Arthritis & Rheumatism
DOI 10.1002/art.38300

Accepted Article

Assessment of Osteoarthritis Candidate Genes in a Meta-analysis of 9 Genome-Wide Association Studies

Cristina Rodriguez-Fontenla [1], Manuel Calaza [1], Evangelos Evangelou [2,3], Ana M. Valdes [3], Nigel Arden [4], Francisco J. Blanco [5], Andrew Carr [6], Kay Chapman [6], Panos Deloukas [7], Michael Doherty [8], Tõnu Esko [9], Carlos M Garces [10], Juan J. Gomez-Reino [1,11], Hafdis Helgadottir [12], Albert Hofman [13,14], Ingileif Jonsdottir [12], Hanneke J.M Kerkhof [13,14], Margreet Kloppenburg [15], Andrew McCaskie [16], Evangelia E. Ntzani [2], William E.R. Ollier [17], Natividad Oreiro [5], Kalliope Panoutsopoulou [7], Stuart H. Ralston [18], Yolande F. Ramos [19], Jose A. Riancho [10], Fernando Rivadeneira [13,14], P. Eline Slagboom [14,19], Unnur Styrkarsdottir [12], Unnur Thorsteinsdottir [12], Gudmar Thorleifsson [12], Aspasia Tsezou [20], André G. Uitterlinden [13,14], Gillian A. Wallis [21], J. Mark Wilkinson [22], Guangju Zhai [3], Yanyan Zhu [23], the arcOGEN consortium, David T Felson [23], John P.A. Ioannidis [2,24], John Loughlin [16], Andres Metspalu [9], Ingrid Meulenbelt [14,19], Kari Stefansson [12], Joyce B. van Meurs [13,14], Eleftheria Zeggini [7], Timothy D. Spector [3] and Antonio Gonzalez [1]

1. Laboratorio Investigacion 10 and Rheumatology Unit, Instituto Investigacion Sanitaria – Hospital Clinico Universitario de Santiago, Santiago de Compostela, Spain.
2. Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece
3. Dept of Twin Research and Genetic Epidemiology, St. Thomas' Hospital, King's College London, London, UK
4. NIHR Biomedical Research Unit, University of Oxford, Oxford, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/art.38300
© 2013 American College of Rheumatology
Received: Mar 05, 2013; Revised: Sep 05, 2013; Accepted: Nov 26, 2013

Accepted Article

5. Rheumatology Division, INIBIC-Complejo Hospitalario Universitario A Coruna, A Coruna, Spain
6. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
7. Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom
8. Academic Rheumatology, Nottingham City Hospital, Nottingham, UK
9. Institute of Molecular and Cell Biology and Estonian Genome Center, University of Tartu, Tartu, Estonia
10. Hospital U. M. Valdecilla-IFIMAV, University of Cantabria, Santander, Spain
11. Department of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain
12. deCODE Genetics, Reykjavik, Iceland
13. Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherland
14. The Netherlands Genomics Initiative-Sponsored Netherlands Consortium for Healthy Aging, Rotterdam, The Netherlands
15. Dept. of Rheumatology and Dept. of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
16. Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK
17. Centre for Integrated Genomic Medical Research, The University of Manchester, Manchester, UK
18. University of Edinburgh, Institute of Genetics & Molecular Medicine, UK
19. Dept. of Molecular Epidemiology, Leiden University Medical Center , Leiden, The Netherlands
20. University of Thessaly, Faculty of Medicine, Department of Biology, Larissa, Greec

Accepted Article

21. Wellcome Trust Centre for Cell-Matrix Research, Faculty of Medical and Human Sciences, University of Manchester, UK

22. Department of Human Metabolism, University of Sheffield, UK

23. Clinical Epidemiology Unit, Boston University School of Medicine, Boston, MA, USA

24. Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, CA, USA

Corresponding author address:

Antonio Gonzalez

Laboratorio Investigacion 10

Instituto de Investigacion Sanitaria - Hospital Clinico Universitario de Santiago

Edificio de consultas, planta -2

Travesia de Choupana, sn

15706 - Santiago de Compostela

Spain.

