DR. JAVIER NARVÁEZ (Orcid ID : 0000-0002-1614-8064)
DR. ISIDORO GONZALEZ-ALVARO (Orcid ID : 0000-0001-9614-5199)

DR. IVAN FERRAZ-AMARO (Orcid ID : 0000-0003-0197-5267)

Article type : Full Length

A genome-wide association study identifies a 3'UTR genetic variant of *RARB* associated with carotid intima-media thickness in rheumatoid arthritis

Raquel López-Mejías, PhD[1¥], F. David Carmona, PhD[2¥], Fernanda Genre, PhD[1], Sara Remuzgo-Martínez, PhD[1], Carlos González-Juanatey, MD, PhD[3], Alfonso Corrales, MD, PhD[1], Esther F. Vicente, MD, PhD[4], Verónica Pulito-Cueto, BSc[1], José A. Miranda-Filloy, MD, PhD[5], Marco A. Ramírez Huaranga, MD, PhD[6], Ricardo Blanco, MD, PhD[1], Montserrat Robustillo-Villarino, MD, PhD[7], Javier Rodríguez-Carrio, PhD[8, 9, 10], Mercedes Alperi-López, MD[10, 11], Juan J. Alegre-Sancho, MD[12], Verónica Mijares, BSc[1], Leticia Lera-Gómez, BSc[1], Eva Pérez-Pampín, MD[13], Antonio González, PhD[13], Rafaela Ortega-Castro, MD[14], Chary López-Pedrera, PhD[14], Mari L. García Vivar, MD[15], Catalina Gómez-Arango, MD[15], Enrique Raya, MD, PhD[16], Javier Narvaez, MD[17], Alejandro Balsa, MD, PhD[18], Francisco J. López-Longo, MD, PhD[19], Patricia Carreira, MD, PhD[20], Isidoro González-Álvaro, MD, PhD[4], Luis Rodríguez-Rodríguez, MD, PhD[21], Benjamín Fernández-Gutiérrez, MD, PhD[21], Iván Ferraz-Amaro, MD, PhD[22], Oreste Gualillo, PhD[23], Santos Castañeda, MD, PhD[4], Javier Martín, MD, PhD[24§], Javier Llorca, MD, PhD[25§], Miguel A. González-Gay, MD, PhD[1, 26, 27§]

[1]Epidemiology, Genetics and Atherosclerosis Research Group on Systemic Inflammatory Diseases, Rheumatology Division, IDIVAL, Santander, Spain.
[2]Departamento de Genética e Instituto de Biotecnología, Universidad de Granada, Granada, Spain.
[3]Cardiology Department, Hospital Universitario Lucus Augusti, Lugo, Spain.
[4]Rheumatology Department, Hospital Universitario de la Princesa, IIS-Princesa, Madrid, Spain.
[5]Division of Rheumatology, Hospital Universitario Lucus Augusti, Lugo, Spain.
[6]Rheumatology Department, Hospital General Universitario de Ciudad Real, Ciudad Real, Spain.
[7]Rheumatology Department, Hospital Universitario de la Plana, Castellón, Spain.
[8]Area of Immunology, Department of Functional Biology, University of Oviedo, Oviedo, Spain.
[9]Bone and Mineral Research Unit, Instituto Reina Sofía de Investigación Nefrológica, REDinREN del ISCIII, Hospital Universitario Central de Asturias, Oviedo, Spain.
[10]Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Spain.
[11]Department of Rheumatology, Hospital Universitario Central de Asturias, Oviedo, Spain.
[12]Rheumatology Department, Hospital Universitario Doctor Peset, Valencia, Spain.
[13]Experimental and Observational Rheumatology and Rheumatology Unit, Instituto de Investigación Sanitaria, Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain.
[14]Instituto Maimónides de Investigación Biomédica de Córdoba (IMIBIC), Hospital Universitario Reina Sofia, Universidad de Córdoba, Córdoba, Spain.

[15]Rheumatology Department, Hospital Universitario de Basurto, Bilbao, Spain.
[16]Rheumatology Department, Hospital Universitario San Cecilio, Granada, Spain.
[17]Department of Rheumatology, Hospital Universitario de Bellvitge, Barcelona, Spain.
[18]Department of Rheumatology, Hospital Universitario La Paz, IdiPAZ, Madrid, Spain.
[19]Department of Rheumatology, Hospital General Universitario Gregorio Marañón, Madrid, Spain.
[20]Department of Rheumatology, Hospital Universitario 12 de Octubre, Madrid, Spain.
[21]Department of Rheumatology, Hospital Clínico San Carlos, IdISSC, Madrid, Spain.
[22]Rheumatology Division, Hospital Universitario de Canarias, Tenerife, Spain.
[23]The NEIRID Group (Neuroendocrine Interactions in Rheumatology and Inflammatory Diseases), Servizo Galego de Saude and Instituto de Investigación Sanitaria de Santiago, Santiago University Clinical Hospital, Santiago de Compostela, Spain.
[24]Instituto de Parasitología y Biomedicina 'López-Neyra', CSIC, PTS Granada, Granada, Spain.
[25]Department of Epidemiology and Computational Biology, School of Medicine, University of Cantabria, and CIBER Epidemiología y Salud Pública (CIBERESP), IDIVAL, Santander, Spain.
[26]School of Medicine, University of Cantabria, Santander, Spain.
[27]Cardiovascular Pathophysiology and Genomics Research Unit, School of Physiology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

[¥]These authors contributed equally to this work.
[§] These authors shared senior authorship in this study.

Address correspondence and reprint requests to: Raquel López-Mejías, PhD and Miguel Angel González-Gay, MD, PhD. Epidemiology, Genetics and Atherosclerosis Research Group on Systemic Inflammatory Diseases, Rheumatology Division, IDIVAL, Avenida Cardenal Herrera Oria s/n, 39011, Santander, Spain. Phone: +34 942 315 515 (74115/4). Fax: +34 942 31 55 17. E-mail: rlopezmejias78@gmail.com and miguelaggay@hotmail.com.

