Assessment of Osteoarthritis Candidate Genes in a Meta-Analysis of Nine Genome-Wide Association Studies


Objective. To assess candidate genes for association with osteoarthritis (OA) and identify promising genetic factors and, secondarily, to assess the candidate gene approach in OA.

Methods. A total of 199 candidate genes for association with OA were identified using Human Genome Epidemiology (HuGE) Navigator. All of their single-nucleotide polymorphisms (SNPs) with an allele frequency of >5% were assessed by fixed-effects meta-analysis of 9 genome-wide association studies (GWAS) that included 5,636 patients with knee OA and 16,972 control subjects and 4,349 patients with hip OA and 17,836 control subjects of European ancestry. An additional 5,921 individuals were genotyped for significantly associated SNPs in the meta-analysis. After correction for the number of independent tests, P values less than $1.58 \times 10^{-5}$ were considered significant.

Results. SNPs at only 2 of the 199 candidate genes ($COL11A1$ and $VEGF$) were associated with OA in the meta-analysis. Two SNPs in $COL11A1$ showed association with hip OA in the combined analysis: rs4907986 ($P = 1.29 \times 10^{-5}$, odds ratio [OR] 1.12, 95% confidence interval [95% CI] 1.06–1.17) and rs1241164 ($P = 1.47 \times 10^{-5}$, OR 0.82, 95% CI 0.74–0.89). The sex-stratified analysis also showed association of $COL11A1$ SNP rs4908291 in women ($P = 1.29 \times 10^{-5}$, OR 0.87, 95% CI 0.82–0.92); this SNP showed linkage disequilibrium with rs4907986. A single SNP of $VEGF$, rs833058, showed association with hip OA in men ($P = 1.35 \times 10^{-5}$, OR 0.85, 95% CI 0.79–0.91). After additional samples were genotyped, association at one of the $COL11A1$ signals was reinforced, whereas association at $VEGF$ was slightly weakened.

Conclusion. Two candidate genes, $COL11A1$ and $VEGF$, were significantly associated with OA in this focused meta-analysis. The remaining candidate genes were not associated.
The etiology of primary osteoarthritis (OA) is multifactorial and includes aging and mechanical, hormonal, and genetic factors (1). Investigations of many of these factors have produced contradictory results, mak-

supported by Leiden University Medical Centre, the Dutch Arthritis Association, and Pfizer Inc., Groton, Connecticut. The genotyping was supported by the NWO (MW 904-61-095, 911-03-016, 917-66-344, and 911-03-012), Leiden University Medical Center, and by the Center of Medical System Biology and the Netherlands Consortium of Healthy Aging in the framework of the Netherlands Genomics Initiative. Additional funding was received from the Dutch Arthritis Association (DAA 2010_017) and the European Union Seventh Framework Programme (FP7/2007-2011, under grant agreement 259679). The arcOGEN Phase I study (http://www.arcogen.org.uk/) was funded by Arthritis Research UK (special purpose grant 18030). The Estonian Genome Center of the University of Tartu (EGCUT) received targeted funding from the Estonian government (SF0180142S08), through the Centre of Excellence in Genomics (EXCEGEN) and University of Tartu (SP1VGARENG).

