.
In this review, a brief history of cardiovascular genetics is reported, from candidate gene to genome wide association studies, that led to the identification of association between CVD and SNPs in the 9p21 region, firstly thought a gene desert without importance.

Contents

Introduction

Ischemic cardiovascular disease and its underlying pathological processes (atherosclerosis and thrombosis) continue to represent a worldwide major cause of mortality and morbidity. The recent advances in our understanding of the molecular and genetic mechanisms coupled with the advances in technology have allowed a more detailed and comprehensive examination of the role of environment and genetics in health and disease and the complex way in which "nature" and "nurture" can interact in determining the individual risk of disease. It is conceivable that the full phenotypic expression of these risk factors and their link to development, severity and type and time of onset of the

Abbreviations: CAD, coronary artery disease; FDR, false discovery rate; GWAS, genome-wide association studies; MI, myocardial infarction; NCBI, National Center for Biotechnology Information; SNP, single nucleotide polymorphism.

☆ XI ETRO Advanced Teaching Course 'Thrombosis: A multidisciplinary approach', Campobasso, Italy, September 2011, 18–23.

* Corresponding author at: Research Laboratories, Fondazione di Ricerca e Cura "Giovanni Paolo II", Università Cattolica del Sacro Cuore, Largo Gemelli 1, 86100 Campobasso, Italy. Tel.: +39 874 312274; fax: +39 874 312710.
E-mail address: licia.iacoviello@rm.unicatt.it (L. Iacoviello).

0049-3848/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.thromres.2011.11.014

clinical manifestations of vascular disease is the result of the combined influence of multiple environmental factors with multiple genes.

The INTERHEART study [1], a case–control study of acute myocardial infarction (MI) in 52 countries throughout the world, found that nine potentially modifiable factors (smoking, hypertension, diabetes, abdominal obesity, lipids and psychosocial factors, conferring higher risk, and moderate alcohol consumption, exercise and consumption of fruits and vegetables, resulting to be protective) accounted for 90% of the risk of MI in men and 94% in women, at population level. Although the population attributable risk for modifiable agents appeared to be almost total, this fact does not mean that genetics would only impact by less than 10% on risk determination. In fact, there is large evidence for a genetic component for coronary artery disease (CAD), with heritability of 57% and 38% for fatal events, for men and women, respectively [2]. The overlapping percentages of risk attributable to these two risk factor categories is due to the fact that the modifiable risk factors studied in the INTERHEART have a certain level of heritability and, at the same time, their effect appear to be modulated by different genetic background. Therefore, great attention should be given to identify genetic polymorphisms and their interaction with environmental factors conferring different susceptibility to cardiovascular disease, aiming at both recognizing subjects at higher risk to which target the treatment and improving the knowledge on gene role.

A Brief History of Cardiovascular Genetics

In these last 20 years, different approaches have been used to associate the effect of a genetic polymorphism to a phenotype, evolving with improvement of genotyping analyses (see Table 1). Indeed, analyses on a single candidate gene were overwhelmed by genome-wide designs, that furnish an enormous quantity of information and decreased time of genotyping, while increased complexity of analyses and costs.

The candidate gene approach

Candidate gene approach was the first and most simple method used, based on the *a priori* hypothesis of the involvement of a gene in pathways playing a role in the determination of an intermediate phenotype, such a molecular and cellular function, or in a clinically evident effect. The first studies focused the attention on single polymorphisms, in single genes, thought having a major role, mainly located in exons, determining functional changing in the transduced protein, or in close regions regulating the gene expression.