This study was supported by European Union FEDER funds and "Fondo de Investigación Sanitaria" (grants PI06/0024, PS09/00748, PI12/00060 and PI15/00525) from 'Instituto de Salud Carlos III' (ISCIII, Health Ministry, Spain). It was also partially supported by RETICS Programs RD12/0009 and RD16/0012 (RIER) from ISCIII, and in part by grants from the European IMI BTCure Program. RL-M is a recipient of a Miguel Servet type I programme fellowship from the ISCIII, co-funded by the European Social Fund (ESF, "Investing in your future") (grant CP16/00033). FDC was supported by the "Ramón y Cajal" programme of the Spanish Ministry of Economy and Competitiveness through the grant RYC-2014-16458. FG is a recipient of a Sara Borrell post-doctoral fellowship from the ISCIII, co-funded by the ESF (grant CD15/00095). SR-M is supported by funds of the RETICS Program (RIER) RD16/0012/0009 (ISCIII, co-funded by ERDF). VP-C is supported by a pre-doctoral grant from IDIVAL (PREVAL 18/01). VM is supported by funds of a Miguel Servet type I programme (grant CP16/00033) (ISCIII, co-funded by ERDF).

Conflicts of interests: The authors declare that they have no conflicts of interest.

ABSTRACT

Objective: A genome-wide association study (GWAS) was conducted to shed light into the genetic background influencing the development of cardiovascular (CV) disease in patients diagnosed with rheumatoid arthritis (RA).

Methods: After quality control and imputation, a total of 6,308,944 polymorphisms across the whole genome were analysed in 2,989 RA patients from European origin. Data on subclinical atherosclerosis, obtained by carotid ultrasonography through assessment of carotid intima-media thickness (cIMT) and presence/absence of carotid plaques, were available for 1,355 individuals.

Results: A genetic variant of the *RARB* gene (rs116199914) was associated with cIMT values at the genome-wide level of significance (minor allele (G): beta (β) coefficient=0.142, P=1.86E-08). Interestingly, rs116199914 overlapped with regulatory elements in tissues related to CV pathophysiology and immune cells. In addition, biological pathway enrichment and predictive protein-protein relationship analyses, including suggestive GWAS signals of potential relevance, revealed a functional enrichment of the collagen biosynthesis network related to the presence/absence of carotid plaques (GO:0032964, P_{FDR}=4.01E-03). Furthermore, our data suggest a potential influence of the previously described candidate CV risk *loci NFKB1, MSRA* and *ZC3HC1* (P=8.12E−04, P=5.94E−04 and P=2.46E-04, respectively).

Conclusion: Our study strongly suggests that genetic variation within *RARB* contributes to the development of subclinical atherosclerosis in patients with RA.

INTRODUCTION

Cardiovascular (CV) disease is the most common cause of morbidity and mortality in patients diagnosed with rheumatoid arthritis (RA) (1-3). This condition develops as a result of an accelerated atherosclerotic process (4). Surrogate markers for subclinical atherosclerosis, increased carotid intima-media thickness (cIMT) values and presence of carotid plaques (5, 6), are excellent predictors of future CV events. Traditional CV risk factors and chronic inflammation do not fully explain the increased CV predisposition observed in patients with RA, but only around 70% of the population attributable risk for CV disease outcomes (7). In this regard, cumulative knowledge clearly suggests that genetic factors may play a relevant role in this phenomenon (8). Nevertheless, the genetic component of CV disease in RA remains elusive.

Genome-wide association studies (GWAS) constitute a hypothesis-free approach in which millions of common genetic variations across the whole genome are interrogated (9). This strategy has been of great help to unravel relevant inroads into the genetics of several complex human diseases (10). The use of this technology has substantially increased the number of established RA susceptibility *loci* from 3 to more than 100 during the last decade (11). Nevertheless, there are currently no available GWAS data specifically focused on CV disease in patients with RA.

Taking together all these considerations, we carried out a multicentric study aimed at conducting the first GWAS on the development of CV disease in RA using a large series of patients in whom both the presence/absence of CV events and subclinical atherosclerosis was evaluated.

PATIENTS AND METHODS

Study Population

A total of 3,433 unrelated Spanish patients of European ancestry diagnosed with RA, according to the 2010 American College of Rheumatology classification criteria (12), were enrolled in the study. Centers involved in patient recruitment included Hospital Universitario Lucus Augusti (Lugo), Hospital Universitario Marqués de Valdecilla (Santander), Hospital Universitario de Basurto (Bilbao), Hospital Universitario Central de Asturias (Oviedo), Hospital Clínico Universitario de Santiago (Santiago de Compostela), Hospital Universitario de Bellvitge (Barcelona), Hospital Universitario San Cecilio (Granada), Hospital Universitario Reina Sofía (Córdoba), Hospital Universitario de Canarias (Tenerife), Hospital Universitario Doctor Peset (Valencia), Hospital General Universitario de Ciudad Real (Ciudad Real) and Hospital Clínico San Carlos, Hospital Universitario La Paz, Hospital Universitario de La Princesa, Hospital General Universitario Gregorio Marañón, and Hospital Universitario 12 de Octubre (Madrid). All patients signed an informed consent before being included in the study according to the declaration of Helsinki. The procedures followed were in accordance with the ethical standards of the responsible committees on human experimentation of all participant centers.

Genotyping and Quality Controls

Genomic DNA was extracted from peripheral blood using standard procedures. Genotyping was conducted in the Human Genotyping Unit of the National Genotyping Center (CEGEN-CENIO, Spain) using the GWAS platform “Infinium® HumanCore

Beadchip" in an iScan System (Illumina), following the manufacturer's protocol. Single nucleotide polymorphisms (SNP) with a cluster separation <0.4 were removed after the calling.

Raw data were subjected to stringent quality-control (QC) filters using the software PLINK v.1.07 (13). Polymorphisms with call rates <0.98 and minor allele frequencies <0.01, as well as and those that deviated from Hardy-Weinberg equilibrium ($p<0.001$) were filtered out. Similarly, samples with less than 95% of successfully called polymorphisms, and one subject per pair of first-degree relatives (identity by descent >0.4) were removed. Sex chromosomes were also excluded from the analysis.

