■ **ORIGINAL ARTICLE**

# Additive Effects of Multiple Genetic Variants on the Risk of Coronary Artery Disease

Carla Lluís-Ganella,[a] Gavin Lucas,[a] Isaac Subirana,[a,b] Mariano Sentí,[a,c] Jordi Jimenez-Conde,[d] Jaume Marrugat,[a] Marta Tomás,[a] and Roberto Elosua[a,b]

[a]Grupo de Epidemiología y Genética Cardiovascular (EGEC-ULEC), Institut Municipal d'Investigació Mèdica (IMIM-Hospital del Mar), Barcelona, Spain
[b]CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
[c]Universitat Pompeu Fabra, Barcelona, Spain
[d]Servicio de Neurología, Hospital del Mar, Barcelona, Spain

**Introduction and objectives.** Coronary artery disease (CAD) has a substantial genetic component and, in recent years, a number of genetic variants associated with the disease have been identified. The objective of this study was to evaluate the magnitude of the association between a genetic risk score, which is based on the accumulated number of risk alleles in all genetic variants of interest, and the presence of CAD.

**Methods.** The study involved in silico data from the Wellcome Trust Case-Control Consortium on 1988 patients with CAD and 5380 controls. The association between the genetic risk score and CAD was assessed using logistic regression analysis.

**Results.** Nine genetic variants independently associated with CAD irrespective of other cardiovascular risk factors were selected. There was a linear association between the number of risk alleles and the risk of presenting with CAD (odds ratio [OR] for an increase of one allele=1.18; 95% confidence interval [CI], 1.15-1.22; $P=2\times10^{-16}$). The OR for CAD for the last quintile of the accumulated number of risk alleles relative to the first was 2.21 (95%CI, 1.87-2.61; $P=5\times10^{-21}$).

**Conclusions.** A genetic risk score based on nine genetic variants independently associated with CAD irrespective of other cardiovascular risk factors was associated with the presence of the disease. Cohort studies are needed to determine whether this genetic risk score can improve the predictive capacity or the risk classification of classical risk functions.

**Key words:** *Coronary artery disease. Genetics. Polymorphisms. Genetic risk. Genetic variants.*

## Efecto aditivo de diferentes variantes genéticas en el riesgo de cardiopatía isquémica

**Introducción y objetivos.** La cardiopatía isquémica (CI) tiene un componente genético significativo y en los últimos años se han identificado variantes genéticas asociadas con ella. El objetivo de este estudio fue evaluar la magnitud de la asociación entre una puntuación individual de riesgo genético, basada en el número de alelos de riesgo acumulados en un conjunto de variantes genéticas de interés, y la presencia de CI.

**Métodos.** Estudio de los datos *in silico* del consorcio de casos y controles del Wellcome Trust. Se incluyó a 1.988 pacientes con CI y 5.380 controles. Se analizó la asociación entre una puntuación de riesgo genética y CI mediante regresión logística.

**Resultados.** Se seleccionaron nueve variantes genéticas asociadas con CI de forma independiente de los factores de riesgo. Existe una asociación lineal entre el número de alelos de riesgo y el riesgo de sufrir CI: *odds ratio* (OR) por incremento de alelo = 1,18 (intervalo de confianza [IC] del 95%, 1,15-1,22; p = 2 × 10⁻¹⁶). La OR de CI del último quintil del número de alelos de riesgo acumulados respecto al primero fue 2,21 (IC del 95%, 1,87-2,61; p = 5 × 10⁻²¹).

**Conclusiones.** Las nueve variantes genéticas asociadas con CI independientemente de los factores de riesgo cardiovascular permiten crear una puntuación de riesgo genético de CI que se asocia con la presencia de la enfermedad. Es necesario realizar estudios de cohorte que analicen si esta puntuación genética puede mejorar la capacidad predictiva o la reclasificación de las funciones de riesgo clásicas.

---

## ABBREVIATIONS

CAD: coronary artery disease
CI: confidence interval
CVRF: classical cardiovascular risk factor
GWA: genome-wide association
OR: odds ratio
SD: standard deviation
WTCCC: Wellcome Trust Case Control
    Consortium

---

## INTRODUCTION

Coronary artery disease (CAD) is a pathology with complex aetiology that involves genetic and environmental factors and the interaction between the two. It is estimated that genetic factors could explain between 40% and 55% of the existing variability among the population (inheritability) in the development of CAD.[1] Genome Wide Association (GWA) studies have been carried out over the last 3 years and have consistently identified new genetic variants associated with different complex diseases, including CAD.[2-8] Whereas some of these genetic variants are also associated with a classic cardiovascular risk factor (CVRF), others are independent. The discovery of these genetic variants independently associated with CAD could make it possible to identify new aetiopathogenic mechanisms of CAD and new therapeutic targets. Furthermore, these variants could be used as new biomarkers to help improve coronary risk prediction, or to reclassify intermediate-risk individuals by adding new data to the classic risk equations.[9] The improvement of risk equations by including new biomarkers is clinically significant since existing risk equation sensitivity is low and a large number of cases affect intermediate-risk patients.[10]

Recently the American Heart Association (AHA) published recommendations which define the different stages in the evaluation of new risk biomarkers and their subsequent application in clinical practice.[11] In the first stage, known as "Proof of Concept," it is necessary to establish whether the level of exposure to the biomarker is different in individuals with and without the disease. In subsequent stages, cohort studies should be carried out to analyze any improvement in ability to discriminate, determine the effect on daily clinical practice and evaluate cost effectiveness.

The aims of this document are: *a)* to select genetic variants that are associated with CAD and independent of CVRFs; and *b)* to apply the first stage of a new biomarker evaluation process proposed by the AHA, which establishes the magnitude of the association between an individual genetic risk score, based on the number of accumulated risk alleles that each individual has within a set of selected genetic variants, and the presence of CAD.