E-mail: antonio.gonzalez.martinez-pedrayo@sergas.es

Telephone: +34 981 950 903, Fax: +34 981 950 906

Running title: OA candidate genes in meta-analysis of GWAs

Competing financial interests: The authors from deCODE Genetics declare competing financial interests: all are employees of the company, and some are shareholders. Hanneke JM Kerkhof is employed at Pfizer B.V. the Netherlands. The research done for this study was part of her PhD before she was employed at Pfizer.

Accepted Article

ABSTRACT

Objectives: To assess osteoarthritis (OA) candidate genes for identification of promising genetic factors and, secondarily, to assess the candidate gene approach in OA.

Methods: 199 published candidate genes for OA were obtained from the HuGe Navigator. All their SNPs with allele frequency >5% were assessed with fixed effect meta-analysis of 9 genome-wide association studies (GWAS) including 5 636 knee OA patients and 16 972 controls, and 4 349 hip OA patients and 17 836 controls of European ancestry. Additional 5 921 individuals were studied for top SNPs in the meta-analysis. Significance was corrected for the number of independent tests at $p < 1.58 \times 10^{-5}$.

Results: SNPs at only two of the 199 candidate genes were associated with OA in the meta-analysis. They were associated with hip OA, *COL11A1* showing two independent associations in the combined analysis (rs4907986, $p = 1.29 \times 10^{-5}$, OR = 1.12; 95 % CI = 1.06-1.17; and rs1241164, $p = 1.47 \times 10^{-5}$, OR = 0.82, CI = 0.74-0.89) and a SNP in linkage disequilibrium with rs4907986 in the female-specific analysis (rs4908291, $p = 1.29 \times 10^{-5}$, OR = 0.87, CI = 0.82-0.92), and *VEGF* associated in male-specific analysis (rs833058, $p = 1.35 \times 10^{-5}$, OR = 0.85, CI = 0.79-0.91). After genotyping additional samples, association at one of the *COL11A1* signals was reinforced, whereas association at *VEGF* was slightly weakened.

Conclusion: Two candidate genes were significantly associated with OA in this focused meta-analysis *COL11A1* and *VEGF*. The remaining candidate genes were not associated.

Accepted Article

INTRODUCTION

The etiology of primary OA is multifactorial, including ageing and mechanical, hormonal and genetic factors (1). Investigation of many of these factors has produced contradictory results making conclusions difficult. Some of the contradictory results are attributable to the design of the studies, but others can be due to unaccounted patient differences (1, 2). It is also notable that susceptibility factors, including genetic ones, are not equally shared by the different joint regions affected by OA (3, 4).

Several studies have tried to identify genetic factors involved in OA with association studies of candidate genes, linkage studies in multicase families or genome-wide association studies (GWAS) (2, 5). Candidate genes have been studied because of their important roles in pathogenesis or their altered expression in OA tissue. Most of these association studies have been inconclusive. There is only one candidate gene, *GDF5*, that is well supported by multiple studies and that has reached genome-wide significance ($p < 5 \times 10^{-8}$) on subsequent replication efforts (6, 7). Linkage studies provided some consistent results, but it proved difficult to progress from them to the identification of the culprit genes (2, 5).

Recently, GWAS have identified 11 additional OA susceptibility loci with a genome-wide level of significance. Two of them in *DVWA/COL6A4* and the *HLA* are associated in Asians, but not in Europeans (8-10). The other nine loci have reached genome-wide significance in Europeans. They include a locus with 6 genes in chromosome *7q22* associated with knee OA (11, 12), the *MCF2L* gene associated with OA of the knee and hip (13), five loci identified in the arcOGEN GWAS (14): *GNL3/GLT8D1* associated with OA of knee and hip, *ASTN2* associated with severe hip OA in women, *FILIP1-SENP6* and *PTHLH* with hip OA and *CHTS11* with severe hip OA; and *DOT1L* that is

Accepted Article

associated with JSW in the hip (15) and with hip OA in men (16). About eight more loci are near this level of association (2, 5, 14).