To ensure reliability of the results, the associated SNP described below was re-genotyped using predesigned TaqMan 5´ SNP genotyping assay (C_154503570_10) in a 7900HT Fast Real-Time PCR System (Applied Biosystems), and the TaqMan types were compared with the corresponding imputed data.

Imputation Methods

After applying the QC filters, whole-genome SNP genotype imputation in autosomal chromosomes was carried out in the Michigan Imputation Server (MIS) (14), using the ShapeIT16 software (version v2.r790) for haplotype reconstruction and the updated Haplotype Reference Consortium (HRC) data (version r1.1) as reference panel, which combines sequencing data from a total of 32,470 individuals from multiple studies (including the 1000 Genomes Project, 1KG) (15). The QC filters mentioned above were also applied to the imputed data with PLINK. Besides, singletons were excluded (r^2 ≤0.2). Finally, possible population sub-stratification was controlled by principal

component (PC) analyses using PLINK and the gcta64 and R-base software under GNU Public license v2. The ten first PCs of each individual were calculated and plotted to identify outliers, and those deviating >4 standard deviations from the cluster centroid were excluded.

After QC, 6,308,944 SNPs and 2,989 RA patients remained for the analysis in the final dataset. A detailed description of the main demographic data, clinical, and CV disease-related characteristics of these patients is displayed in **Table 1**. Definitions of traditional CV risk factors were established elsewhere (3, 6). Information related to CV events was obtained from the medical records of each patient, according to definitions previously described (3, 6). Briefly, ischaemic heart disease (IHD) was diagnosed if any of the following criteria were satisfied: a recorded diagnosis of ischemic cardiopathy, on account of some acute coronary syndrome (acute myocardial infarction or unstable angina), the presence of pathological Q waves in the electrocardiogram, and/or coronary images showing >50% stenosis of at least one coronary vessel. A patient was considered to have a heart failure based on the Framingham criteria. Cerebrovascular accident was diagnosed if patients had a stroke and/or transient ischemic attacks (TIAs). Strokes were classified according to their clinical features and they were confirmed by computed tomography and/or magnetic resonance imaging. TIAs were diagnosed if the symptoms were self-limited in less than 24 hours, without residual neurological damage. Finally, peripheral arterial disease was considered to be present if it was confirmed by Doppler and arteriography (3, 6).

Subclinical Atherosclerosis Examination

Data on subclinical atherosclerosis was available for 1,355 RA patients of the filtered datasets. Subclinical atherosclerosis examination was assessed by a carotid ultrasound (US) technique (by evaluation of cIMT and presence/absence of carotid plaques). The hospitals from Santander, Bilbao, Granada, Córdoba, Tenerife, Valencia, Ciudad Real and Madrid performed the US examination by using a commercial scanner (16, 17). Patients from Lugo were assessed by high-resolution B-mode ultrasound (18). cIMT was measured at the far wall of the right and left common carotid arteries over the proximal 15 mm-long segment. cIMT was determined as the average of three measurements in each common carotid artery. Consistency between these two US methods was previously reported (19), supporting the fact that the use of two different instruments to collect cIMT data did not influence on the results derived from this analysis. In addition, experts with a high reproducibility, excellent inter-observer reliability and close collaboration in the assessment of subclinical atherosclerosis in RA performed these studies. Plaque criteria, in the accessible extracranial carotid tree, were defined as described by Touboul *et al* (20).

Statistical Analyses

A statistical power estimation of our study was obtained with the CaTS Power Calculator for Genetic Studies software, which implements the methods described by Skol *et al* (21) (**Supplementary Tables 1-4**).

All the statistical analyses were conducted with PLINK. First, we compared the genotype frequencies of every SNP according to a continuous CV disease outcome variable (cIMT values) by lineal regression assuming an additive model. The ten first PCs, age at the time of the carotid US examination, and gender were included in the model as covariates. Subsequently, we compared the genotype frequencies of every SNP according to binary CV disease outcome variables (presence/absence of CV events, IHD and carotid plaques) by logistic regression on the best-guess genotypes assuming an additive model. The ten first PCs, age at the time of the RA diagnosis, and gender were included as covariates for the presence/absence of the CV events and IHD analyses, whereas the ten first PCs, age at the time of the carotid US examination, and gender were included as covariates for the presence/absence of the carotid plaques analysis. Finally, P-values, beta coefficients (β), standard errors (SE), odds ratios (OR), and 95% confidence intervals (CI) were calculated. The statistical threshold was set at the genome-wide level of significance (P<5E-08).

Functional Annotations of the Associated Variants

In a further step, we evaluated the putative functional implications of the identified CV risk signals by implementing our data with functional annotation data available in public databases using different bioinformatics approaches.

For that purpose, we first identified all the potential polymorphisms in high linkage disequilibrium (LD; r^2>0.8) of the associated signals of our GWAS using the European populations of the 1KG and PLINK. All those potential polymorphism taggers would be considered equally as candidates for prioritizing casualty or hypothesizing possible

molecular causes of the observed associations in the subsequent bioinformatic approaches. Then, the online tools RegulomeDB (22), HaploReg v.4.1. (23) and Capture HiC Plotter (CHi-CP) (24) were used to evaluate the possible regulatory effect of the associated signals and their possible implication on the analysed clinical phenotypes.

Candidate Genomic regions and Pathway Enrichment Analysis

Finally, we assessed the statistical significance in our GWAS of previously described CV risk associated genomic regions (±100 Kbp 3´ and 5´ of the reported gene) through candidate genes studies (8) and a recently published meta-analysis of Immunochip data (25).

Regarding the human leukocyte antigen (HLA) region, a more comprehensive analysis was conducted. We extracted the extended HLA region (29,000,000 to 34,000,000 bp in chromosome 6) and imputed SNPs, classical HLA alleles at two- and four-digits, and polymorphic amino acid positions, as previously described (26-28).

Additionally, a biological pathway enrichment analysis involving those genes showing suggestive P-values in our study (P<1E-04) was performed by using the tool for that purpose of the Gene Ontology (GO) reference genome project (29, 30), powered by the Protein Analysis Through Evolutionary Relationships (PANTHER) Classification System (31). Moreover, we conducted a predictive protein-protein interaction analysis amongst these same markers using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (32).