Cristina Rodriguez-Fontenla, MSc, Manuel Calaza, PhD, Antonio Gonzalez, MD, PhD: Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain; Evangelos Evangelou, BSc, PhD, Evangelia E. Ntzani, MD, PhD: University of Ioannina School of Medicine, Ioannina, Greece, and King’s College London, London, UK; Ana M. Valdes, PhD, Guangji Zhai, PhD, Timothy D. Spector, MD, MSc, FRCP: King’s College London, London, UK; Nigel Arden, MD, Andrew Carr, MD, Kay Chapman, PhD: University of Oxford, Oxford, UK; Francisco J. Blanco, MD, Natividad Oreiro, MD: Instituto de Investigación Biomédica–Complejo Hospitalario Universitario A Coruña, A Coruña, Spain; Panos Deloukas, PhD, Kalliopi Panoutsopoulou, PhD, Eleftheria Zeggini, PhD: Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; Michael Doherty, MA, MD, FRCP: Nottingham City Hospital, Nottingham, UK; Tonu Esko, PhD: Andres Metspalu, MD, PhD: University of Tartu, Tartu, Estonia; Carlos M. García Aleta, MD, Jose A. Riancho, MD: University of Cantabria, Santander, Spain; Juan J. Gomez-Reino Carnota, MD, PhD: Hospital Clínico Universitario de Santiago and University of Santiago de Compostela, Santiago de Compostela, Spain; Hafidis Helgadottir, BSc, Ingileif Jonsdottir, PhD, Unnur Thorsteinsdottir, PhD, Gudmar Thorleifsson, PhD, Kari Stefansson, MD, PhD: deCODE Genes, Reykjavik, Iceland; Albert Hofman, MD, PhD, Hanneke J. M. Kerkhof, MD, PhD (current address: Pfizer, Ijssel, The Netherlands), Fernando Rivadeneira, MD, PhD, André G. Uitterlind, PhD, Joyce B. van Meurs, PhD: Erasmus Medical Center and The Netherlands Genomics Initiative–Sponsored Netherlands Consortium for Healthy Aging, Rotterdam, The Netherlands; Margreet Kloppenburg, MD, PhD, Yolande F. Ramos, BSc: Leiden University Medical Centre, Leiden, The Netherlands; Andrew McCaskie, MD, John Loughlin, PhD: Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK; William E. R. Ollier, PhD: Centre for Integrated Genomic Medical Research, University of Manchester, Manchester, UK; Stuart H. Ralston, MD: Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK; P. Eline Slagboom, PhD, Ingrid Meulenbelt, PhD: Leiden University Medical Center, Leiden, The Netherlands, and The Netherlands Genomics Initiative–Sponsored Netherlands Consortium for Healthy Aging, Rotterdam, The Netherlands; Aspasia Tsou, PhD: University of Thessaly, Larissa, Greece; Gillian A. Wallis, PhD: Wellcome Trust Centre for Cell Matrix Research, University of Manchester, Manchester, UK; Mark Wilkinson, PhD, FRCS: University of Sheffield, Sheffield, UK; Yanan Zhu, PhD, MPH, David T. Felson, MD, PhD: Boston University School of Medicine, Boston, MA; John P. A. Ioannidis, MD, PhD: University of Ioannina School of Medicine, Ioannina, Greece, and Stanford University School of Medicine, Stanford, CA. The results of most of these association studies have been inconclusive. The association of only one candidate gene, GDF5, was well supported by data from multiple studies and reached genome-wide significance ($P < 5 \times 10^{-8}$) in subsequent replication analysis (6,7). Linkage studies have provided some consistent results, but progression from these results to identification of the causal genes has proven to be difficult (2,5).

Recently, GWAS have identified 11 additional OA susceptibility loci with genome-wide significance levels. Two of them, in DWWA/COL6A4 and a region containing HLA class II/III genes, showed association in Asians but not in Europeans (8–10). The other 9 loci reached genome-wide significance in Europeans. They include a locus on chromosome 7q22 (which is located in a large linkage disequilibrium [LD] block that contains 6 genes) associated with knee OA (11,12); MCF2L associated with knee and hip OA (13); 5 loci that were identified in the arcOGEN GWAS (GNL3/GLT8D1 associated with knee and hip OA, ASTM2 associated with severe hip OA in women, FILIP1–SENP6 and PTHLH associated with hip OA, and CHTS11 associated with severe hip OA) (14); and DOTIL associated with joint space width of the hip (15) and with hip OA in men (16). Approximately 8 more loci are near this level of association (2,5,14).

All of the studies reaching genome-wide signifi-

Dr. Blanco has received consulting fees, speaking fees, and/or honoraria from Gebro Pharma, Bioiberica Farma, and Pfizer (less than $10,000 each). Dr. Gonzalez has received consulting fees, speaking fees, and/or honoraria from Bioiberica Farma (less than $10,000).

Address correspondence to Antonio Gonzalez, MD, PhD, Laboratorio Investigacion 10, Instituto de Investigacion Sanitaria–Hospital Clinico Universitario de Santiago, Edificio de Consultas, Planta 2, Travesia de Choupana s/n, 15706 Santiago de Compostela, Spain. E-mail: antonio.gonzalez.martinez-pedrayo@sergas.es.

Submitted for publication March 5, 2013; accepted in revised form November 26, 2013.
cance have used meta-analysis of data from multiple sample collections. This is a very efficient approach to increase power. In addition, it is also very useful for discovery of new associations when applied to GWAS, because each study provides information for most single-nucleotide polymorphisms (SNPs) in the genome, either directly or through imputation, and therefore add to the overall result (17). Another way to favor discovery of new loci is by focusing analysis on particular subsets of genes for which there is prior supporting evidence, thereby increasing the prior probability of association and reducing the burden of multiplicity and thus the stringency required for claiming association (18,19).

The aim of this study was to identify new OA genetic factors, using meta-analysis of GWAS and focused analysis of OA candidate genes. A secondary aim was to assess the validity of the candidate gene approach in OA. To this end, we performed a meta-analysis of 9 GWAS that included patients with knee or hip OA and control subjects of European descent: the deCODE study from Iceland (20), 3 collections from the Rotterdam Study (Rotterdam Study I, II, and III) (21), the Genetics Osteoarthritis and Progression (GARP) collection from The Netherlands (22), arcOGEN phase I (23) and TwinsUK (24) from the UK, the Framingham Osteoarthritis Study from the US (25), and EGCUT (Estonian Genome Center of the University of Tartu) from Estonia (26). Additional sample collections not involved in the GWAS were used for an extension study of significant results (Table 1). They included collections from the north of Spain (27–29), the center of Greece (30), and the Nottingham (31) and Genetics in Osteoarthritis and Lifestyle (GOAL) (32) studies from the UK.