All the studies reaching genome-wide significance 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 SNPs in the genome, either directly or through imputation, and therefore all add to the whole result (17). Another way to favor discovery of new loci is by focusing analysis on particular subsets of genes, for which there are 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).

Here, we aimed to identify new OA genetic factors taking advantage of the two above-mentioned approaches, meta-analysis of GWAS and focused analysis of OA candidate genes. A secondary aim of the study was to assess validity of the candidate gene approach in OA. To this end, we have used a meta-analysis of 5 636 knee and 4 349 hip OA patients from nine GWAS and explored association with knee or hip OA of more than 24 000 SNPs corresponding to 199 previously reported OA candidate genes. Significant associations were found at two candidate genes, *COL11A1* and *VEGF*. The remaining 197 candidate genes were not significantly associated with OA.

MATERIAL AND METHODS

Sample collections: We performed a meta-analysis of SNP-level GWAS results from 9 large sample collections of European descent including patients with knee or hip OA and controls (Table 1): deCODE from Iceland (20), three collections from the Rotterdam Study (Rotterdam Study I, II and III) (21) and the GARP collection (22)

Accepted Article

from the Netherlands, arcOGEN phase I(23) and Twin UK (24) from the UK, the Framingham Osteoarthritis Study from the USA (25), and EGCUT from Estonia (26). Additional sample collections not involved in GWAS were used as extension study of the significant results (Table 1). They included collections from the North of Spain (27-29), the centre of Greece (30) and the Nottingham (31) and GOAL (32) studies from the UK.

All these sample collections have been described in detail in previous reports. In brief, the deCODE study included joint replacement, TKR and THR, patients with OA and population controls, which excluded all individuals on OA susceptibility lists (hand, hip, knee) obtained from hospitals and health care centers in Iceland (20). The Rotterdam Study, Framingham Osteoarthritis Study and Twins UK included patients with radiographic OA and controls from the same cohort without radiographic signs of OA defined according to standardized phenotypes with cases having a Kellgren and Lawrence (KL) grade of ≥2 and controls a KL grade of <2 (33). The GARP cohort consisted of clinical and radiographically (KL grade of ≥2) confirmed OA at two or more joint sites (22) and were compared to random controls. The arcOGEN phase 1 included knee and hip OA determined by radiographic evidence of disease and clinical evidence of OA that in many patients required joint replacement (23). Controls for this study were from an early release of the Wellcome Trust Case Controls Consortium 2 data. EGCUT included OA confirmed by radiograph (KL score > 2) and controls were free of any OA symptoms. All collections of samples used for the extension study were from case-control studies. Patients were ascertained by hip joint replacement due to symptomatic and radiographically confirmed hip OA except in GOAL, where severe symptomatic hip OA was the selection criterion. All the sample collections and the

Accepted Article

genetic studies have 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 **(34)** (April 29, 2013) to identify candidate genes that have been studied for their possible association with OA without additional filtering. Query terms were: osteoarthritis, spinal osteophytosis and intervertebral disk displacement. The candidate genes of the two latter keywords showed large overlapping with the osteoarthritis list (90 % and 72 %, respectively). The 17 non-overlapping genes were considered 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 GWA designs. *FRAH1H* was excluded because the bibliographic reference was wrong. All the genes with GWAS association in Europeans ($p < 5 \times 10^{-8}$) 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 MAF >5% in the CEU data set corresponding to the candidate gene loci were retrieved from HapMap (phases I+II + III, 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 meta-analysis were different and have been described in detail (11, 20, 23, 25, 35). Imputation of untyped SNPs was done based on HapMap phases I+II release 22 CEU data. Summary information on genotyping and imputation is shown in Supplementary Table S1. Genotyping in the extension study collections was done at

Accepted Article

Santiago by single-base extension with the SNaPshot Multiplex Kit (Applied Biosystems) for the Spanish and Greek collections or was carried out by Kbioscience Ltd, Hertfordshire, UK, using the KASPar chemistry for the UK collections.