P-values lower than 0.05 after multiple testing correction were considered statistically significant.

RESULTS

Testing for association with CV disease outcomes

Figure 1 and **Supplementary Figure 1** summaries the overall results obtained for each CV disease outcome analysis performed.

Interestingly, a statistically significant signal at the genome-wide level of significance was associated with cIMT values (**Figure 1**). This signal corresponded to the genetic variant rs116199914, which maps to the 3´UTR region of the retinoic acid receptor beta (*RARB*) gene (**Table 2**). In this regard, the minor allele (G) of this SNP was significantly related to increased cIMT values (β=0.142, P=1.86E-08) (**Table 2**). To discard that a possible bias due to wrong genotyping or imputation could be affecting our results, we obtained direct genotypes using TaqMan probes for rs116199914. The overall concordance reached after comparing TaqMan types with the corresponding imputed data was 99.94%. Based on previous studies that reported association between anti-citrullinated protein antibodies (ACPA) presence and CV disease in RA (33, 34), the potential association between the genetic variant rs116199914 and ACPA status was evaluated in our study. However, no statistically significant results were observed (data not shown). Furthermore, several suggestive associations with cIMT values were also detected, although none of them reached the statistical genome-wide level of significance (**Figure 1**). Amongst them, intronic variants of both the *RARB* and PR domain containing 10 (*PRDM10)* genes as well as a disequilibrium block of intergenic polymorphisms at chromosome 12 showed the most suggestive P-values.

Likewise, several trends of association were observed when the presence/absence of CV events, IHD and carotid plaques were analysed (**Table 2, Supplementary Figure**

1). Regarding the presence/absence of CV events, two intergenic variants in high LD at chromosome 1 exhibited the lowest P-values. According to the presence/absence of IHD, an intronic variant of the Kinesin Family Member 26B (*KIF26B)* gene and two disequilibrium blocks of polymorphisms at chromosome 1 and 7 represented the strongest signals. Similarly, an intronic variant of the formin 2 (*FMN2)* gene and intergenic polymorphisms located at chromosomes 4, 9 and 17 exhibited the lowest P-values regarding the presence/absence of carotid plaques.

Similar results were obtained when the analyses were also performed considering traditional CV risk factors as covariates (smoking habit, diabetes mellitus, hypertension, obesity and dyslipidemia). In this regard, a statistically significant signal at the genome-wide level of significance that corresponded to *RARB* rs116199914 was associated with cIMT values (minor allele (G): β=0.137, P=4.35E-08). In addition, trends of association were observed when the presence/absence of CV events, IHD and carotid plaques were analysed (data not shown).

Functional annotations of the associated variants

We evaluated the possible functional implications of the associated genetic variant rs116199914 by integrating our data with public databases. First, we searched for proxies (r^2>0.8) of rs116199914 in the five populations of European origin of the 1KG project (Iberian Population in Spain, Utah residents with Northern and Western European ancestry from the CEPH collection, British in England and Scotland, Toscani in Italy, and Finnish in Finland). Since no proxies were identified, we functionally

annotated just the rs116199914 polymorphism. As this SNP is located in the 3´UTR region of the *RARB* gene, we used bioinformatic tools aimed at exploring annotations of the noncoding genome with putative regulatory effects on gene expression (including effect on regulatory motifs, chromatin state, protein binding and expression from eQTL studies) in Gene Expression Omnibus (GEO), the Encyclopedia of DNA Elements (ENCODE), the Roadmap Epigenomics, as well as promoter CHi-C datasets, and published literature.

Interestingly, the RegulomeDB results suggested that rs116199914 may represent a DNA element with relevant regulatory effects (score=6). Additional functional implications were suggested by both HaploReg v.4.1. and CHi-C tools. In particular, overlapping with histone marks in tissues related to CV pathophysiology and cells of the immune system were observed (**Figure 2**). Specifically, rs116199914 was described to overlap with the enhancer histone mark H3K4me1 and the promoter histone mark H3K9ac in fetal heart, and with histone marks enriched at promoters and enhancers in immune cells (**Figure 2**) (23). Furthermore, as derived from the CHi-C datasets, rs116199914 was described to interact, amongst others, with the nuclear factor kappa-light-chain-enhancer of activated B cells (NFKB) inhibitor interacting Ras like 1 (*NKIRAS1*) gene in total CD4 *Mycosis fungoides* cells and total CD8 cells (35) (**Supplementary Figure 2**). In addition, rs116199914 was described to affect the sequence-specific binding for the nuclear factor of activated T-cells (NFAT) (23).

Candidate Genes and Pathway Analysis

We also checked the statistical significance in our GWAS of previously described CV risk genes by candidate studies (8) and a recently published meta-analysis of Immunochip data (25). P-values <0.05 were observed across most of the evaluated *loci* (**Supplementary Table 5**). Amongst them, the lowest P-values were detected within the *NFKB1* and methionine sulfoxide reductase A (*MSRA*) regions with the presence of CV events (P=8.12E-04 and P=5.94E-04, respectively), as well as the zinc finger C3HC-type containing 1 (*ZC3HC1*) region with cIMT values (P=2.46E-04) (**Supplementary Table 5**). The association between *NFKB1* and CV events remained statistically significant after multiple testing correction (rs227361, P_{FDR}=4.50E-02). Regarding the HLA system, no statistically significant results were observed across this genomic region (**Supplementary Figure 3**).

In addition, analysis of possible biological pathway enrichments and predictive protein-protein relationships were performed considering the gene products of those *loci* showing P-values of potential relevance in our study (P<1E-04). In this regard, the molecular network of the selected proteins related to the presence/absence of carotid plaques had significantly more interactions than expected (number of nodes: 51, number of edges=8, average node degree=0.314, clustering coefficient=0.235; expected number of edges=3, protein-protein interaction enrichment P=1.68E-02; **Figure 3**). In accordance to the functional enrichments of the network, the most significantly associated GO term corresponded to “collagen biosynthetic process” (GO:0032964, P_{FDR}=4.01E-03). No statistically significant results were obtained when

these analyses were performed according to cIMT values, presence/absence of CV events, and IHD.