All of these sample collections have been described in detail previously. Briefly, the deCODE study included patients with knee OA or hip OA and control subjects of European descent. In addition, it is also very useful for discovery of new associations when applied to GWAS, because each study provides information for most single-nucleotide polymorphisms (SNPs) in the genome, either directly or through imputation, and therefore add to the overall result (17). Another way to favor discovery of new loci is by focusing analysis on particular subsets of genes for which there is prior supporting evidence, thereby increasing the prior probability of association and reducing the burden of multiplicity and thus the stringency required for claiming association (18,19).

The aim of this study was to identify new OA genetic factors, using meta-analysis of GWAS and focused analysis of OA candidate genes. A secondary aim was to assess the validity of the candidate gene approach in OA. To this end, we performed a meta-analysis of 9 GWAS that included patients with knee or hip OA and control subjects of European descent: the deCODE study from Iceland (20), 3 collections from the Rotterdam Study (Rotterdam Study I, II, and III) (21), the Genetics Osteoarthritis and Progression (GARP) collection from The Netherlands (22), arcOGEN phase I (23) and TwinsUK (24) from the UK, the Framingham Osteoarthritis Study from the US (25), and EGCUT (Estonian Genome Center of the University of Tartu) from Estonia (26). Additional sample collections not involved in the GWAS were used for an extension study of significant results (Table 1). They included collections from the north of Spain (27–29), the center of Greece (30), and the Nottingham (31) and Genetics in Osteoarthritis and Lifestyle (GOAL) (32) studies from the UK.

All of these sample collections have been described in detail previously. Briefly, the deCODE study included patients with knee OA or hip OA and control subjects of European descent. In addition, it is also very useful for discovery of new associations when applied to GWAS, because each study provides information for most single-nucleotide polymorphisms (SNPs) in the genome, either directly or through imputation, and therefore add to the overall result (17). Another way to favor discovery of new loci is by focusing analysis on particular subsets of genes for which there is prior supporting evidence, thereby increasing the prior probability of association and reducing the burden of multiplicity and thus the stringency required for claiming association (18,19).

### PATIENTS AND METHODS

#### Sample collections

We performed a meta-analysis of 9 GWAS that included patients with knee or hip OA and control subjects of European descent: the deCODE study from Iceland (20), 3 collections from the Rotterdam Study (Rotterdam Study I, II, and III) (21), the Genetics Osteoarthritis and Progression (GARP) collection from The Netherlands (22), arcOGEN phase I (23) and TwinsUK (24) from the UK, the Framingham Osteoarthritis Study from the US (25), and EGCUT (Estonian Genome Center of the University of Tartu) from Estonia (26). Additional sample collections not involved in the GWAS were used for an extension study of significant results (Table 1). They included collections from the north of Spain (27–29), the center of Greece (30), and the Nottingham (31) and Genetics in Osteoarthritis and Lifestyle (GOAL) (32) studies from the UK.

All of these sample collections have been described in detail previously. Briefly, the deCODE study included patients with knee OA or hip OA and control subjects of European descent. In addition, it is also very useful for discovery of new associations when applied to GWAS, because each study provides information for most single-nucleotide polymorphisms (SNPs) in the genome, either directly or through imputation, and therefore add to the overall result (17). Another way to favor discovery of new loci is by focusing analysis on particular subsets of genes for which there is prior supporting evidence, thereby increasing the prior probability of association and reducing the burden of multiplicity and thus the stringency required for claiming association (18,19).

The aim of this study was to identify new OA genetic factors, using meta-analysis of GWAS and focused analysis of OA candidate genes. A secondary aim was to assess the validity of the candidate gene approach in OA. To this end, we performed a meta-analysis of 9 GWAS that included patients with knee or hip OA and control subjects of European descent: the deCODE study from Iceland (20), 3 collections from the Rotterdam Study (Rotterdam Study I, II, and III) (21), the Genetics Osteoarthritis and Progression (GARP) collection from The Netherlands (22), arcOGEN phase I (23) and TwinsUK (24) from the UK, the Framingham Osteoarthritis Study from the US (25), and EGCUT (Estonian Genome Center of the University of Tartu) from Estonia (26). Additional sample collections not involved in the GWAS were used for an extension study of significant results (Table 1). They included collections from the north of Spain (27–29), the center of Greece (30), and the Nottingham (31) and Genetics in Osteoarthritis and Lifestyle (GOAL) (32) studies from the UK.