## METHOD

### Study Design

A case-control study with in silico data from the Wellcome Trust Case Control Consortium (WTCCC).[8]

### Study Sample

The cases (n=1988) were selected from the first phase of the WTCCC study.[8] As described in the original study, the CAD cases presented a history of myocardial infarction or coronary revascularization (including coronary bypass surgery or coronary angioplasty) before the age of 66. The control subjects were selected from 2 studies included in the WTCCC project: *a)* 2706 control subjects from a British cohort of people born in 1958 (1958 Birth Cohort Controls [58C])[12]; and *b)* 2674 control subjects who were blood donors (UK Blood Services Controls [NBS])[8]. Although data from 3004 control subjects were presented in the original study,[8] recently the individual level genotype data for 5380 control subjects were made available to the scientific community[13] and included in this analysis.

### Selection of Genetic Variants

The genetic variants were selected from the database of GWA studies (Catalog of GWA Studies) of the National Human Genome Research Institute (NHGRI).[14] The selection criteria were that the genetic variants needed to be associated with CAD and be independent of CVRFs according to the data from this catalogue.[14] To identify the variants associated with CAD, the following clinical terms were included in the search of the NHGRI Catalog of GWA Studies: Coronary Artery Disease, Coronary Disease, Myocardial Infarction, Early Onset Myocardial Infarction. Variants with a *P* value $>1\times10^{-6}$ in the variant discovery study were not included. Of the genetic variants associated with one another because of linkage disequilibrium

($r^2$>0.9, HapMap CEU [residents of Utah with ancestors in Northern and Eastern Europe]), only one was selected. In order to exclude variants associated with CVRFs, any possible association between CVRFs and the genetic variant of interest or the chromosomal region where the variant could be found was explored in the NHGRI Catalog of GWA Studies; moreover, variants located within genes associated with CVRFs were excluded, taking into consideration the findings of previous studies (for example: the low density lipoprotein receptor gene [LDL], *LDLR*).

Quality control methods were applied to the results for all the genetic variants selected, verifying Hardy-Weinberg equilibrium in control subjects and that the frequency of the minority allele was greater than 1%. Quality control methods were also applied to individuals, excluding those participants who had no data for at least one of the genetic variants, as no score could be obtained for all possible risk alleles (19 individuals excluded).

A risk allele was defined for each of the variants, in other words, the nucleotide whose presence confers a higher risk of suffering a coronary event, according to the data provided from the variant discovery study.

## Estimation of the Individual Genetic Risk Score

To calculate the genetic risk score, the number of accumulated risk alleles in each individual within the set of selected genetic variants was counted. Each individual might have 0, 1, or 2 risk alleles in each of the variants. By calculating the sum of the accumulated risk alleles within the set of selected variants (n=9), each individual is given a score from 0 to 18. It was assumed that the effect of the different alleles analyzed was similar and therefore they were all weighted the same.

The individual genetic risk score was also calculated, weighting each of the selected variants by the magnitude of their individual effect, which was extracted from the data presented in the variant discovery studies.

## Statistical Analysis

In an initial analysis the selected genetic variants were analyzed individually and the association between each variant and the presence of CAD was established using an additive genetic model estimating the OR of the CAD per increase in risk allele frequency.

In order to consider the additive effect of the set of the different genetic variants selected, first the distribution of the number of accumulated risk alleles in cases and controls was analyzed and the difference in the mean genetic risk score between cases and controls was compared using the Student *t*-test.

Subsequently, logistic regression was used to estimate the association between the genetic risk score and the risk of CAD, and different types of analyses were carried out:

*1.* Firstly, the genetic risk variable was considered a categorical variable, taking the group of carriers of 7 risk alleles as the reference group, because this is the closest value to the median in controls and this group contains the highest number of individuals, which makes it possible to obtain more accurate risk estimators. The OR was estimated for each genetic risk score category (group of carriers of 8 alleles vs carriers of 7 alleles; carriers of 9 alleles vs 7; carriers of 6 alleles vs 7; and so on).

*2.* Secondly, the OR per increase in the frequency of risk alleles was calculated, treating this variable as continuous, assuming that risk is constant per increase in the frequency of alleles in the range of values observed; an analysis was carried out to determine whether the association between the number of alleles and the risk of CAD was linear.

*3.* Thirdly, the analysis described in paragraph *1* of this section was repeated, defining groups based on the genetic risk score quintiles obtained from the control group and using the first quintile as reference.

Any interaction between the genetic variants of interest was also analyzed to establish whether the combined effect of these was greater than the additive effect.

In all the analyses carried out, a *P*<.05 was considered statistically significant. The analyses were carried out using the R program.[15]

## RESULTS

A total of 21 unique genetic variants was identified (Table 1) in the NHGRI Catalog of GWA Studies,[14] and in the selected phenotypes. Figure 1 shows a flow diagram of the selection process of the genetic variants of interest in this study. Nine genetic variants associated with CAD independently of the CVRFs were selected using this procedure.

Table 2 shows the results of the association of each of the 9 variants selected with CAD in the larger WTCCC sample analyzed in this study. In this analysis, most of the variants were associated with CAD, although there was an association in the