DISCUSSION

During the last decade, the genetic basis of the increased CV predisposition observed in RA patients has been thoroughly investigated following a candidate gene strategy (8). However, until the development of this study, no GWAS data has been generated and analysed. Therefore, the results presented here may represent a turning point for a better understanding of the pathogenic mechanisms underlying this severe complication of RA.

A genetic marker of the *RARB* gene (rs116199914) was associated at the genome wide level of significance with subclinical atherosclerosis, assessed by cIMT. Interestingly, this signal overlaps with promoter and enhancer histone marks in fetal heart and immune cells. In addition, rs116199914 was described to interact with the *NKIRAS1* gene. These data are striking, as *NKIRAS1* encodes a crucial protein for the inhibition of NF-κB (36, 37), one of the most relevant molecules involved in inflammation processes (38) that is considered a key regulator of several atherosclerosis genes (39). In this sense, a previous candidate gene study demonstrated the influence of a promoter genetic variant in the NF-κB coding gene (*NFKB1*) on the risk of developing CV events in patients with RA (39). Likewise, the use of drugs blocking cytokines of the NF-κB signaling pathway has been described as a promising therapeutic strategy to attenuate the heightened CV risk in patients with RA (40, 41) and to provide a beneficial effect on surrogate CV disease markers in those

patients (42, 43). In addition, the associated variant identified in our study was shown to affect the sequence-specific binding for NFAT, which regulates the inducible gene transcription during the immune response (44-46). Originally, NFAT was described as a specific protein for activated T cells (46) and other immune cells (45). Currently, regulatory roles of NFAT in blood vessels and heart tissues are well-stablished (47-49). Besides, an implication of this molecule in angiogenic processes has been confirmed (48). In line with this, cumulative knowledge clearly indicates that the chronic inflammation observed in patients with RA, critical for the development of atherosclerosis, is often accompanied by imbalanced angiogenesis (50). In accordance with that, increased serum levels of the angiogenic molecule angiopoietin-2 were previously described to correlate with the development of CV events in patients with RA (50). Our results suggest a functional impact of the genetic variant *RARB* rs116199914. In this regard, it could be speculated that the interaction described between this polymorphism and *NKIRAS1* modulates the expression of the latter, affecting the inhibition of NF-κB. This may trigger the regulation of genes encoding pro-inflammatory cytokines, adhesion molecules, chemokines, and the inducible nitric oxide synthase, thus contributing to endothelial damage and subsequently to CV disease. Similarly, since *RARB* rs116199914 was described to affect the sequence-specific binding of NFAT, it may be reasonable to consider that this phenomenon modulate the expression of genes related to angiogenic processes in atherosclerosis.

On the other hand, suggestive signals of potential relevance were observed when both the presence/absence of CV events (including IHD) and subclinical atherosclerosis were tested. However, those signals did not reach the genome-wide level of

significance, probably due to an insufficient statistical power to detect risk variants with a low to moderate effect. Consequently, biological pathway enrichment and protein-protein interaction analyses revealed a functional enrichment of the collagen biosynthesis network according to the presence/absence of carotid plaques. This result is consistent with the fact that collagen constitutes the main component of the fibrous cap of the carotid plaque, and contributes to its structural integrity and vulnerability (51). Indeed, a recent Metabochip performed in American patients with RA revealed a suggestive association between a genetic variant in the collagen type IV alpha 1 chain (*COL4A1*) gene and carotid plaques (52).

Finally, our study supports the implication of the previously reported candidate CV risk *locus NFKB1*, and suggests a potential influence of both *MSRA* and *ZC3HC1* in the development of CV disease in RA. On the contrary, a relevant influence of the HLA region in this process, as previously suggested, was not inferred from our data (8).

There is evidence that current CV risk screenings and management strategies underestimate the actual CV predisposition in patients with RA. In this context, genetic markers related to the development of CV disease underlying RA may be used as additional tools to identify those patients at high CV risk, who may definitively benefit from active therapy to prevent CV events. Accordingly, the results derived from our study may help to develop efficient predictive tools that could anticipate the development of CV disease in RA patients based on his/her genetic background.

It should be noted that the potential major limitation of our study is the lack of replication of our discovery findings in an independent cohort of patients with RA. In addition, our study could be underpowered to detect associations of small effect size.

Because of that, further studies aimed to confirm our results are needed. Interestingly, Karpouzas *et al.* described that there are more unstable, non-calcified plaques in patients with RA (53). Unstable plaques are very dangerous since they are particularly susceptible to disruption. Vulnerable plaques are generally characterized as those having a thin inflamed fibrous cap over a very large lipid core. Since the conventional carotid US technique performed in our study did not allow us to identify the presence of unstable plaques, we consider that further studies aimed to identify the potential role of the genetic variant *RARB* rs116199914 in the risk of unstable plaques should be conducted.

In summary, through a whole-genome screening of common genetic variation, we have identified *RARB* rs116199914 as the main genetic variant associated with cIMT values in patients with RA.

REFERENCES

1. Aviña-Zubieta JA, Choi HK, Sadatsafavi M, Etminan M, Esdaile JM, Lacaille D. Risk of cardiovascular mortality in patients with rheumatoid arthritis: a meta-analysis of observational studies. Arthritis Rheum 2008;59:1690-7.
2. Castañeda S, Nurmohamed MT, Gonzalez-Gay MA. Cardiovascular disease in inflammatory rheumatic diseases. Best Pract Res Clin Rheumatol 2016;30:851-69.
3. Gonzalez-Gay MA, Gonzalez-Juanatey C, Lopez-Diaz MJ, Piñeiro A, Garcia-Porrua C, Miranda-Filloy JA, et al. HLA-DRB1 and persistent chronic

inflammation contribute to cardiovascular events and cardiovascular mortality in patients with rheumatoid arthritis. Arthritis Rheum 2007;57:125-32.

