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,d

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

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 and 64(6):509–514

Rev Esp Cardiol. 2011;64(6):509–514

© 2011 Sociedad Española de Cardiología. Published by Elsevier España, S.L. All rights reserved.

Keywords:
Cardiovascular risk
Coronary artery disease
Myocardial infarction
9p21

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.

© 2011 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L. Todos los derechos reservados.

ABSTRACT

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

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 genómica, 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.

© 2011 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L. Todos los derechos reservados.

© 2011 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L. Todos los derechos reservados.

INTRODUCTION

© 2011 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L. Todos los derechos reservados.
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. 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, ALOX5AP, and TNFSF4. 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.

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, ATIR 1166 A/C, ApoB [Xba, EcoRI, Ins/Del], ApoE e4/e4, ACE DD, LPL Ser447Ter), but many of these associations are false positives.

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. 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 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.

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 and MI 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. 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.

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...
annotated genes, close to cluster INK4a/ARF composed of the tumor suppressor genes CDKN2A and CDKN2B which are implicated in causing various cancers. Adjacent to these, an antisense noncoding RNA gene known as ANRIL has been identified. Various groups have evaluated the expression of CDKN2A/2B/ANRIL and its association with risk SNPs on 9p21.

Broadbent et al. using QRT-PCR, detected the expression of the ANRIL transcript (D4Q85453) in primary coronary smooth muscle, macrophages and in carotid endarterectomy and abdominal aortic aneurysm tissue samples. Jaronina et al. demonstrated that ANRIL has 4 evolutionarily conserved sequences, one of which (CNS3) potentiates gene expression when amplified from the risk homozygote compared to the reference homozygote. Fokkensen et al. identified 8 novel ANRIL transcripts in lymphoblastoid cells and carotid artery, medial aorta, and mammary artery plaque tissue. Holdt et al. analyzed gene expression in peripheral blood mononuclear cells of CAD patients and carotid, aortic, and femoral plaques. These studies confirm that the only gene that manifests differential expression is ANRIL in association with risk SNPs on 9p21 and with atherosclerotic severity.

In addition, Mathews et al. demonstrated in coronary endarterectomy plaque samples that smooth muscle cells undergo senescence associated with DKN2A (p16) and p21 protein overexpression as measured by Western blotting. Furthermore, knockout mice for several tumor suppressor genes, such as p19ARF, p53, p27Kip1, and pRb, present aggravated atherosclerosis, indicating the role of cell cycle control proteins in this disease.

In summary, gene expression studies in 9p21 show that the genes CDKN2A, CDKN2B, and ANRIL are expressed in atherosclerotic tissue and are transcriptionally correlated. The ANRIL gene has several transcripts whose expression levels demonstrate a stronger correlation with risk SNPs genotypes in 9p21 than with CDKN2A/2B. In addition, GWAS have identified ANRIL as a susceptibility gene for diabetes type 2, glioma, and basal cell carcinoma. C/EBP transcription factor binding site, altering the hepatic expression of SORT1. By using gene overexpression and siRNA knockdown for SORT1 in mouse liver, these authors demonstrated that this gene alters LDLc levels. The importance of this novel pathway is based on a 40% difference in MI risk between alternative p13 homoygotes, an effect comparable to the common variants of LDLR and PCSK9. The p13 minor allele frequency is around 30% in Europeans and is present in other populations, and thus this locus is considered a global genetic determinant of MI. In another study, the risk allele A of rs599839 was associated with increased susceptibility for developing CAD (1.29; 95% CI, 1.18-1.40). The intronic polymorphism rs6922269 is located in gene MTHFD1L. This gene codes a mitochondrial enzyme responsible for synthesizing forms of tetrahydrofolate. The possibility of a functional connection between MTHFD1L and CAD is supported by the fact that its activity influences plasma homocysteine levels.