**TABLE 1. Description of the Genetic Variants Associated with Coronary Artery Disease Identified in the Catalogue of Published Genome-Wide Association Studies and the Reasons for their Exclusion from our Study**

| Ref. | Genetic Variant | Chromosomal Region[a] | Genes Close to the Variant Identified[a] | Alleles | MAF[a] | Minor Allele[b] | Risk Allele[a] | OR (95% CI)[a] | P[a] | Reason for Exclusion |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | rs646776 | 1p13.3 | *CELSR2, PSRC1, SORT1* | T/C | 0.19 | C | T | 1.19 (1.13-1.26) | $8\times10^{-12}$ | 3 |
| 6 | rs599839 | 1p13.3 | *PSRC1* | A/G | 0.23 | G. | A | 1.29 (1.18-1.4) | $4\times10^{-9}$ | 3 |
| 4 | rs11206510 | 1p32.3 | *PCSK9* | C/T | 0.19 | C | T | 1.15 (1.10-1.21) | $1\times10^{-8}$ | 3 |
| 4.6 | rs17465637 | 1q41 | *MIA3* | A/C | 0.29 | A | C | 1.20 (1.12-1.3) | $1\times10^{-6}$ | – |
| 8 | rs17672135 | 1q43 | Intergenic | C/T | 0.13 | C | C | 1.43 (1.23-1.64) | $2\times10^{-6}$ | 1 |
| 4 | rs6725887 | 2q33.1 | *WDR12* | C/T | 0.14 | C | C | 1.17 (1.11-1.23) | $1\times10^{-8}$ | – |
| 6 | rs2943634 | 2q36.3 | *IRS1* | A/C | 0.35 | A | C | 1.21 (1.13-1.3) | $2\times10^{-7}$ | 3 |
| 2 | rs9818870 | 3q22.3 | *MRAS* | C/T | 0.15 | T | T | 1.15 (1.11-1.19) | $7\times10^{-13}$ | – |
| 4 | rs12526453 | 6p24.1 | *PHACTR1* | C/G | 0.35 | G. | C | 1.12 (1.08-1.17) | $1\times10^{-9}$ | – |
| 6 | rs6922269 | 6q25.1 | *MTHFD1L* | A/G | 0.25 | A | A | 1.23 (1.15-1.33) | $3\times10^{-8}$ | – |
| 3 | rs10757278 | 9p21.3 | *CDKN2A/2B* | A/G | 0.45 | A | G. | 1.28 (1.22-1.35) | $1\times10^{-20}$ | 2 |
| 4 | rs4977574 | 9p21.3 | *CDKN2A/2B* | A/G | 0.44 | A | G. | 1.29 (1.25-1.34) | $3\times10^{-44}$ | 2 |
| 6.8 | rs1333049 | 9p21.3 | *CDKN2A/2B* | C/G | 0.47 | C | C | 1.47 (1.27-1.7) | $1\times10^{-13}$ | – |
| 4 | rs1746048 | 10q11.21 | *CXCL12* | C/T | 0.16 | T | C | 1.17 (1.11-1.24) | $7\times10^{-9}$ | – |
| 6 | rs501120 | 10q11.21 | *CXCL12* | T/C | 0.13 | C | T | 1.33 (1.20-1.48) | $9\times10^{-8}$ | 2 |
| 2 | rs2259816 | 12q24.31 | *HNF1A* | T/G | 0.36 | T | T | 1.08 (1.05-1.11) | $5\times10^{-7}$ | 4 |
| 6 | rs17228212 | 15q22.33 | *SMAD3* | C/T | 0.30 | C | C | 1.21 (1.13-1.3) | $2\times10^{-7}$ | – |
| 8 | rs8055236 | 16q23.3 | Intergenic | G/T | 0.20 | T | G. | 1.91 (1.33-2.74) | $6\times10^{-6}$ | 1 |
| 4 | rs1122608 | 19p13.2 | *LDLR* | G/T | 0.25 | T | G. | 1.15 (1.10-1.2) | $2\times10^{-9}$ | 4 |
| 4 | rs9982601 | 21q22.11 | *SLC5A3, MRPS6, KCNE2* | C/T | 0.13 | T | T | 1.20 (1.14-1.27) | $6\times10^{-11}$ | – |
| 8 | rs688034 | 22q12.1 | Intergenic | C/T | 0.31 | T | T | 1.11 (0.98-1.25) | $4\times10^{-6}$ | 1 |

MAF indictaes minor allele frequency; OR (95% CI), odds ratio (95% confidence interval); –, variants not excluded.
[a]Data from the variant discovery study.
[b]Data from the CEU population (residents of Utah with ancestors from Northern and Eastern Europe) from the HapMap Project.
The information about associations has been extracted from the NHGRI Catalogue of GWA Studies,[14] and has been reviewed in articles reporting the association.
Reasons for exclusion:
1. *P* value of the variant discovery study $>1\times10^{-6}$.
2. Another genetic variant already included which captures this variant (value of $r^2>0.9$, according to the data obtained from our CEU sample from the HapMap Project).
3. Genetic variant is associated with cardiovascular risk or linkage disequilibrium (value of $r^2>0$. according to the data obtained from our CEU sample from the HapMap Project) with another genetic variant/other genetic variants associated with cardiovascular risk factors in this catalogue.
4. Genetic variant is located within a gene associated with cardiovascular risk factors (LDL receptor [LDLR]: genetic variations in this gene cause familial hypercholesterolaemia; HNF1A: defects in this gene are associated with diabetes [MODY3]), according to the data obtained from the NCBI [RefSeq].

expected direction for 2 variables, but did not reach statistical significance.

Figure 2A shows the distribution of the number of risk alleles in cases and controls. The distribution follows the same trend for both groups although there is a shift of data toward the right-hand side of the graph indicating the number of alleles in cases (mean risk alleles [typical deviation]: 7.9 (1.8) in cases and 6.8 [1.8] in controls, *P* value = $2\times10^{-16}$). In order to establish whether the increased risk was different in the groups defined by the number of alleles, the magnitude of the risk of CAD was calculated for each group, using the group with 7 risk alleles as the reference group (median in controls). Figure 2B shows the OR and the regression coefficient value (β) for each group in relation to the reference group, the difference in value of the regression coefficients (Δβ) between consecutive categories, and the *P* value of the Fisher's exact test for differences between cases and controls. Figure

2A shows the regression coefficient values obtained (β) on a graph.

The increased risk per increase in the frequency of alleles seems to be constant within the range of values observed. To quantify this increase per frequency of alleles the variable number of risk alleles was considered as a continuous variable and it was estimated that the OR per increase in the frequency of alleles was 1.18 (95% confidence interval,1.15-1.22; $P=2\times10^{-16}$); this linear model explained 92% of the variability of the β values obtained.