Statistical analyses: Each team performed association testing for knee OA and hip OA under a per-allele model. The lambda inflation factor was calculated per gender-specific effect size using the genomic control method (36) and the standard errors were corrected by the square root of the lambda inflation factor ($SE_{corrected} = SE_{observed} x\sqrt{\lambda}$). Robust standard errors were estimated to adjust for family relationships in the deCODE and GARP studies. For meta-analysis, the effect size for each SNP (odds ratio per copy of minor allele according to HapMap) was calculated using inverse-variance fixed effects models synthesizing all effect sizes and the corrected standard errors. Heterogeneity was assessed with the inconsistency I^2 statistic and when low or moderate no random-effet meta-analysis was done (37). Meta-analysis of the GWAS was performed using METAL (38) considering six strata with two joint levels (knee and hip) and three gender levels (all, women and men) . Two research centers (Ioannina, Greece and Erasmus MC Rotterdam, the Netherlands) performed both the Quality Control (QC) and meta-analyses for the whole GWAS. A QC protocol was set up including validation of the results file format, reports for range of values and elimination of potential biases (i.e., extremely large beta′s or SEs). Files were cross-validated between the two research centers after QC and after meta-analyses to check for inconsistencies. SNPs with a MAF <1%, imputation quality <0.30 (MACH) or <0.40 (IMPUTE) and beta′s >4 or <-4 and SNPs that were not available in >4 studies were excluded for further analysis.

The significance threshold for claiming association was determined considering the number of independent tests performed. This number was estimated with a modification

Accepted Article

of the simpleM algorithm (39) applied to the genotypes of the CEU collection in HapMap. The modification consisted in replacing the observed correlation matrix for the nearest positive semidefinite matrix, as implemented in the R package corpcor, to correct for biases introduced by missing genotypes. It was estimated that the number of independent tests represented by these SNPs was 3 156 (Supplementary Table S2). The number of independent tests was used to define a significance threshold of $p = 1.58 \times 10^{-5}$ according to the Bonferroni multiplicity correction. No additional correction was done for stratification by joint and gender because there is known heterogeneity in OA genetics across these strata (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 with the Mantel-Haenszel approach (40). Combination of the extension study with the GWAS data was done using a fixed effects model with R software (41). Power estimates were obtained with the Power and Sample size software (42). A full analysis of power is provided in Supplementary Figure 1. As an illustrative example, power was 80% to detect association with a SNP of MAF = 20 % and OR = 1.15 for knee OA or 1.16 for hip OA assuming no heterogeneity between the GWAS.

RESULTS

Systematic identification of candidate genes. A total of 199 genes (Table 2) have been investigated for their association with OA in humans according to the HuGe Navigator (34). The HuGe Navigator included 542 bibliographic references with the genetic studies of the candidate genes. Some of the genes are from loci that had been 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*). They were

Accepted Article

included because the identification of many of them as putative susceptibility genes partially came from *post hoc* analyses of their potential biological role and because none has reached genome wide significant association in Europeans (none with $p < 5 \times 10^{-8}$). The 199 genes were grouped in 158 non-overlapping genome segments that contained 27 501 autosomal SNPs (MAF>5%) with known genotypes in the CEU population of HapMap.

Meta-analysis of association of all SNPs in the candidate genes: Effect sizes from each of the nine GWAS corresponding to SNPs in the candidate genes were obtained and combined in a meta-analysis. Genotypes were available for 25 839 SNPs of the 27 501 included in candidate genes after applying quality control filters (A supplementary results file with all the OR and P values is available for download). These genotypes had been directly typed or imputed. In the knee OA meta-analysis there were no significant associations.