The SNP rs2943634 is located in the region 2q36.3 where only a pseudogene (ENSG00000197218) has been annotated. The SNP rs501120 is located upstream of CXXC12, a chemokine with a core role in tissue regeneration in ischemic heart disease and angiogenesis via endothelial progenitor cell recruitment. The SNP rs2943634 has been associated with hypertension and low HDLC levels in the MORGAN cohort.

The SNP rs71228212 is located in gene SMAD3, a transcriptional modulator activated by TGFβ which participates in cell growth and inhibition, both of which are fundamental in the progression of atherosclerotic plaque. This SNP has been associated with cholesterol other than HDLC in the MORGAN study, and it is assumed that its effect on CAD is mediated by its association with this CRF. The SNP rs9818870 is found in gene MREAS, which belongs to the RAS superfamily of GTP-binding proteins and is highly expressed in the heart. It is thought to participate in adhesion molecule signaling activity, an important process in the initial phase of atherosclerotic disease. rs9982601 is a variant located in SLC5A3, which participates in Na+ and myo-inositol transport as a response to hypertonic stress.

The SNP rs12526453 is located in gene PHACTR1. This gene encodes for an inhibitor of enzyme activity of protein phosphatase 1, responsible for the dephosphorylation of serine and threonine residues in proteins.

Polymorphism rs17465637 is located in gene MIA3, which is involved in cell growth and inhibition. Finally, SNP rs3184504 is located in gene SH2B3, which encodes for an adapting protein of the intracellular signaling pathway of T lymphocyte activation. Table 2 shows a list of the SNPs mentioned and others, extracted from the GWAS catalog by searching for the terms “coronary heart disease” and “myocardial infarction.” It is of note that after identifying the target loci other trials are required (gene and protein expression, knock-out/down, biological activity) to explain the associations and identify therapeutic targets of potential clinical interest.

**GENETIC PREDICTION OF CARDIOVASCULAR RISK**

Cardiovascular risk tables are mathematical models based on prospective cohort studies that model the risk of suffering CAD over time given certain risk factors. In the mid-20th century, the Framingham study developed the first risk table and since then others have been created, such as PROCAM, QRISK, ASSIGN, etc. However, the measurement of CRFs related to lipid values alone has proven to be an inefficient marker for the prediction of cardiovascular risk, as these are affected by many variables. On the other hand, genetic polymorphisms remain constant throughout...
life, underlining the immense value of genetic screening. Several groups have included SNPs in cardiovascular risk tables to create models with greater predictive power compared to the use of CRFs alone, allowing the reclassification of subjects into more accurate risk categories. Reclassification is based on a statistical analysis to assess whether the incorporation of a given genetic risk score (composed of several SNPs) moves subjects toward high risk categories more often than toward low risk categories, and controls toward low risk categories more frequently than toward high risk ones. Humphries et al. investigated coronary risk prediction by analyzing subjects from the Northwick Park Heart Study II over 10.8 years. The value of area under the ROC curve for a model composed of the CRFs age, triglycerides, total cholesterol, smoking, hypertension, and diabetes was 0.79 for the combined genetic plus CRF model compared to 0.76 for the CRF model alone.

### Table 2

<table>
<thead>
<tr>
<th>SNP (dbSNP)</th>
<th>Gene/chromosome</th>
<th>Frequency of the risk allele in controls</th>
<th>Odds ratio</th>
<th>Trend, P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17672123 C/T</td>
<td>FMAN2</td>
<td>Allele T 86%</td>
<td>1.32 (0.79-2.22)</td>
<td>$1.04 \times 10^{-4}$</td>
</tr>
<tr>
<td>rs383830 A/T</td>
<td>5q21</td>
<td>Allele T 78%</td>
<td>1.92 (1.40-2.63)</td>
<td>$5.72 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs6922269 A/G</td>
<td>MTHHD1L</td>
<td>Allele A 25%</td>
<td>1.65 (1.32-2.06)</td>
<td>$6.33 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs8055236 G/T</td>
<td>19q12</td>
<td>Allele G 80%</td>
<td>2.23 (1.56-3.17)</td>
<td>$9.73 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs888034 C/T</td>
<td>SEZ6L</td>
<td>Allele T 31%</td>
<td>1.62 (1.34-1.95)</td>
<td>$6.90 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs1333049 C/G</td>
<td>9p21</td>
<td>Allele C 47%</td>
<td>1.90 (1.61-2.24)</td>
<td>$1.79 \times 10^{-14}$</td>
</tr>
<tr>
<td>rs7250581</td>
<td>19q12</td>
<td>Allele G 78%</td>
<td>1.40 (1.05-1.86)</td>
<td>$9.12 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