When these analyses were carried out, weighting each of the selected variants according to the magnitude of their individual effect, the results were almost identical (OR per increase in the frequency of alleles = 1.18; 95% CI, 1.14-1.21).

No statistically significant interaction between genetic variants was observed, a result that also supports the additive effect of the variants analyzed.
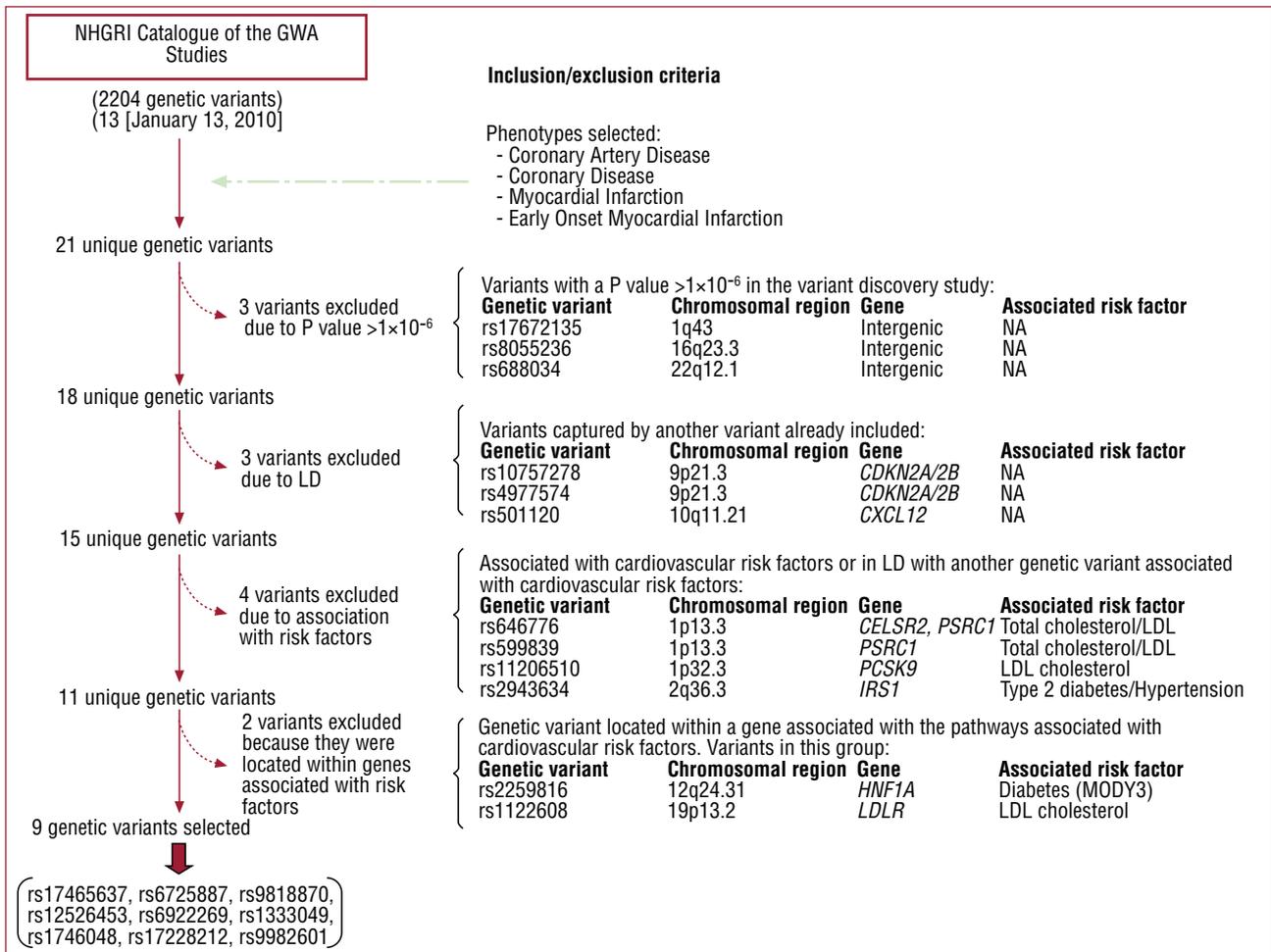
**NHGRI Catalogue of the GWA Studies**

(2204 genetic variants)
(13 [January 13, 2010]

**Inclusion/exclusion criteria**

Phenotypes selected:
- Coronary Artery Disease
- Coronary Disease
- Myocardial Infarction
- Early Onset Myocardial Infarction

21 unique genetic variants

3 variants excluded due to P value >1×10⁻⁶

Variants with a P value >1×10⁻⁶ in the variant discovery study:

| Genetic variant | Chromosomal region | Gene | Associated risk factor |
|---|---|---|---|
| rs17672135 | 1q43 | Intergenic | NA |
| rs8055236 | 16q23.3 | Intergenic | NA |
| rs688034 | 22q12.1 | Intergenic | NA |

18 unique genetic variants

3 variants excluded due to LD

Variants captured by another variant already included:

| Genetic variant | Chromosomal region | Gene | Associated risk factor |
|---|---|---|---|
| rs10757278 | 9p21.3 | *CDKN2A/2B* | NA |
| rs4977574 | 9p21.3 | *CDKN2A/2B* | NA |
| rs501120 | 10q11.21 | *CXCL12* | NA |

15 unique genetic variants

4 variants excluded due to association with risk factors

Associated with cardiovascular risk factors or in LD with another genetic variant associated with cardiovascular risk factors:

| Genetic variant | Chromosomal region | Gene | Associated risk factor |
|---|---|---|---|
| rs646776 | 1p13.3 | *CELSR2, PSRC1* | Total cholesterol/LDL |
| rs599839 | 1p13.3 | *PSRC1* | Total cholesterol/LDL |
| rs11206510 | 1p32.3 | *PCSK9* | LDL cholesterol |
| rs2943634 | 2q36.3 | *IRS1* | Type 2 diabetes/Hypertension |