Meta-analysis of hip OA GWAS showed significant association at two candidate genes. Two SNPs, rs4907986 and rs1241164, in the 5' and 3' extremes of *COL11A1*, respectively, were associated in the unstratified analysis (Table 3 and Figure 1A). They comprised two independent associations as manifested by the low pairwise r^2 (0.09) between them. A SNP of *COL11A1*, rs2615977, was highlighted in the previously reported analysis of the arcOGEN phase 1 study (with $p = 1.1 \times 10^{-5}$) (23), which overlaps with the current meta-analysis. However, none of the two independent top associated SNPs showed strong r^2 with rs2615977 (pairwise $r^2 < 0.4$) and rs2615977 was not among the most associated SNPs at the meta-analysis. Stratified analysis by gender showed also association of a *COL11A1* SNP, rs4908291, in women. This SNP showed LD with rs4907986 ($r^2 = 0.68$) but not with rs1241164.

Accepted Article

The second candidate gene associated with hip OA was *VEGF*. It showed only a SNP, rs833058, reaching association over the required threshold in men (Table 3). No other SNP with strong or modest (r^2 >0.5) LD with rs833058 was present in the meta-analysis (Figure 1B). 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 hip OA association of three SNPs, two from *COL11A1* representing the top SNPs of the two independent associations in the combined analysis and the top SNP in *VEGF*. This part of the study was not intended as replication of the results due to the relative small number of independent samples that were available.

For the *COL11A1* SNPs 1929 samples from Spanish and Greek individuals (784 hip OA patients and 1142 controls) were studied. No significant association was found in this underpowered analysis (Table 4; study in additional samples showed independent association, Hanneke J.M Kerkhof and Joyce B. van Meurs, personal communication) but when these data were combined with the meta-analysis of GWAS 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 (Table 4).The second signal, corresponding to rs4907986, was slightly weakened in the combined analysis because the risk allele in the meta-analysis was only of risk in the women of the extension study.

For the *VEGF* polymorphism, rs833058, we have genotyped 5921 additional individuals (3303 hip OA patients and 2618 controls). No significant association was found in men (1466 hip OA patients and 1263 controls), but direction of change was the same in the extension samples than in the meta-analysis. Summary results in men showed association slightly below the found in the meta-analysis (2.6×10^{-5}). This SNP

Accepted Article

showed weak association in the combined analysis of men and women in the extension study (p = 0.03).

DISCUSSION

Focused analysis of candidate genes within a large meta-analysis of GWAS was able to uncover associations with hip OA showing strong statistical support in two genes *COL11A1* that showed two independent signals and *VEGF* that was associated in men. This was the main objective of the study, 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 have found 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 the increase in sample size, in coverage of analyzed SNPs, 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 OR > 1.15 and allele frequencies over 0.2, and also GWAS plus imputation gave us a very complete coverage of genetic variation in them. In addition, focused analysis in candidate genes should increase the chances of uncovering associations worth pursuing by two mechanisms: increased prior likelihood of association and more tolerant threshold to claim association (18, 19). A series of studies of other diseases have capitalized in the first aspect either by considering candidate genes from previous genetic studies, as done here, or by defining a set of genes of high relevance for the disease using bibliographic analysis. The second aspect,

Accepted Article

a more tolerant threshold for association, is related to a problem that affects all complex diseases: effect sizes of most genetic factors are below OR = 1.20 and therefore large sample sizes are required to find association at the genome level. In this context, the use of focused analysis in a subset of genes allows selecting SNPs not reaching genome-wide significance for further validation.

Another point worth to comment is that small variations between sample collections result in large differences in statistical power. This is due to the low effect sizes of most genetic factors and the dramatic decrease in statistical power when the effect size approaches 1.0 (Supplementary Figure 1). It is very likely that this extreme sensitivity to variation of the effect size in the 1.0-1.15 range explains that confirmed genetic factors fail to show significant association even in some large studies given that they are not large enough as to offset these small fluctuations. In addition, there is a dramatic decrease in power for SNPs of low minor allele frequency < 10 % (Supplementary Figure 1). Thus we cannot exclude that some additional candidate genes may have associations with OA with OR ≤ 1.10 or if their minor allele frequencies are low.