All of these sample collections have been described in detail previously. Briefly, the deCODE study included patients with knee OA or hip OA and control subjects of European descent. In addition, it is also very useful for discovery of new associations when applied to GWAS, because each study provides information for most single-nucleotide polymorphisms (SNPs) in the genome, either directly or through imputation, and therefore add to the overall result (17). Another way to favor discovery of new loci is by focusing analysis on particular subsets of genes for which there is prior supporting evidence, thereby increasing the prior probability of association and reducing the burden of multiplicity and thus the stringency required for claiming association (18,19).

### Table 1. Characteristics of the sample collections included in the GWAS meta-analysis and the extension study

<table>
<thead>
<tr>
<th></th>
<th>Knee OA</th>
<th>Hip OA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>Controls</td>
<td></td>
</tr>
<tr>
<td>OA definition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TJR and radiographic</td>
<td>4,096</td>
<td>4,896</td>
</tr>
<tr>
<td>TJR and clinical</td>
<td>2,318</td>
<td>4,896</td>
</tr>
</tbody>
</table>

**PATIENTS AND METHODS**

**Sample collections.** We performed a meta-analysis of 9 GWAS that included patients with knee or hip OA and control subjects of European descent: the deCODE study from Iceland (20), 3 collections from the Rotterdam Study (Rotterdam Study I, II, and III) (21), the Genetics Osteoarthritis and Progression (GARP) collection from The Netherlands (22), arcOGEN phase I (23) and TwinsUK (24) from the UK, the Framingham Osteoarthritis Study from the US (25), and EGCUT (Estonian Genome Center of the University of Tartu) from Estonia (26). Additional sample collections not involved in the GWAS were used for an extension study of significant results (Table 1). They included collections from the north of Spain (27–29), the center of Greece (30), and the Nottingham (31) and Genetics in Osteoarthritis and Lifestyle (GOAL) (32) studies from the UK.

All of these sample collections have been described in detail previously. Briefly, the deCODE study included patients with knee OA or hip OA and control subjects of European descent. In addition, it is also very useful for discovery of new associations when applied to GWAS, because each study provides information for most single-nucleotide polymorphisms (SNPs) in the genome, either directly or through imputation, and therefore add to the overall result (17). Another way to favor discovery of new loci is by focusing analysis on particular subsets of genes for which there is prior supporting evidence, thereby increasing the prior probability of association and reducing the burden of multiplicity and thus the stringency required for claiming association (18,19).

### Table 1. Characteristics of the sample collections included in the GWAS meta-analysis and the extension study

<table>
<thead>
<tr>
<th>Knee OA</th>
<th>Hip OA</th>
<th>OA definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TJR and radiographic</td>
<td>4,096</td>
<td>4,896</td>
</tr>
<tr>
<td>TJR and clinical</td>
<td>2,318</td>
<td>4,896</td>
</tr>
</tbody>
</table>

**PATIENTS AND METHODS**

**Sample collections.** We performed a meta-analysis of 9 GWAS that included patients with knee or hip OA and control subjects of European descent: the deCODE study from Iceland (20), 3 collections from the Rotterdam Study (Rotterdam Study I, II, and III) (21), the Genetics Osteoarthritis and Progression (GARP) collection from The Netherlands (22), arcOGEN phase I (23) and TwinsUK (24) from the UK, the Framingham Osteoarthritis Study from the US (25), and EGCUT (Estonian Genome Center of the University of Tartu) from Estonia (26). Additional sample collections not involved in the GWAS were used for an extension study of significant results (Table 1). They included collections from the north of Spain (27–29), the center of Greece (30), and the Nottingham (31) and Genetics in Osteoarthritis and Lifestyle (GOAL) (32) studies from the UK.

All of these sample collections have been described in detail previously. Briefly, the deCODE study included patients with knee OA or hip OA and control subjects of European descent. In addition, it is also very useful for discovery of new associations when applied to GWAS, because each study provides information for most single-nucleotide polymorphisms (SNPs) in the genome, either directly or through imputation, and therefore add to the overall result (17). Another way to favor discovery of new loci is by focusing analysis on particular subsets of genes for which there is prior supporting evidence, thereby increasing the prior probability of association and reducing the burden of multiplicity and thus the stringency required for claiming association (18,19).
at ≥2 joint sites (22). They were compared with population controls. The areOGEN phase 1 cohort included cases of knee or hip OA, as determined by radiographic evidence of disease or clinical evidence of disease “to a level requiring joint replacement” (23); controls were derived from an early release of the Wellcome Trust Case Control Consortium 2 data. EG CUT included cases of radiographically confirmed OA (K/L score of >2) and controls that were free of any OA symptoms. All sample collections used for the extension study were derived from case–control studies. Patients were ascertained on the basis of having undergone hip joint replacement due to symptomatic and radiographically confirmed hip OA, except in GOAL, in which severe symptomatic hip OA was the selection criterion. All of the sample collections and the genetic studies received approval by the relevant ethics committees, and the samples were obtained with the written informed consent of the participants.