11 unique genetic variants

2 variants excluded because they were located within genes associated with risk factors

Genetic variant located within a gene associated with the pathways associated with cardiovascular risk factors. Variants in this group:

| Genetic variant | Chromosomal region | Gene | Associated risk factor |
|---|---|---|---|
| rs2259816 | 12q24.31 | *HNF1A* | Diabetes (MODY3) |
| rs1122608 | 19p13.2 | *LDLR* | LDL cholesterol |

9 genetic variants selected

rs17465637, rs6725887, rs9818870,
rs12526453, rs6922269, rs1333049,
rs1746048, rs17228212, rs9982601

**Figure 1.** Flow diagram of the selection process of the genetic variants in the study.

**TABLE 2. Association Between the Selected Genetic Variants and the Risk of Coronary Artery Disease in the Larger WTCCC (Wellcome Trust Consortium of Cases and Controls) Sample**

| | Results obtained in the Variant Discovery Study and Registered in the NHGRI Catalogue of GWA Studies[a] | | | | | | Association Data Obtained in this Analysis[b] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ref. | Genetic Variant | Chromosomal Region | Genes Close at the Variant Identified | Alleles | Risk Alleles | OR (95% IC) | P Value | Minor Allele | MAF Controls | MAF Cases | OR (95% IC) | P Value | HWE |
| 4.6 | rs17465637[c] | 1q41 | *MIA3* | C/A | C | 1.2 (1.12-1.3) | 1×10⁻⁶ | A | 0.287 | 0.251 | 1.24 (1.13-1.36) | 1×10⁻⁵ | .894 |
| 4 | rs6725887 | 2q33.1 | *WDR12* | T/C | C | 1.17 (1.11-1.23) | 1×10⁻⁸ | C | 0.133 | 0.142 | 1.09 (0.98-1.21) | 1×10⁻¹ | .058 |
| 2 | rs9818870 | 3q22.3 | *MRAS* | C/T | T | 1.15 (1.11-1.19) | 7×10⁻¹³ | T | 0.153 | 0.169 | 1.12 (1.02-1.23) | 2×10⁻² | .528 |
| 4 | rs12526453 | 6p24.1 | *PHACTR1* | C/G | C | 1.12 (1.08-1.17) | 1×10⁻⁹ | G | 0.341 | 0.316 | 1.13 (1.04-1.24) | 4×10⁻³ | .927 |
| 6 | rs1474787[d] | 6q25.1 | *MTHFD1L* | C/T | T | 1.23 (1.15-1.33) | 3×10⁻⁸ | T | 0.27 | 0.297 | 1.09 (0.99-1.19) | 1×10⁻¹ | .557 |
| 3-6.8 | rs1333049[c] | 9p21.3 | *CDKN2A/2B* | G/C | C | 1.47 (1.27-1.7) | 1×10⁻¹³ | G | 0.478 | 0.445 | 1.43 (1.28-1.6) | 7×10⁻¹⁷ | .428 |
| 4.6 | rs1746048[c] | 10q11.21 | *CXCL12* | C/T | C | 1.17 (1.11-1.24) | 7×10⁻⁹ | T | 0.128 | 0.115 | 1.18 (1.06-1.31) | 2×10⁻² | .179 |
| 6 | rs17228212 | 15q22.33 | *SMAD3* | C/T | C | 1.21 (1.13-1.3) | 2×10⁻⁷ | C | 0.298 | 0.331 | 1.16 (1.06-1.27) | 1×10⁻⁴ | .745 |
| 4 | rs9978407[d] | 21q22.11 | SLC5A3, MRPS6 | G/A | A | 1.2 (1.14-1.27) | 6×10⁻¹¹ | A | 0.134 | 0.159 | 1.24 (1.12-1.37) | 1×10⁻⁴ | .481 |