The first candidate gene study on cardiovascular disease was published in Nature in 1992 [3]. This study showed that homozygotes for a deletion in the region of the angiotensin converting enzyme gene, previously associated with higher levels of circulating enzyme, were at higher risk of MI. This association was more evident in subjects without some well-established non-genetic risk factors, suggesting the importance of studying genetic factors independently and in addition to environmental factors. Many studies were published later on, focused on a single genetic polymorphism. It is noteworthy that in a gene of 30 Kb there are approximately 10–100 polymorphisms, all of them potentially responsible for variation of the gene effect. Then, some studies started to genotype more than one polymorphism in the same gene. The first study on cardiovascular disease focusing on more gene variants was published in 1998 [4] in the New England Journal of Medicine, focused on two polymorphisms in distant regions of the gene encoding for coagulation factor VII. An association between MI and both studied genetic variants was found, suggesting also a higher effect for the presence of both risk alleles.

A further step was the identification of plausible mechanisms underlying the associations between gene polymorphisms and diseases, also in order to strengthen the reliability of the observed results, which could be affected by chance. In the candidate gene approach, differently from other genetic designs, a linked intermediate phenotype which could confirm the hypothesis was easily to find, since the selection of gene was based on *a priori* hypothesis of involvement of its function in a certain phenotype. In the study mentioned above [4], an association between the risk alleles and higher levels of factor VII clotting activity was found in the same sample, showing an association between factor VII levels and MI risk. These results suggested the protein level as a link between genotype and disease. Besides levels of encoded protein in basal conditions, the link between genetic variants and phenotypic variability could be represented by different response to an environmental stimulus. An environmental risk factor, before damaging its final target, interacts with a number of molecular pathways, which are in turn regulated by genetic factors. As a consequence, the effect of an environmental risk factor could be modified by the different genetic background. In another study from our group on the association between IL-1β polymorphisms and MI and stroke risk at younger age, we observed that subjects with different genotypes showed a different level of cytokine production by blood monocytes in response to LPS, corresponding to a different risk of MI and stroke [5]. This is a classical example of how genetic could impact in disease pathogenesis.

The gene-wide approach

A limit of the candidate gene approach with few tested polymorphisms is that, among all polymorphisms present in a gene region, there could be many possible variants exerting the observed effect, even located thousands of base pairs away from the gene (regulating gene expression) or within the introns (regulating the alternative splicing of the transcript). The pattern of linkage disequilibrium (the way in which single nucleotide polymorphism -SNP- alleles in the same gene are combined) among multiple polymorphic sites varies considerably within the same gene and between genes [6] and several SNPs within a gene are in strong linkage disequilibrium. Such complex genetic structure requires the identification of all possible polymorphisms within genes, the prediction of a haplotypes structure by computer simulation to assure the detection of all possible variants with functional effects and a haplotype analysis [7]. In this way it is possible to calculate, in each gene, the minimal subset of SNPs (tag-SNPs) required to extract the maximum amount of information regarding the locus, computational approaches based on allele frequency, linkage disequilibrium, and position on the gene. This was the approach of the SeattleSNPs project, focusing on identifying, genotyping, and modeling the associations of SNPs in candidate genes and pathways that underlie inflammatory responses in humans [8].

Using the SeattleSNPs database approach, we could show that not simply a single SNP but different haplotypes of tissue factor or IL-1β genes were associated with the risk of MI or stroke [9,10]. Haplotypes, indeed, determine the genetic variance of a locus more accurately than individual SNPs, accounting for the synergic contribution of all SNPs alleles along the same DNA strand. Haplotype-pairs analysis, moreover, provide finest information on both chromosomes.

The International Haplotype Map Project [11] moved the attention from single gene to the entire genome, developing a validated marker set using validated SNPs derived from dbSNP, the NCBI archive for genetic variation collecting genotype data from different sources. The ultimate goal of this project was to genotype up to 5,000,000 SNPs (~1 SNP/600 bp), in different populations from Europe, Asia, and the Americas. These data were then used to develop a linkage disequilibrium based map, divided into blocks, which define sets of correlated SNPs, to produce a common resource for genetic association studies.

The genome-wide approach

The SNPs associated with disease identified through candidate gene studies explained only a minor part of trait variance, suggesting