Gene and SNP selection. We used the Phenopedia tool of the Human Genome Epidemiology (HuGE) Navigator (35) to identify candidate genes that have been studied for their possible association with OA, without additional filtering. The query terms were as follows: osteoarthritis, spinal osteophytosis, and intervertebral disk displacement. The candidate genes derived from the spinal osteophytosis and intervertebral disk displacement lists showed a large degree of overlap with those from the osteoarthritis list (90% and 72%, respectively). The 17 nonoverlapping genes were considered to be of interest for the discovery of new OA loci. This database covers genetic studies published since 2000. Genes in chromosome X were excluded, because it is impossible to impute the genotypes needed for meta-analysis across different GWAS designs. FRA1H1H was excluded because the bibliographic reference was incorrect. All of the genes with genome-wise significance in Europeans (P < 5 × 10^-8) were also excluded. Duplicates were removed.

Map positions of loci encompassing the candidate genes and 50 kb downstream of their stop codon and upstream of their start codon were obtained from the Ensembl database. Overlapping loci were fused as a single locus. All SNPs with a minor allele frequency (MAF) of >5% in the CEU data set corresponding to the candidate gene loci were retrieved from HapMap (phases 1, 2, and 3; release 27) with in-house Perl programs interacting with the HapMart server. All SNPs were aligned according to the positive strand, to avoid ambiguities.

Genotyping and imputation of untyped SNPs. Genotyping technologies for the GWAS included in the meta-analysis were different and have been previously described in detail (11,20,23,25,36). Imputation of untyped SNPs was performed based on CEU data from HapMap (phases 1 and 2; release 22). Summary information on genotyping and imputation is shown in Supplementary Table 1 (available on the Arthritis & Rheumatology web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38300/abstract). Genotyping of the extension study collections was performed at Hospital Clínico Universitario de Santiago, using single-base extension with a SNaPshot Multiplex Kit (Applied Biosystems) (Spanish and Greek collections) or was carried out by KBioscience, using a KASPar system (UK collections).

Statistical analysis. Each team contributing a GWAS performed association testing for knee OA and hip OA under a per-allele model. The lambda inflation factor was calculated per sex-specific effect size using the genomic control method (37), and the SEs were corrected by the square root of the lambda inflation factor (SEcorrected = SEobserved × \sqrt{λ}). Robust SEs were estimated to adjust for family relationships (DECODE and GARP studies). For meta-analysis, the effect size for each SNP (odds ratio [OR] per copy of minor allele as per HapMap) was calculated using fixed-effects inverse variance models, synthesizing all effect sizes and the corrected SEs.

Heterogeneity was assessed with the I^2 statistic; when low or moderate heterogeneity was observed, no random-effects meta-analysis was performed (38). Meta-analysis of the GWAS was performed using METAL software (39), considering 6 strata with 2 joint levels (knee and hip) and 3 sex levels (all, women, and men). Two research centers (Ioannina, Greece and Erasmus MC Rotterdam, The Netherlands) performed both the quality control and meta-analyses for the whole GWAS. A quality control protocol was set up that included validation of the results file format, reports for range of values, and elimination of potential biases (i.e., extremely large beta values or SEs). Files were cross-validated between the 2 research centers after quality control and meta-analyses to check for inconsistencies. SNPs with a MAF of <1%, imputation quality of <0.30 (MACH program) or <0.40 (IMPUTE program) and beta values of >4 or <-4, and SNPs that were not available in >4 studies were excluded from further analysis.

The significance threshold for claiming association was determined considering the number of independent tests performed. This number was estimated using a modification of the simpleM algorithm (40) applied to the genotypes of the CEU collection in HapMap. The modification consisted of replacing the observed correlation matrix for the nearest positive semidefinite matrix, as implemented in the R software package corpcor (http://www.R-project.org), to correct for biases introduced by missing genotypes. It was estimated that the number of independent tests represented by these SNPs was 3,156 (see Supplementary Table 2, available on the Arthritis & Rheumatology web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38300/abstract). The number of independent tests was used to define a significance threshold of P = 1.58 × 10^-3, according to Bonferroni multiplicity correction. No additional correction was performed for stratification according to joint and sex, because there is known heterogeneity across these strata in OA genetics (2–5) and, therefore, no correction of this type is used in OA genetic studies (6,8,9,11,13,14,16).

Results of the extension study were combined using a Mantel-Haenszel approach (40). Combination of the extension study data with the GWAS data was done using a fixed-effects model with R software (http://www.R-project.org). Power estimates were obtained with Power and Sample Size software (42). A full analysis of power is shown in Supplementary Figure 1 (available on the Arthritis & Rheumatology web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38300/abstract). For example, the power to detect association with a SNP of MAF 20% and OR 1.15 for knee OA or 1.16 for hip OA was 80%, assuming no heterogeneity between the GWAS.