HWE indicates P value obtained in the test to analyse Hardy-Weinberg equilibrium in controls; MAF, minor allele frequency; OR (95% CI), odds ratio (95% confidence interval).
The chromosomal region and the gene presented were extracted from the catalogue of published genome-wide association (GWA) Studies of the NHGRI,[14] which in turn extracted the information from the UCSC Genome Browser, as described. The associations that are presented are for an additive inheritance model, in other words, the risk of CAD increases with each copy of the risk allele.
[a]Values described in the variant discovery study.
[b]Data obtained in the sample of 1988 cases and 5380 controls from the WTCCC8 study analyzed.
[c]Genetic variants that have been associated in more than one GWA study. The data included in this table belong to the first study that defined the association.
[d]Genetic variants that respectively capture the variants rs6922269, rs9982601 described in the GWA studies database.
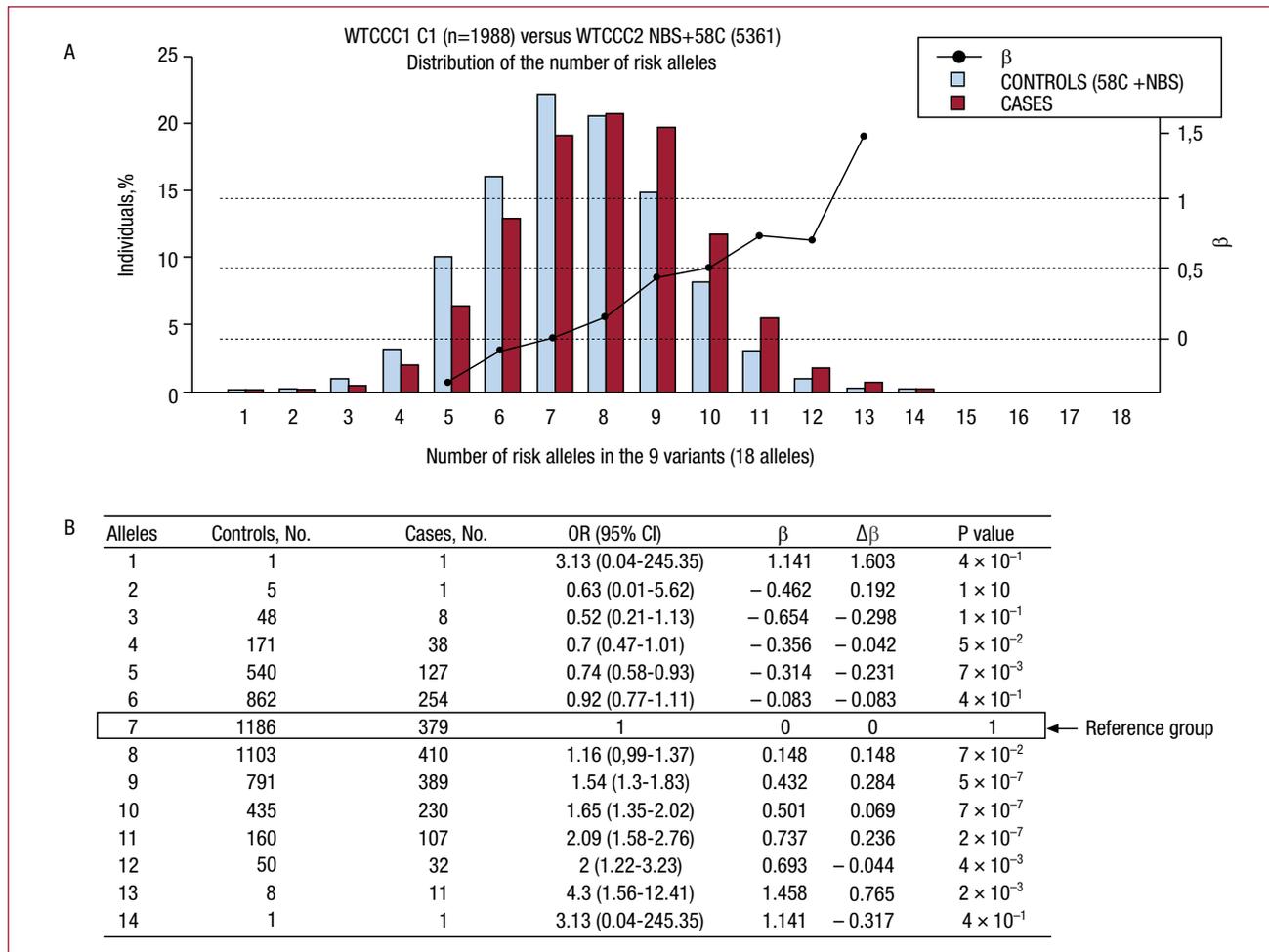
**A:** WTCCC1 C1 (n=1988) versus WTCCC2 NBS+58C (5361). Distribution of the number of risk alleles.

**B:**

| Alleles | Controls, No. | Cases, No. | OR (95% CI) | $\beta$ | $\Delta\beta$ | P value |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 3.13 (0.04-245.35) | 1.141 | 1.603 | $4 \times 10^{-1}$ |
| 2 | 5 | 1 | 0.63 (0.01-5.62) | − 0.462 | 0.192 | $1 \times 10$ |
| 3 | 48 | 8 | 0.52 (0.21-1.13) | − 0.654 | − 0.298 | $1 \times 10^{-1}$ |
| 4 | 171 | 38 | 0.7 (0.47-1.01) | − 0.356 | − 0.042 | $5 \times 10^{-2}$ |
| 5 | 540 | 127 | 0.74 (0.58-0.93) | − 0.314 | − 0.231 | $7 \times 10^{-3}$ |
| 6 | 862 | 254 | 0.92 (0.77-1.11) | − 0.083 | − 0.083 | $4 \times 10^{-1}$ |
| 7 | 1186 | 379 | 1 | 0 | 0 | 1 | ← Reference group |
| 8 | 1103 | 410 | 1.16 (0,99-1.37) | 0.148 | 0.148 | $7 \times 10^{-2}$ |
| 9 | 791 | 389 | 1.54 (1.3-1.83) | 0.432 | 0.284 | $5 \times 10^{-7}$ |
| 10 | 435 | 230 | 1.65 (1.35-2.02) | 0.501 | 0.069 | $7 \times 10^{-7}$ |
| 11 | 160 | 107 | 2.09 (1.58-2.76) | 0.737 | 0.236 | $2 \times 10^{-7}$ |
| 12 | 50 | 32 | 2 (1.22-3.23) | 0.693 | − 0.044 | $4 \times 10^{-3}$ |
| 13 | 8 | 11 | 4.3 (1.56-12.41) | 1.458 | 0.765 | $2 \times 10^{-3}$ |
| 14 | 1 | 1 | 3.13 (0.04-245.35) | 1.141 | − 0.317 | $4 \times 10^{-1}$ |

**Figure 2.** A: the distribution of the percentage of individuals according to the number of risk alleles in the sample of cases and controls analyzed. B: the magnitude of association between the exposure to different categories of the number of risk alleles and the presence of coronary artery disease, defining the group with 7 risk alleles as the reference group. $\beta$ indicates regression coefficient; $\Delta\beta$, difference in the regression coefficient between consecutive categories of the number of risk alleles; OR (95% CI), odds ratio (95% confidence interval); 58C, control subjects in the study: 1958 Birth Cohort Controls[8,23]; NBS, controls in the study: UK Blood Services Controls.[8]

The results of the analyses defining 5 groups according to the quintiles of the number of risk alleles and using the quintile with the lowest number of alleles as reference are shown in figure 3A. The increases in risk ($\Delta\beta$) between the consecutive quintiles are similar (mean increase of $\beta$, 0.198), and the OR between extreme quintiles is 2.21 (Figure 3B).