COL11A1 codes for a minor component of the cartilage matrix whose importance has been shown by the chondrodysplasia mouse mutation, the skeletal abnormalities of Stickler type II (OMIM 604841) and Marshall (OMIM 154780) syndromes, and mutations in patients with fibrochondrogenesis and by association with susceptibility to lumbar disc herniation (43). The two top SNPs observed in our analysis were independent and different from the reported in a previous OA study (23). This raises the possibility of multiple variants with an effect in OA susceptibility. One of them increased its association with hip OA after our limited extension study.

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

Accepted Article

plate development, endochondral ossification and articular cartilage formation (44). 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 (45). Although it has been considered an OA candidate gene, no previous study showed significant association.

For the remaining candidate genes, we did not find association in spite of the large number of samples assembled, the largest ever studied for most candidate genes, and the more tolerant significance threshold allowed by the focused analysis. This does not exclude SNPs of weak effect or 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 *ASPN* (46) or *BMP5* (47, 48). Other reported associations have been described as specific of an OA subphenotype not included in our meta-analysis, as for the association of *MATN3* with OA in the first carpometacarpal joint (49), or *DIO2* in women with severe OA (50). In addition, GDF5 that has been also a candidate gene has not been included in this analysis because it is already confirmed as OA susceptibility locus at the genome-wide significance level (6, 7). All the previous caveats apply, but they do not negate the conclusion of a general lack of reproducibility of OA candidate genes. This is particularly so for genes highlighted in candidate gene studies of small size and showing large effect sizes. It has become clear that the effect sizes and odds ratios reported were widely overestimated as we would have found most associations with odds ratios of over 1.20 and most reported odds ratios in these studies were around two-fold. Our findings suggest that traditionally conducted candidate gene studies are unlikely to be fruitful in OA genetics. However, candidate gene studies including accurate genotyping technologies, sufficient quality

Accepted Article

control, large collections of samples, replication and sound statistical analysis could be useful to validate or refine loci from agnostic genome-wide or sequencing platforms. In summary, our candidate gene meta-analysis of 9 OA GWAS highlighted the association of *COL11A1* and *VEGF* with hip OA and showed lack of association for the other 197 candidate genes.

ACKNOWLEDGMENTS

This work has been funded by EU Project Framework 7 grant TreatOA (grant 200800) and by the Instituto de Salud Carlos III (ISCIII, Spain) grants PS09/01431 and PI11/01048 that are partially financed by the European Regional Development Fund of the EU. C. R-F is the recipient of a predoctoral contract of the ISCIII (Spain). MC salary is provided by the Isabel Barreto program of the Xunta of Galicia (Spain). The Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) (project nr. 050-060-810) and the Erasmus Medical Center and Erasmus University, Rotterdam. The Leiden University Medical Centre, the Dutch Arthritis Association and Pfizer Inc., Groton, CT, USA support the GARP study. The genotyping was supported by the Netherlands Organization of Scientific Research (MW 904-61-095, 911-03-016, 917-66-344 and 911-03-012), Leiden University Medical Centre, and by the "Centre of Medical System Biology" and the "Netherlands Consortium of Healthy Aging" in the framework of the Netherlands Genomics Initiative (NGI). Furthermore, the research leading to these results has received funding from the Dutch Arthritis Association (DAA 2010_017) and

Accepted Article

the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement n° 259679. arcOGEN (http://www.arcogen.org.uk/) was funded by a special purpose grant from Arthritis Research UK (grant 18030). For the Newcastle arm of arcOGEN we acknowledge the Freeman Hospital arthroplasty surgical and nursing teams. We acknowledge the support of the National Institute for Health Research (NIHR) Newcastle Biomedical Research Centre based at Newcastle upon Tyne Hospitals NHS Foundation Trust and Newcastle University, and the support of the Northumberland, Tyne and Wear CLRN. EGCUT received targeted financing from Estonian Government SF0180142s08, Center of Excellence in Genomics (EXCEGEN) and University of Tartu (SP1GVARENG). We acknowledge EGCUT technical personnel, especially Mr V. Soo and S. Smit. Data analyzes were carried out in part in the High Performance Computing Center of University of Tartu. JAR received funding from ISCIII-FIS PI 06-0034. EZ and KP are supported by the Wellcome Trust grant [098051].

