Rev Esp Cardiol. 2011;64(6):509–514

Review article

Genetic Variants, Cardiovascular Risk and Genome-Wide Association Studies

Osmel Companioni,[a] Francisco Rodríguez Esparragón,[a] Alfonso Medina Fernández-Aceituno,[a,b] and José Carlos Rodríguez Pérez[a,c,*]

[a] Unidad de Investigación, Hospital Universitario de Gran Canaria Dr. Negrín, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[b] Servicio de Cardiología, Hospital Universitario de Gran Canaria Dr. Negrín, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[c] Servicio de Nefrología, Hospital Universitario de Gran Canaria Dr. Negrín, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

Article history:
Received 3 August 2010
Accepted 27 January 2011
Available online 6 May 2011

Keywords:
Cardiovascular risk
Coronary artery disease
Myocardial infarction
9p21

ABSTRACT

Genome-wide association studies have shown an association between single nucleotide polymorphisms (SNPs) and coronary artery disease and myocardial infarction in new chromosomal regions: 1p13.1, 2q36.3, 9p21 and 10q11.21. The SNPs from the 9p21 region constitute a risk haplotype due to the strong linkage disequilibrium in this area. These SNPs have been extensively replicated in several European and Asian populations, and are associated with other pathologies such as abdominal aortic and intracranial aneurysms, and with intermediate phenotypes such as arterial stiffness and coronary calcium. The risk haplotype of 9p21 is located in a region without annotated genes, near *CDKN2A* and *CDKN2B,* known tumor suppressor genes encoding for inhibitors of cell cycle kinases. In the remaining regions the SNPs are located in genes with known roles in atherosclerosis as well as others with new roles. It has been shown that the incorporation of genetic information in the form of SNPs slightly improves the prediction of long-term cardiovascular risk estimated by the Framingham function, allowing the reclassification of individuals into more precise categories. Gene expression studies have found that expression levels of *CDKN2A/CDKN2B/ANRIL* are co-regulated and associated with the risk haplotype and atherosclerosis severity.

Variantes genéticas, riesgo cardiovascular y estudios de asociación de genoma completo

Palabras clave:
Riesgo cardiovascular
Enfermedad arterial coronaria
Infarto de miocardio
9p21

RESUMEN

Estudios de genoma completo han demostrado asociación entre polimorfismos de nucleótido simple (SNP) y enfermedad coronaria e infarto agudo de miocardio en diversas regiones cromosómicas: 1p13.1, 2q36.3, 9p21 y 10q11.21. Los SNP de 9p21 conforman un haplotipo de riesgo; las asociaciones detectadas en esta región han sido replicadas en diversas poblaciones y se los ha encontrado asociados con otras afecciones como aneurisma aórtico abdominal e intracraneal, rigidez arterial y calcio coronario. El haplotipo en 9p21 está localizado en una zona sin anotación génica, cercana a los genes reguladores del ciclo celular *CDKN2A* y *CDKN2B.* En las restantes regiones, los SNP asociados se encuentran en genes con funciones conocidas en la enfermedad aterosclerótica. Se ha demostrado que la incorporación de información genética de los SNP de riesgo de 9p21 mejora la predicción del riesgo cardiovascular a largo plazo estimado por medio del *score* de Framingham y permite la reclasificación de individuos en categorías más precisas. Se han realizado estudios de expresión de *CDKN2A, CDKN2B* y *ANRIL* que han demostrado que están corregulados y se asocian con SNP de 9p21, así como con la severidad aterosclerótica, lo que apunta a la relevancia de esta región en la enfermedad coronaria.

INTRODUCTION

Cardiovascular disease is the leading cause of mortality in industrialized countries. Since 1996, coronary artery disease (CAD) has been responsible for approximately 31% of cardiovascular mortality in Spain, the most common cause being myocardial infarction (MI; 61%). In 2005, there were 72 950 cases of MI in Spain; 60% of these patients were admitted to hospital and 40% died before arrival. The health care costs of coronary disease were 1953 million Euros in 2003.[1]

Epidemiological and animal studies have determined the cardiovascular risk factors (CRF) predisposing to CAD, such as low-density lipoprotein cholesterol (LDLc), age, obesity, smoking, low levels of high-density lipoprotein cholesterol (HDLc), triglycerides, diabetes mellitus, and hypertension.[2] Furthermore, a genetic component has been confirmed by studies on monozygotic twins and other studies in which a family history of CAD has been associated with coronary events.

* Corresponding author: Unidad de Investigación, Servicio de Nefrología, Hospital Universitario de Gran Canaria Dr. Negrín, Universidad de Las Palmas de Gran Canaria, 35010 Las Palmas de Gran Canaria, Spain.
E-mail address: jrodperd@gobiernodecanarias.org (J.C. Rodríguez Pérez).

1885-5857/$ – see front matter
doi:10.1016/j.rec.2011.01.011

Abbreviations

CAD: coronary artery disease
CNV: copy number variation
CRF: cardiovascular risk factor
GWAS: genome-wide association studies
MI: myocardial infarction
RR: risk ratio
SiRNA: small interference RNA
SNPs: single nucleotide polymorphism

From the genetic standpoint, CAD is classified as a complex disease, although there are forms of presentation with a simple Mendelian inheritance pattern, such as familial hypercholesterolemia caused by mutations in the LDL receptor gene, *PCSK9* and *ApoB*.[2] The main difference lies in the fact that Mendelian diseases cause mutations in a gene that lead to a dysfunctional change in the coded protein, thus increasing the risk of the disease, whereas complex diseases are caused by multiple gene polymorphisms with a small effect size interacting with environmental risk factors.