4. Gonzalez-Gay MA, Gonzalez-Juanatey C, Martin J. Rheumatoid arthritis: a disease associated with accelerated atherogenesis. Semin Arthritis Rheum 2005;35:8-17.
5. Evans MR, Escalante A, Battafarano DF, Freeman GL, O'Leary DH, del Rincón I. Carotid atherosclerosis predicts incident acute coronary syndromes in rheumatoid arthritis. Arthritis Rheum 2011;63:1211-20.
6. Gonzalez-Juanatey C, Llorca J, Martin J, Gonzalez-Gay MA. Carotid intima-media thickness predicts the development of cardiovascular events in patients with rheumatoid arthritis. Semin Arthritis Rheum 2009;38:366-71.
7. Crowson CS, Rollefstad S, Ikdahl E, Kitas GD, van Riel PLCM, Gabriel SE, et al. Impact of risk factors associated with cardiovascular outcomes in patients with rheumatoid arthritis. Ann Rheum Dis 2018;77:48-54.
8. López-Mejías R, Castañeda S, González-Juanatey C, Corrales A, Ferraz-Amaro I, Genre F, et al. Cardiovascular risk assessment in patients with rheumatoid arthritis: The relevance of clinical, genetic and serological markers. Autoimmun Rev 2016;15:1013-30.
9. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008;9:356-69.
10. López-Mejías R, Carmona FD, Castañeda S, Genre F, Remuzgo-Martínez S, Sevilla-Perez B, et al. A genome-wide association study suggests the HLA Class

II region as the major susceptibility locus for IgA vasculitis. Sci Rep 2017;7:5088.

11. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376-81.
12. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62:2569-81.
13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.
14. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-7.
15. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48:1279-83.
16. Corrales A, González-Juanatey C, Peiró ME, Blanco R, Llorca J, González-Gay MA. Carotid ultrasound is useful for the cardiovascular risk stratification of patients with rheumatoid arthritis: results of a population-based study. Ann Rheum Dis 2014;73:722-7.
17. Corrales A, Parra JA, González-Juanatey C, Rueda-Gotor J, Blanco R, Llorca J, et al. Cardiovascular risk stratification in rheumatic diseases: carotid ultrasound is more sensitive than Coronary Artery Calcification Score to detect subclinical

atherosclerosis in patients with rheumatoid arthritis. Ann Rheum Dis 2013;72:1764-70.

18. Gonzalez-Juanatey C, Llorca J, Garcia-Porrua C, Martin J, Gonzalez-Gay MA. Effect of anti-tumor necrosis factor alpha therapy on the progression of subclinical atherosclerosis in severe rheumatoid arthritis. Arthritis Rheum 2006;55:150-3.
19. Naredo E, Möller I, Gutiérrez M, Bong DA, Cobo T, Corominas H, et al. Multi-examiner reliability of automated radio frequency-based ultrasound measurements of common carotid intima-media thickness in rheumatoid arthritis. Rheumatology (Oxford) 2011;50:1860-4.
20. Touboul PJ, Hennerici MG, Meairs S, Adams H, Amarenco P, Bornstein N, et al. Mannheim carotid intima-media thickness consensus (2004-2006). An update on behalf of the Advisory Board of the 3rd and 4th Watching the Risk Symposium, 13th and 15th European Stroke Conferences, Mannheim, Germany, 2004, and Brussels, Belgium, 2006. Cerebrovasc Dis 2007;23:75-80.
21. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 2006;38:209-13.
22. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012;22:1790-7.

23. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature 2015;526:68-74.
24. Schofield EC, Carver T, Achuthan P, Freire-Pritchett P, Spivakov M, Todd JA, et al. CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. Bioinformatics 2016;32:2511-3.
25. Leonard D, Svenungsson E, Dahlqvist J, Alexsson A, Ärlestig L, Taylor KE, et al. Novel gene variants associated with cardiovascular disease in systemic lupus erythematosus and rheumatoid arthritis. Ann Rheum Dis 2018;77:1063-9.
26. Carmona FD, Mackie SL, Martín JE, Taylor JC, Vaglio A, Eyre S, et al. A large-scale genetic analysis reveals a strong contribution of the HLA class II region to giant cell arteritis susceptibility. Am J Hum Genet 2015;96:565-80.
27. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet 2014;94:47-61.
28. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 2012;44:291-6.
29. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res 2015;43:D1049-56.
30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25-9.

31. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 2003;13:2129-41.
32. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.
33. Sokolove J, Brennan MJ, Sharpe O, Lahey LJ, Kao AH, Krishnan E, et al. Brief report: citrullination within the atherosclerotic plaque: a potential target for the anti-citrullinated protein antibody response in rheumatoid arthritis. Arthritis Rheum 2013;65:1719-24.
34. Mackey RH, Kuller LH, Deane KD, Walitt BT, Chang YF, Holers VM, et al. Rheumatoid Arthritis, Anti-CCP Positivity, and Cardiovascular Disease Risk in the Women's Health Initiative. Arthritis Rheumatol 2015;67:2311-22.
35. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell 2016;167:1369-84.
36. Dieguez-Gonzalez R, Akar S, Calaza M, Perez-Pampin E, Costas J, Torres M, et al. Genetic variation in the nuclear factor kappaB pathway in relation to susceptibility to rheumatoid arthritis. Ann Rheum Dis 2009;68:579-83.
37. Fenwick C, Na SY, Voll RE, Zhong H, Im SY, Lee JW, et al. A subclass of Ras proteins that regulate the degradation of IkappaB. Science 2000;287:869-73.
38. Baldwin AS. Control of oncogenesis and cancer therapy resistance by the transcription factor NF-kappaB. J Clin Invest 2001;107:241-6.

39. López-Mejías R, García-Bermúdez M, González-Juanatey C, Castañeda S, Miranda-Filloy JA, Gómez-Vaquero C, et al. NFKB1-94ATTG ins/del polymorphism (rs28362491) is associated with cardiovascular disease in patients with rheumatoid arthritis. Atherosclerosis 2012;224:426-9.

40. Barnabe C, Martin BJ, Ghali WA. Systematic review and meta-analysis: anti-tumor necrosis factor alpha therapy and cardiovascular events in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2011;63:522-9.