RESULTS

Systematic identification of candidate genes. A total of 199 genes (Figure 1) have been investigated for
their association with OA in humans, according to the HuGe Navigator (35). The HuGe Navigator included 542 bibliographic references for the genetic studies of the candidate genes. Some of the genes are from loci that were associated with OA for the first time in GWAS (HLA class II/III, A2BP1, and LCRH1) or in genome-wide linkage studies (MATN3, DIO2, FRZB, and BMP5). These genes were included because the identification of many of them as putative susceptibility genes partially came from post hoc analyses of their potential biologic role, and because none has reached a genome-wide significant association in Europeans. The 199 genes were grouped in 158 nonoverlapping genome segments that contained 27,501 autosomal SNPs (MAF of >5%) with known genotypes in the CEU (Utah residents with ancestry from northern and western Europe) population of HapMap.

**Meta-analysis of association of all SNPs in the candidate genes.** The effect sizes for each of the 9 GWAS corresponding to SNPs in the candidate genes were obtained and combined in a meta-analysis. Genotypes were available for 25,839 of the 27,501 SNPs included in candidate genes after applying quality control filters (association results are available from the corresponding author). These genotypes had been directly typed or imputed. No significant associations were observed in the knee OA meta-analysis.

Meta-analysis of hip OA GWAS showed significant association at 2 candidate genes. Two SNPs, rs4907986 and rs1241164, in the 5′ and 3′ ends of COL11A1, respectively, were associated in the unstratified analysis (Table 2 and Figure 2A). They comprised 2 independent associations, as manifested by the low pairwise correlation coefficient (r² = 0.09) between them. A SNP of COL11A1, rs2615977, was highlighted in the previously reported analysis of the arcOGEN phase I study (with P < 1.1 × 10⁻⁵) (23), which overlaps with the current meta-analysis. However, none of the 2 independent top associated SNPs showed strong

![Figure 1. Osteoarthritis candidate genes selected from Human Genome Epidemiology Navigator.](image)

<p>| Table 2. SNPs independently associated with hip osteoarthritis in all samples and in samples stratified by sex* |
|---------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|</p>
<table>
<thead>
<tr>
<th>Sample</th>
<th>Chr.</th>
<th>Position</th>
<th>Alleles</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>I²</th>
<th>P for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>All subjects COL11A1</td>
<td>rs4907986</td>
<td>1</td>
<td>103322221</td>
<td>T/C</td>
<td>1.12 (1.06–1.17)</td>
<td>1.29 × 10⁻⁵</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>rs1241164</td>
<td>1</td>
<td>103129065</td>
<td>T/C</td>
<td>0.82 (0.74–0.89)</td>
<td>1.47 × 10⁻⁵</td>
<td>0.92</td>
</tr>
<tr>
<td>Women COL11A1</td>
<td>rs4908291</td>
<td>1</td>
<td>103345324</td>
<td>A/T</td>
<td>0.87 (0.82–0.92)</td>
<td>1.29 × 10⁻⁵</td>
<td>33.2</td>
</tr>
<tr>
<td>Men VEGF</td>
<td>rs833058</td>
<td>6</td>
<td>43839832</td>
<td>T/C</td>
<td>0.85 (0.79–0.91)</td>
<td>1.35 × 10⁻⁵</td>
<td>18.6</td>
</tr>
</tbody>
</table>

* Single-nucleotide polymorphisms (SNPs) below the threshold of significance (P = 1.58 × 10⁻⁵) are shown. The odds ratios (ORs) and 95% confidence intervals (95% CIs) are relative to the first listed allele at each SNP. Chr. = chromosome.
correlation with rs2615977 (pairwise \( r^2 \approx 0.4 \)), and rs2615977 was not among the most-associated SNPs in the meta-analysis. An analysis stratified by sex also showed association of \( COL11A1 \) SNP rs4907986 in women. This SNP showed LD with rs4907986, with white diamonds representing maximal correlation, red diamonds representing maximal correlation, and orange diamonds representing intermediate correlation.

The second candidate gene associated with hip OA was \( VEGF \). Only a single SNP, rs833058, reached association over the required threshold in men (Table 2). No other SNP with strong or modest \( (r^2 > 0.5) \) LD with rs833058 was observed in the meta-analysis (Figure 2B). None of the other 197 candidate genes showed association with knee or with hip OA at the requested level of significance.

**Extension and summary studies.** We attempted to further establish the association of hip OA with 3 SNPs, 2 from \( COL11A1 \) representing the top SNPs of the 2 independent associations in the combined analysis, and the top SNP in \( VEGF \). This part of the study was not intended as a replication of the results due to the relatively small number of independent samples that were available.