Linkage and association studies have been used to map the causal genes of CAD. The former follow a family study design in which microsatellites distributed throughout the genome are genotyped (Table 1) leading to the identification of genes such as *MEF2A*,[3] *ALOX5AP*,[4] and *TNFSF4*.[5] Association studies usually follow a case-control design and investigate candidate genes assumed to be involved in the physiology of the disease, and are thus based on a previous hypothesis. By comparing the frequency of genetic polymorphisms in the cases and controls it can be determined whether the gene is associated with the disease or not. The type of polymorphism most commonly used in these studies are single-nucleotide polymorphisms (SNPs) consisting of a variation in a single base pair in the DNA sequence. The most important limitations of association studies include the following: the lack of replicability in different populations, the small sample size and consequent low statistical power, genotyping errors, inaccurate clinical characterization of the disease, inadequate case and control selection, and the presence of a population substructure.[6]

Metaanalysis of case-control studies has detected the association of some polymorphisms with MI (*MTHFR*-C677T, *CETP*-TaqIB, *PON1*-Q192R, *eNOS* Glu298Asp, *Prothrombin* G20210A, *F5, AT1R* 1166 A/C, *ApoB* [Xba, EcoRI, Ins/Del], *ApoE* ε4/ε4, *ACE* DD, *LPL* Ser447Ter), but many of these associations are false positives.[7]

The aim of the present article is to review new advances regarding the genetic component of CAD, based on genome-wide association studies (GWAS) and their potential clinical usefulness.

GENOME-WIDE ASSOCIATION STUDIES

Recently, GWAS have been successful in the discovery of *loci* associated with CAD, MI, diabetes mellitus type 2, rheumatoid arthritis, Crohn disease, bipolar disorder, and others detailed in a catalog of GWAS.[8] These studies are based on the genetic analysis of large case-control samples through genotyping thousands of SNPs distributed throughout the whole genome using DNA microchips (these studies are not based on previous hypotheses). The results of the HapMap project on genotype frequency and haplotype structure have enabled the selection of the minimum SNPs necessary for genotyping in GWAS[9] to capture most of the common genetic variability in the human genome.

The advantages of GWAS compared to case-control genetic studies derive from the use of large samples, automated SNPs genotyping of the whole genome (and thus not restricted to candidate genes), sex testing, proof of family relationships and ancestry, and sample quality control to exclude duplication. Furthermore, quality control excludes from the analysis SNPs with missing data, those with deviations from Hardy-Weinberg equilibrium in the controls, any with an allele frequency $< 1\%$, and those in which the genotyping rate is $<80\%$-90%. The SNPs are chosen using a criterion for statistical significance that keeps the false positive rate within acceptable limits; also, the results have to be replicable in other populations for an association to be considered definitive.[9,10]

The limitations of these studies include the difficulty in detecting *loci* with a small effect size, the preferential choice of SNPs (and thus other polymorphisms such as copy-number variations (CNV) and microsatellites are not analyzed), and the fact that the contribution of less common SNPs is not assessed and that these studies have usually been conducted in European individuals, which means that other population groups have been little studied.

Genome-Wide Association Studies for Coronary Artery Disease and Myocardial Infarction (Region 9p21)

Recent GWAS have found a new region without annotated genes (9p21) associated with CAD[9,11,12] and MI[13] independently of its association with CRF. Region 9p21 contains various SNPs in linkage disequilibrium, such as rs1333049, rs10757274, rs10757278, rs2383206, and rs2383207; of these rs1333049 has shown the greatest evidence of an association (odds ratio [OR] = 1.24; 95% confidence interval [CI], 1.20-1.29). A metaanalysis of some 40 000 subjects showed that 25% of European individuals have 2 copies of the rs1333049 risk allele, which increases the risk of CAD by 1.6.[14] In addition, these SNPs have been associated with abdominal and intracranial aortic aneurysm, arterial rigidity, myocardial damage caused by coronary spasm, severe coronary stenosis, and coronary artery calcification.[6]

Gene Expression Studies at Chromosome Region 9p21

Of all the loci associated with CAD and MI, region 9p21 has been the most replicated and has been shown to have the strongest association, motivating the attempt to determine the molecular mechanism underlying this relationship. The risk haplotype for cardiovascular disease on 9p21 is located in an area without

Table 1
Glossary of Genetic Terms Used in the Article

Term	Concept
Genotyping	Determining the genotype of an individual for a genetic variation, whether a polymorphism or a mutation
Transcript	Each of the mRNA variants resulting from the alternative splicing characteristic of human genes
Hardy-Weinberg equilibrium	Mathematical model establishing that the genetic composition of a population remains constant over generations in the absence of mutation, natural selection, and other evolutionary forces
Minor allele	Given 2 alleles A (0.7) and B (0.3) of a polymorphism, the minor allele is the one with the lowest relative frequency (B)