41. Greenberg JD, Kremer JM, Curtis JR, Hochberg MC, Reed G, Tsao P, et al. Tumour necrosis factor antagonist use and associated risk reduction of cardiovascular events among patients with rheumatoid arthritis. Ann Rheum Dis 2011;70:576-82.

42. Hürlimann D, Forster A, Noll G, Enseleit F, Chenevard R, Distler O, et al. Anti-tumor necrosis factor-alpha treatment improves endothelial function in patients with rheumatoid arthritis. Circulation 2002;106:2184-7.

43. Mäki-Petäjä KM, Hall FC, Booth AD, Wallace SM, Yasmin, Bearcroft PW, et al. Rheumatoid arthritis is associated with increased aortic pulse-wave velocity, which is reduced by anti-tumor necrosis factor-alpha therapy. Circulation 2006;114:1185-92.

44. Macian F. NFAT proteins: key regulators of T-cell development and function. Nat Rev Immunol 2005;5:472-84.

45. Rao A, Luo C, Hogan PG. Transcription factors of the NFAT family: regulation and function. Annu Rev Immunol 1997;15:707-47.

46. Shaw JP, Utz PJ, Durand DB, Toole JJ, Emmel EA, Crabtree GR. Identification of a putative regulator of early T cell activation genes. Science. 1988. 241: 202-205. J Immunol 2010;185:4972-5.

47. Crabtree GR, Olson EN. NFAT signaling: choreographing the social lives of cells. Cell 2002;109 Suppl:S67-79.

48. Graef IA, Chen F, Crabtree GR. NFAT signaling in vertebrate development. Curr Opin Genet Dev 2001;11:505-12.

49. Hogan PG, Chen L, Nardone J, Rao A. Transcriptional regulation by calcium, calcineurin, and NFAT. Genes Dev 2003;17:2205-32.

50. López-Mejías R, Corrales A, Genre F, Hernández JL, Ochoa R, Blanco R, et al. Angiopoietin-2 serum levels correlate with severity, early onset and cardiovascular disease in patients with rheumatoid arthritis. Clin Exp Rheumatol 2013;31:761-6.

51. Ross R. Atherosclerosis--an inflammatory disease. N Engl J Med 1999;340:115-26.

52. Arya R, Escalante A, Farook VS, Restrepo JF, Battafarano DF, Almeida M, et al. A genetic association study of carotid intima-media thickness (CIMT) and plaque in Mexican Americans and European Americans with rheumatoid arthritis. Atherosclerosis 2018;271:92-101.

53. Karpouzas GA, Malpeso J, Choi TY, Li D, Munoz S, Budoff MJ. Prevalence, extent and composition of coronary plaque in patients with rheumatoid arthritis without symptoms or prior diagnosis of coronary artery disease. Ann Rheum Dis 2014;73:1797-804.

FIGURE LEGENDS

Figure 1. Manhattan plot representation of the analysis of carotid intima-media thickness values as cardiovascular disease outcome. The -log10 of the P-values are plotted against their physical chromosomal position. The red line represents the genome-wide level of significance (P<5E-08).

Figure 2. Regulatory chromatin annotations of *RARB* rs116199914 in tissues related to cardiovascular pathology and cells of the immune system according to the Encyclopedia of DNA Elements (ENCODE) data. The chromatin 15-state model was developed using 5 marks and 127 epigenomes from the Roadmap Epigenomics Project.

Figure 3. Interaction network formed by the encoded proteins of genes showing P-values of potential relevance in our study (P<1E-04) according to the presence/absence of carotid plaques. The width of the blue lines indicates the reliability of each interaction. Proteins of the collagen biosynthetic process pathway (GO:0032964) are highlighted in red.

Table 1. Demographic, clinical, and CV disease-related characteristics of the 2,989 RA patients whose samples were included in the filtered dataset.

Variables	[% (n/N)]
Age at the time of disease onset [years, mean ± SD]	49.8 ± 14.9
Follow-up time [years, mean ± SD]	11.7 ± 9.1
Women [%]	74.7
RF positive*	65.2 (1,585/2,432)
Anti-CCP antibodies positive	59.7 (1,365/2,286)
Erosions	52.4 (1,125/2,148)
Extra-articular manifestations**	28.8 (575/1,994)
Traditional CV risk factors	
Hypertension	39.4 (1,018/2,585)
Diabetes mellitus	12.3 (318/2,585)
Dyslipidemia	43.4 (1,122/2,585)
Obesity	23.4 (605/2,585)
Smoking habit	37.0 (957/2,585)
CV events	15.6 (467/2,989)
Ischaemic heart disease	7.5 (224/2,989)
Heart failure	4.9 (146/2,989)
Cerebrovascular accident	4.2 (125/2,989)
Peripheral arterial disease	2.0 (60/2,989)

CV: cardiovascular; RA: rheumatoid arthritis; SD: standard deviation; RF: rheumatoid factor; Anti-CCP antibodies: anti-cyclic citrullinated peptide antibodies.

*At least two determinations at different times were required for analysis of this result.

**Extra-articular manifestations of the disease (if patients with RA experienced at least one of the following manifestations: nodular disease, Felty's syndrome, pulmonary fibrosis, rheumatoid vasculitis, or secondary Sjögren's syndrome) (3).