REFERENCES

1. Blagojevic M, Jinks C, Jeffery A, Jordan KP. Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. Osteoarthritis Cartilage. 2010;18:24-33.
2. Loughlin J. Genetics of osteoarthritis. Curr Opin Rheumatol. 2011;23:479-83.
3. Hunter DJ, Demissie S, Cupples LA, Aliabadi P, Felson DT. A genome scan for joint-specific hand osteoarthritis susceptibility: The Framingham Study. Arthritis Rheum. 2004;50:2489-96.
4. MacGregor AJ, Li Q, Spector TD, Williams FM. The genetic influence on radiographic osteoarthritis is site specific at the hand, hip and knee. Rheumatology (Oxford). 2009;48:277-80.
5. Valdes AM, Spector TD. The contribution of genes to osteoarthritis. Med Clin North Am. 2009;93:45-66, x.
6. Miyamoto Y, Mabuchi A, Shi D, Kubo T, Takatori Y, Saito S, et al. A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. Nat Genet. 2007;39:529-33.

Accepted Article

7. Valdes AM, Evangelou E, Kerkhof HJ, Tamm A, Doherty SA, Kisand K, et al. The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. Ann Rheum Dis. 2011;70:873-5.
8. Miyamoto Y, Shi D, Nakajima M, Ozaki K, Sudo A, Kotani A, et al. Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. Nat Genet. 2008;40:994-8.
9. Nakajima M, Takahashi A, Kou I, Rodriguez-Fontenla C, Gomez-Reino JJ, Furuichi T, et al. New sequence variants in HLA class II/III region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. PLoS One. 2010;5:e9723.
10. Meulenbelt I, Chapman K, Dieguez-Gonzalez R, Shi D, Tsezou A, Dai J, et al. Large replication study and meta-analyses of DVWA as an osteoarthritis susceptibility locus in European and Asian populations. Hum Mol Genet. 2009;18:1518-23.
11. Kerkhof HJ, Lories RJ, Meulenbelt I, Jonsdottir I, Valdes AM, Arp P, et al. A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. Arthritis Rheum. 2010;62:499-510.
12. Evangelou E, Valdes AM, Kerkhof HJ, Styrkarsdottir U, Zhu Y, Meulenbelt I, et al. Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. Ann Rheum Dis. 2011;70:349-55.
13. Day-Williams AG, Southam L, Panoutsopoulou K, Rayner NW, Esko T, Estrada K, et al. A variant in MCF2L is associated with osteoarthritis. Am J Hum Genet. 2011;89:446-50.
14. Zeggini E, Panoutsopoulou K, Southam L, Day-Williams A, Lopes M, Boraska V, et al. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet. 2012;380:815-23.
15. Castano Betancourt MC, Cailotto F, Kerkhof HJ, Cornelis FM, Doherty SA, Hart DJ, et al. Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. Proc Natl Acad Sci U S A. 2012;109:8218-23.
16. Evangelou E, Valdes AM, Castano-Betancourt MC, Doherty M, Doherty S, Esko T, et al. The DOT1L rs12982744 polymorphism is associated with osteoarthritis of the hip with genome-wide statistical significance in males. Ann Rheum Dis. 2013.
17. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. Pharmacogenomics. 2009;10:191-201.
18. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, Sklar P, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 2009;5:e1000534.
19. Khoury MJ, Bertram L, Boffetta P, Butterworth AS, Chanock SJ, Dolan SM, et al. Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases. Am J Epidemiol. 2009;170:269-79.
20. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, Ingvarsson T, et al. New sequence variants associated with bone mineral density. Nat Genet. 2009;41:15-7.
21. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, et al. The Rotterdam Study: objectives and design update. Eur J Epidemiol. 2007;22:819-29.
22. Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio le Graverand MP, Rosendaal FR, et al. Evidence for familial aggregation of hand, hip, and spine but not