For the \( COL11A1 \) SNPs, 1,929 samples from Spanish and Greek individuals (787 patients with hip OA and 1,142 control subjects) were studied. No significant association was observed in this underpowered analysis; however, analysis of additional samples showed independent association (Kerkhof HJM, van Meurs JB; unpublished observations). In the combined analysis, the significance of rs1241164 association was clearer \((P = 5.3 \times 10^{-6})\) than before, given that the direction of change and effect size were similar in the meta-analysis and the extension analysis (Table 3). The second signal, corresponding to rs4907986, was slightly weakened in the combined analysis, because the risk allele identified in the meta-analysis showed the same direction only in women in the extension study, not in men.

For \( VEGF \) SNP rs833058, we genotyped 5,921 additional individuals (3,303 patients with hip OA and 2,618 control subjects). No significant association was observed in men (1,466 patients with hip OA and 1,263 control subjects), but the direction of change in the extension study was the same as that in the meta-analysis. Summary results in men showed association slightly below that in the meta-analysis \((2.6 \times 10^{-5})\). This SNP showed weak association in the combined analysis of men and women in the extension study \((P = 0.03)\).

**DISCUSSION**

This focused analysis of candidate genes within a large meta-analysis of GWAS identified associations...
with hip OA in 2 genes: COL11A1, which showed 2 independent signals, and VEGF, which showed association in men. The main objective of the study was to highlight SNPs for further study and confirmation as new OA genetic factors. A secondary aim of the study was to assess the validity of the candidate gene approach in the study of OA. In this respect, we observed that none of the other 197 candidate genes showed association in our meta-analysis of GWAS.

Discovery of genetic associations has been greatly advanced by GWAS due to increased sample sizes, increased coverage of analyzed SNPs, and increases in quality control standards and in the requirements to claim association that have accompanied them. These positive characteristics of the GWAS are further potentiated by their combination through meta-analysis (17). The current meta-analysis had an unprecedented power to analyze most of the candidate genes (at least for an OR of >1.15 and allele frequencies of >0.2), and GWAS plus imputation provided very complete coverage of genetic variation in them.

In addition, focused analysis of candidate genes should increase the chances of uncovering associations worth pursuing, by the following 2 mechanisms: increased prior likelihood of association and more tolerant threshold to claim association (18,19). A series of studies of other diseases have capitalized on the first mechanism, either by considering candidate genes from previous genetic studies, as was done here, or by defining a set of genes of high relevance for the disease, using bibliographic analysis. The second mechanism, a more tolerant threshold for association, is related to a problem that affects all complex diseases: the effect sizes of most genetic factors are below an OR of 1.20, and therefore large sample sizes are required to identify association at the genome level. In this context, the use of focused analysis in a subset of genes allows the selection of SNPs not reaching genome-wide significance for further validation.

Another point that deserves comment is that small variations between sample collections result in large differences in statistical power. This is attributable to the low effect sizes of most genetic factors and the dramatic decrease in statistical power when the effect size approaches 1.0 (see Supplementary Figure 1 available on the Arthritis & Rheumatology web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38300/abstract). It is very likely that this extreme sensitivity to variation of the effect size in the 1.0–1.15 range explains why confirmed genetic factors fail to show significant association even in some large studies, given that they are not large enough to offset these small fluctuations. In addition, there is a dramatic decrease in power for SNPs with a low MAF (<10%) (Supplementary Figure 1). Thus, we cannot exclude the possibility that some additional

---

**Table 3.** Association of the 2 top independent SNPs in COL11A1 and the top SNP in VEGF with hip OA*  

<table>
<thead>
<tr>
<th>Gene, SNP, sex</th>
<th>GWAS meta-analysis</th>
<th>Extension study</th>
<th>GWAS meta-analysis plus extension study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR_M-H (95% CI)</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>COL11A1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1241164</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.82 (0.74–0.89)</td>
<td>0.85 (0.68–1.06)</td>
<td>0.82 (0.75–0.89) 5.3 × 10⁻⁶</td>
</tr>
<tr>
<td>Female</td>
<td>0.79 (0.70–0.89)</td>
<td>0.96 (0.70–1.32)</td>
<td>0.81 (0.72–0.91) 2.4 × 10⁻⁴</td>
</tr>
<tr>
<td>Male</td>
<td>0.84 (0.74–0.97)</td>
<td>0.77 (0.56–1.06)</td>
<td>0.11 (0.83–0.94) 3.7 × 10⁻³</td>
</tr>
<tr>
<td>rs4907986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1.12 (1.06–1.17)</td>
<td>0.98 (0.85–1.12)</td>
<td>1.09 (1.05–1.15) 5.8 × 10⁻⁵</td>
</tr>
<tr>
<td>Female</td>
<td>1.13 (1.07–1.21)</td>
<td>1.15 (0.95–1.39)</td>
<td>1.13 (1.07–1.21) 3.2 × 10⁻⁵</td>
</tr>
<tr>
<td>Male</td>
<td>1.1 (1.03–1.19)</td>
<td>0.95 (0.77–1.18)</td>
<td>0.67 (1.01–1.16) 0.015</td>
</tr>
<tr>
<td>VEGF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs833058</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.92 (0.88–0.97)</td>
<td>0.92 (0.85–0.99)</td>
<td>0.92 (0.88–0.96) 1.9 × 10⁻⁴</td>
</tr>
<tr>
<td>Female</td>
<td>0.99 (0.93–1.06)</td>
<td>0.91 (0.81–1.01)</td>
<td>0.97 (0.91–1.02) 0.21</td>
</tr>
<tr>
<td>Male</td>
<td>0.85 (0.79–0.91)</td>
<td>0.94 (0.84–1.06)</td>
<td>0.31 (0.87–0.93) 2.6 × 10⁻⁵</td>
</tr>
</tbody>
</table>