For each SNP, the frequency of the risk allele in controls is provided, along with the odds ratio and its 95% confidence interval. The trend, P value reflects the statistical significance of the odds ratio. The dbSNP, single nucleotide polymorphism database; SNP, single nucleotide polymorphism.
and systolic blood pressure was 0.66 (0.61–0.70). However, when the model was composed of SNPs in genes UCP2, ApoE, LPL, and ApoA4, the ROC value was 0.62 (0.58–0.66). The model including genotypic and environmental variables (area under the ROC curve, 0.72 [0.68–0.76]) showed a significant increase in the predictive power of the Framingham equation compared to isolated models.

In a follow-up study of 10,000 patients from the ARIC (Atherosclerosis Risk in Communities) cohort conducted over 14.6 years, Brahmbhatt et al.35 analyzed the prediction of MI, coronary revascularization and cardiac death, and genotyped one SNP on rs921 (rs10757274). The authors reclassified 1.3% and 0.8% of the subjects compared to the Framingham equations; the rs921 allele had the greatest influence on the intermediate risk categories. Morrison et al.36 conducted a follow-up study of the ARIC cohort for the onset of cardiovascular events over an average of 13 years and genotyped 116 SNPs; they created a score that was significantly associated with CAD in black individuals (risk ratio [RR] = 1.2; 95% CI, 1.11–1.29) and European individuals (RR = 1.1; 95% CI, 1.06–1.14). The area under the ROC curve calculated with the risk score and CRFs was significantly larger than that which only took into account these factors in black individuals; however, the result was not statistically significant in Europeans.

Paynter et al.37 conducted a 10.2-year follow-up of 22,129 European women who were health professionals and part of the Women’s Genome Health Study. They genotyped the SNP rs10757274, which was associated with the presence of cardiovascular events (RR = 1.25; 95% CI, 1.04–1.51). However, adding this SNP to a predictive model using CRFs (C-reactive protein and a family history of MI) had no effect on the discriminative power of the model.

McGeachie et al.38 used the prospective study MESA based on subclinical atherosclerosis markers. The authors created a predictive model of coronary calcification with 13 SNPs and an ancestry informative marker. The clinical variables were sex, age, weight, smoking, and diabetes mellitus. When the model (SNP plus clinical variables) was compared to other models that included SNPs alone (77%) or clinical variables (78.3%), the area under the ROC curve increased to 85%. In Spain, 2 private companies (Ferrer inCode and GenDiag) have created a test for cardiovascular risk called Cardioincode (http://www.ferrerincode.com) that improves predictions based on CRFs and 11 associated SNPs according to GWAS.

Although the statistical differences obtained in these studies are not sufficient to warrant immediate clinical interest, and the prediction models can only be used at the population level rather than the individual level, they are still relevant given that the associated SNPs, despite representing moderate risk, are frequently found in the general population and thus provide valuable epidemiological information. Since the GWAS are an active area of research, in the future it may be possible to create better predictive models that could include a greater number of SNPs associated with CAD, MI, and CRF, as well as other types of polymorphisms such as the CNVs recently investigated in GWAS.

CONFLICTS OF INTEREST

None declared.

REFERENCES