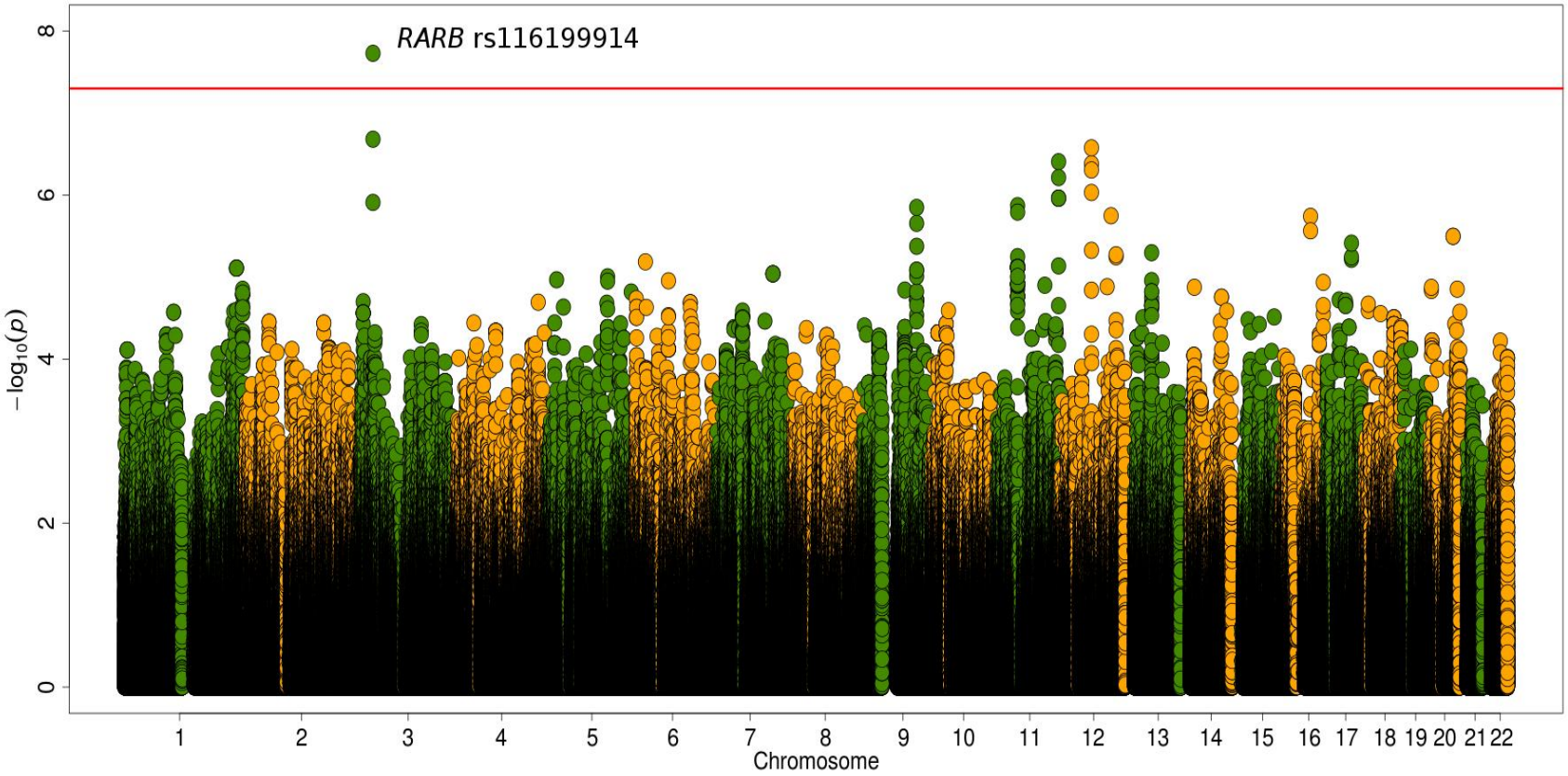
Table 2. Index signals showing the lowest P-values according to the different CV disease outcomes.

CHR	Position in CHR (GRCh37)	SNP ID	GENCODE Gene	Change	Minor allele	MAF	P	OR [CI 95%]	CV disease outcome
3	25.638.355	rs116199914	*RARB* (3´UTR)	G<A	G	0.012	**1.86E-08**	0.142 (0.025)*	cIMT values
3	25.622.694	rs77388418	*RARB* (intronic)	C<T	C	0.014	2.07E-07	0.124 (0.024)*	cIMT values
12	63.337.536	rs1695024	8kb 3' of Y_RNA	A<G	A	0.230	2.64E-07	0.031 (0.006)*	cIMT values
11	129.852.180	rs111703287	*PRDM10* (intronic)	T<C	T	0.014	3.90E-07	0.119 (0.023)*	cIMT values
1	166.485.891	rs6684311	27kb 5' of RP11-276E17.2	G<C	G	0.189	2.85E-07	1.68 [1.38-2.05]	CV events
1	245.338.976	rs112844193	*KIF26B* (intronic)	T<C	T	0.054	1.35E-07	2.67 [1.85-3.85]	IHD
7	120.966.790	rs3779381	*WNT16* (intronic)	G<A	G	0.283	2.09E-07	1.77 [1.23-2.19]	IHD
1	156.057.417	rs112941217	*LMNA (intronic)*	C<T	C	0.030	4.67E-07	4.81 [2.61-8.87]	IHD
17	15.008.430	rs8066891	123bp 3' of RP11-924A14.1	G<A	G	0.171	4.47E-06	0.58 [0.46-0.73]	Carotid plaques
9	29.148.449	rs12683261	259kb 5' of MIR873	A<G	A	0.031	4.57E-06	0.25 [0.14-0.45]	Carotid plaques
1	240.599.906	rs9727451	*FMN2* (intronic)	A<G	A	0.087	4.69E-06	2.13 [1.54-2.95]	Carotid plaques
4	166.579.647	rs2611206	26kb 5' of RP11-340B18.1	A<G	A	0.126	4.84E-06	0.53 [0.41-0.69]	Carotid plaques

CV: cardiovascular; CHR: chromosome; SNP: single nucleotide polymorphism; MAF: minor allele frequency; OR: odds ratio; CI: confidence interval, UTR: untranslated region; cIMT: carotid intima-media thickness; IHD: ischaemic heart disease.

The statistically significant P-value at the genome-wide level of significance is highlighted in **bold**.

*These results are expressed as β (SE).



Tissue / cell type	Roadmap core 15-state model (enhancers)	Enhancer mark H3K4me1	Enhancer mark H3K27ac	Promoter mark H3K9ac
Fetal heart				
Monocytes-CD14+ RO01746 primary cells				
Primary hematopoietic stem cells G-CSF-mobilized female				
Primary mononuclear cells from peripheral blood				
Pancreatic Islets				
Primary T CD8+ naive cells from peripheral blood				
Primary T helper cells from peripheral blood				
Primary T helper cells PMA-I stimulated				
Primary T helper naive cells from peripheral blood				
Primary T regulatory cells from peripheral blood