* Single-nucleotide polymorphisms (SNPs) rs1241164 and rs4907986 were genotyped only in samples from Spain and Greece. The odds ratios (ORs) and 95% confidence intervals (95% CIs) are relative to the first listed allele for each SNP in Table 2. OA = osteoarthritis; GWAS = genome-wide association study; OR_M-H = OR as determined using the Mantel-Haenszel approach. No significant heterogeneity (P < 0.05) was detected in any of the combined analyses.
candidate genes may have associations with OA with an OR of \( \leq 1.10 \) or if their minor allele frequencies are low. 

\textit{COL11A1} codes for a minor component of the cartilage matrix, the importance of which has been shown by the chondrodysplasia mouse mutation, the skeletal abnormalities of Stickler syndrome, type II (OMIM ID 604841) and Marshall syndrome (OMIM ID 154780), and mutations in patients with fibrochondrogenesis and by association with susceptibility to lumbar disc herniation (http://omim.org/entry/120280). The 2 top SNPs associated with OA in our analysis were independent and different from those observed in a previous OA study (23). This raises the possibility of multiple variants with an effect in OA susceptibility.

\textit{VEGF} was associated with hip OA only in men, according to our meta-analysis. This gene codes for a very important angiogenic factor that is involved in normal growth plate development, endochondral ossification, and articular cartilage formation (43). It is one of the overexpressed markers of hypertrophic chondrocytes. It contributes to OA changes in animal models by stimulating chondrocyte proliferation, apoptosis, and production of catabolic mediators (44). Although it has been considered an OA candidate gene, no previous study showed significant association.

For the remaining candidate genes, we did not observe association in spite of the large number of samples assembled (the largest ever studied for most candidate genes) and the more tolerant threshold for significance allowed by the focused analysis. This does not exclude SNPs of weak effect or those showing heterogeneity between the sample collections. Also, it is possible that some gene variants were not adequately covered, especially those with low frequency or multiallelic polymorphisms, as in \textit{ASPN} (45) or \textit{BMP5} (46,47). Other reported associations have been described as specific for an OA subphenotype not included in our meta-analysis, such as the association of \textit{MATN3} with OA in the first carpometacarpal joint (48) or \textit{DIO2} in women with severe OA (49). In addition, \textit{GDF5}, which has been also a candidate gene, was not included in this analysis because it is already confirmed as an OA susceptibility locus at the genome-wide significance level (6,7).

All of the previous caveats apply, but they do not negate the conclusion of a general lack of reproducibility of OA candidate genes. This is particularly true for genes highlighted in candidate gene studies of small size and showing large effect sizes. It has become clear that the effect sizes and ORs reported were widely overestimated, because we would have observed most associations with ORs >1.20, and most reported ORs in these studies were \( \sim 2 \)-fold. Our findings suggest that traditionally conducted candidate gene studies are unlikely to be fruitful in OA genetics. However, such candidate gene studies are likely to continue to be useful to validate or refine loci from hypothesis-free genome-wide studies.

In summary, our candidate gene meta-analysis of 9 OA GWAS highlighted the association of \textit{COL11A1} and \textit{VEGF} with hip OA and showed a lack of association for the other 197 candidate genes.

**ACKNOWLEDGMENTS**

We thank the Newcastle arm of arcOGEN and the Freeman Hospital arthroplasty surgical and nursing teams. We are grateful for the support of the National Institute of Health Research (NIHR) Newcastle Biomedical Research Centre (Newcastle upon Tyne Hospitals NHS Foundation Trust and Newcastle University) and the Northumberland, Tyne and Wear Comprehensive Local Research Network. Finally, we thank EGCUT technical personnel, especially Mr. V. Soo and S. Smit. Data analyses were carried out in part at the High Performance Computing Center, University of Tartu.

**AUTHOR CONTRIBUTIONS**

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Gonzalez had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.


**ADDITIONAL DISCLOSURES**

Mr. Helgadottir and Drs. Jonsdottir, Styrkarsdottir, Thorsteinsdottir, Thorleifsson, and Stefansson are employees of deCODE Genetics.

**REFERENCES**

3. Hunter DJ, Demissie S, Cupples LA, Aliabadi P, Felson DT. A