## DISCUSSION

Nine genetic variants associated with CAD and which, according to current scientific knowledge, are not related to CVRFs were selected for this study. In accordance with the AHA recommendations regarding the evaluation of new risk biomarkers, the data of this study are consistent with the first stage of this evaluation; it has been observed that the number of accumulated risk alleles of these genetic variants is associated with the risk of presenting CAD.

We have demonstrated that the association between the number of risk alleles and the risk of CAD is linear and direct, therefore the higher the number of risk alleles, the greater the probability of presenting CAD. This linear association has been observed in other phenotypes like blood pressure[16] and diabetes.[17] In this study each allele was considered to be of similar weight and we did not weight the effect of each genetic variant according to the magnitude of its association with CAD, which is defined by the regression coefficient. We weighted the variants according to the coefficients obtained in our analysis in a sensitivity analysis and found that the results were similar to those presented. This similarity in the results obtained once the effect
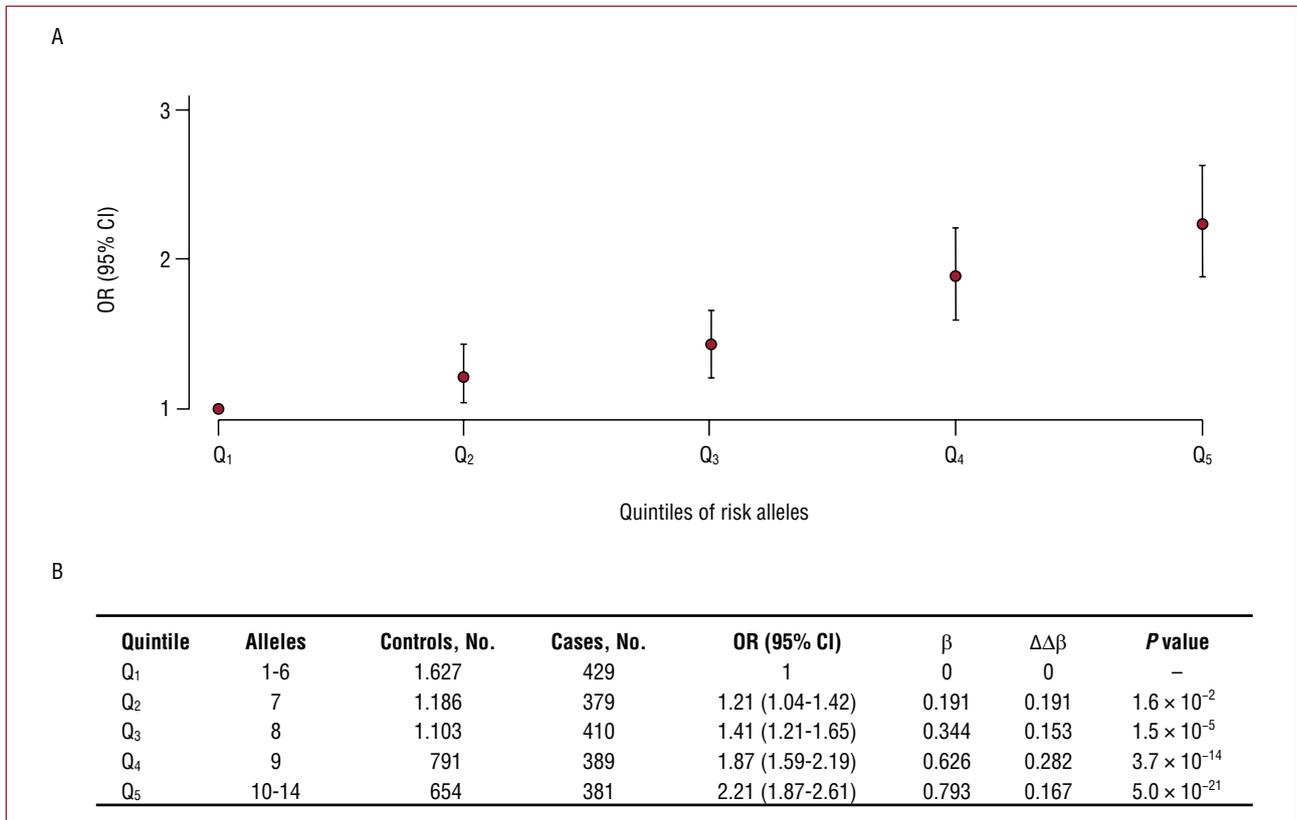
**Figure 3.** A: odds ratio and 95% confidence interval of the association between quintiles of the number of risk alleles and the presence of coronary artery disease, taking the quintile with the lowest number of risk alleles as the reference category. B: the magnitudes of said associations. β indicates regression coefficient; Δβ, difference in the regression coefficient between contiguous groups; OR (95% CI), odds ratio (95% confidence interval).

of the various alleles had been analyzed and after weighting them according to the magnitude of their association with CAD has also been observed in other phenotypes, such as diabetes.[18]

We have observed that the magnitude of the association observed, determined by the regression coefficient, is similar to that observed in some CVRFs included in the risk equations. For example, in the Framingham risk equation and its adaptations,[19] the difference in men between the regression coefficient at the extremes of exposure of the cholesterol variable (<160 and ≥280mg/dL) is 1.32 (which corresponds to an OR=3.74) and of the blood pressure variable (optimal and grade I-III hypertension) is 0.62 (OR=1.86). In our analysis, the difference between the regression coefficients of the extreme categories of quintiles of the number of alleles is 0.79, (OR=2.2) which is similar to the blood pressure values; considering the number of alleles and taking the group of 4 and 12 risk alleles as extremes, the difference in the regression coefficients is 1.05 (OR=2.86), which is slightly lower than the cholesterol values. It is important to consider that one of the limitations of this comparison is that we are comparing regression coefficients obtained in case-control studies that tend to slightly overestimate the magnitude of the association.

The next stage in the evaluation of these genetic biomarkers is to examine cohort studies to establish whether they can help to improve coronary risk prediction, or to reclassify individuals, in particular those in the intermediate-risk group.[9] There are already some cohort studies that have included a genetic variant of the chromosome 9p21 in the risk equations but have not observed any significant improvement in the discriminating ability of the predictive models,[20-22] probably because the magnitude of association of the variant analyzed was small (relative risk, RR<1.3). Other studies have included a genetic risk score based on the number of accumulated risk alleles in an individual,[4,23-26] with the aim of increasing the magnitude of the association observed. Among these are 4 cohort studies,[23-26] the first of which was based on candidate genes and highlighted that the inclusion of a genetic risk score can improve the predictive ability of classic risk equations.[23] Two of them,[24-25] which included various polymorphisms in genes involved in lipid metabolism, demonstrated that the genetic risk score was associated with the risk of presenting a

cardiovascular event, and even though the predictive ability assessed by the C statistic did not improve, the reclassification of individuals in risk categories did improve slightly. One of the limitations of these studies is that by including markers in risk equations that are associated with variables or factors that are already included, no additional or complementary information is contributed, since no new biological pathways or mechanisms associated with the disease are considered. Of the 21 genetic variants associated with CAD that are included in the NHGRI Catalog of GWA Studies, we have identified 9 that are independent of CVRFs and good candidates for inclusion in a genetic risk score in coronary risk equations. The results of a third cohort study, the Nurse Health Study,[26] which followed 19 313 women and had a similar focus to that of our study, have been published recently. More than 100 genetic variants were included in the first analysis, and in the second only 12 were included, the majority of which were independent of CVRFs. These were analyzed to establish whether a genetic score based on the number of risk alleles of these genetic variants would improve predictive ability or the reclassification of coronary risk. The results of this study have been negative, although the external validity of these results is debatable.

One of the limitations of our study is associated with the analysis of the in silico data from the WTCCC[8] study, which gave us access to individual-level data about the participant's genotype and whether or not they presented CAD; however, we did not have access to other personal data like age or sex, which would have allowed us to adjust or stratify the results using these kinds of clinical variables. In any case, there are no data to suggest that the association between the selected genetic variants and CAD is different in men than in women.

## CONCLUSIONS

The results of this study suggest that this type of genetic risk score for CAD, based on the additive and combined effect of a number of risk alleles in different genetic markers that are independent of CVRFs, is associated with an increased risk of CAD and is consistent with the first stage of the evaluation of a new biomarker proposed by the American Heart Association. It is now necessary to continue with the next stages of the evaluation of this type of biomarker by carrying out population-based cohort studies in order to establish whether there is any improvement in the discriminating ability of the equations used in the primary prevention of CAD, or the reclassification of individuals, especially those belonging to the intermediate-risk group. It is also important to define the relevance of this type of biomarker in clinical practice.

## REFERENCES

1. Elosua R, Lluís C, Lucas G. Estudio del componente genético de la cardiopatía isquémica: de los estudios de ligamiento al genotipado integral del genoma. Rev Esp Cardiol Supl. 2009;9 Suppl B:B24-38.
2. Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet. 2009;41:280-2.
3. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007;316:1491-3.
4. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet. 2009;41:334-41.
5. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007;316:1488-91.
6. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007;357:443-53.
7. Tregouet DA, Konig IR, Erdmann J, Munteanu A, Braund PS, Hall AS, et al. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. Nat Genet. 2009;41:283-5.
8. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447:661-78.
9. Gerszten RE and Wang TJ. The search for new cardiovascular biomarkers. Nature. 2008;451:949-52.
10. Marrugat J, D'Agostino R, Sullivan L, Elosua R, Wilson P, Ordovas J, et al. An adaptation of the Framingham coronary heart disease risk function to European Mediterranean areas. J Epidemiol Community Health. 2003;57:634-8.
11. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MS, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. Circulation. 2009;119:2408-16.
12. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int J Epidemiol. 2006;35:34-41.
13. EMBL-EBI. EMBL-EBI website [cited Dic 2, 2009]. Available from: http://www.ebi.ac.uk/ega/page.php?page= studies&name=WTCCCe
14. Hindorff LA, Junkins HA, Mehta JP, Manolio TA. A catalog of published genome-wide association studies [cited Jan 13, 2010]. Available from: www.genome.gov/gwastudies
15. Gonzalez JR, Armengol L, Sole X, Guino E, Mercader JM, Estivill X, et al. SNPassoc: an R package to perform whole genome association studies. Bioinformatics. 2007;23:644-5.
16. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009;41:666-76.

17. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42:105-16.

18. Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, et al. Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. Ann Intern Med. 2009;150:541-50.

19. Marrugat J, Solanas P, D'Agostino R, Sullivan L, Ordovás J, Cordón F. Estimación del riesgo coronario en España mediante la ecuación de Framingham calibrada. Rev Esp Cardiol. 2003;56:253-61.

20. Brautbar A, Ballantyne CM, Lawson K, Nambi V, Chambless L, Folsom AR, et al. Impact of adding a single allele in the 9p21 locus to traditional risk factors on reclassification of coronary heart disease risk and implications for lipid-modifying therapy in the Atherosclerosis Risk in Communities study. Circ Cardiovasc Genet. 2009;2:279-85.

21. Paynter NP, Chasman DI, Buring JE, Shiffman D, Cook NR, Ridker PM. Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. Ann Intern Med. 2009;150:65-72.

22. Talmud PJ, Cooper JA, Palmen J, Lovering R, Drenos F, Hingorani AD, et al. Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of CHD in healthy middle-aged men. Clin Chem. 2008;54:467-74.

23. Morrison AC, Bare LA, Chambless LE, Ellis SG, Malloy M, Kane JP, et al. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. Am J Epidemiol. 2007;166:28-35.

24. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med. 2008;358:1240-9.

25. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet. 2009;41:47-55.

26. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, et al. Association between a literature-based genetic risk score and cardiovascular events in women. JAMA. 2010;303:631-7.