Accepted Article

knee osteoarthritis in siblings with multiple joint involvement: the GARP study. Ann Rheum Dis. 2005;64:438-43.
23. Panoutsopoulou K, Southam L, Elliott KS, Wrayner N, Zhai G, Beazley C, et al. Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. Ann Rheum Dis. 2011;70:864-967.
24. Spector TD, MacGregor AJ. The St. Thomas' UK Adult Twin Registry. Twin Res. 2002;5:440-3.
25. Kiel DP, Demissie S, Dupuis J, Lunetta KL, Murabito JM, Karasik D. Genome-wide association with bone mass and geometry in the Framingham Heart Study. BMC Med Genet. 2007;8 Suppl 1:S14.
26. Metspalu A, Kohler F, Laschinski G, Ganten D, Roots I. [The Estonian Genome Project in the context of European genome research]. Dtsch Med Wochenschr. 2004;129 Suppl 1:S25-8.
27. Rodriguez-Lopez J, Pombo-Suarez M, Liz M, Gomez-Reino JJ, Gonzalez A. Further evidence of the role of frizzled-related protein gene polymorphisms in osteoarthritis. Ann Rheum Dis. 2007;66:1052-5.
28. Rego-Perez I, Fernandez-Moreno M, Fernandez-Lopez C, Arenas J, Blanco FJ. Mitochondrial DNA haplogroups: role in the prevalence and severity of knee osteoarthritis. Arthritis Rheum. 2008;58:2387-96.
29. Velasco J, Zarrabeitia MT, Prieto JR, Perez-Castrillon JL, Perez-Aguilar MD, Perez-Nunez MI, et al. Wnt pathway genes in osteoporosis and osteoarthritis: differential expression and genetic association study. Osteoporos Int. 2010;21:109-18.
30. Kostopoulou F, Gkretsi V, Malizos KN, Iliopoulos D, Oikonomou P, Poultsides L, et al. Central role of SREBP-2 in the pathogenesis of osteoarthritis. PLoS One. 2012;7:e35753.
31. Valdes AM, Lories RJ, van Meurs JB, Kerkhof H, Doherty S, Hofman A, et al. Variation at the ANP32A gene is associated with risk of hip osteoarthritis in women. Arthritis Rheum. 2009;60:2046-54.
32. Holliday KL, McWilliams DF, Maciewicz RA, Muir KR, Zhang W, Doherty M. Lifetime body mass index, other anthropometric measures of obesity and risk of knee or hip osteoarthritis in the GOAL case-control study. Osteoarthritis Cartilage. 2011;19:37-43.
33. Kerkhof HJ, Meulenbelt I, Akune T, Arden NK, Aromaa A, Bierma-Zeinstra SM, et al. Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. Osteoarthritis Cartilage. 2011;19:254-64.
34. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. Nat Genet. 2008;40:124-5.
35. Valdes AM, Loughlin J, Timms KM, van Meurs JJ, Southam L, Wilson SG, et al. Genome-wide association scan identifies a prostaglandin-endoperoxide synthase 2 variant involved in risk of knee osteoarthritis. Am J Hum Genet. 2008;82:1231-40.
36. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55:997-1004.
37. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557-60.
38. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010.
39. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol. 2008;32:361-9.

Accepted Article

40. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959;22:719-48.
41. R Core Team. R: A Language and Environment for Statistical Computing. 2013 [cited; Available from: http://www.R-project.org
42. Dupont WD, Plummer WD, Jr. Power and sample size calculations. A review and computer program. Control Clin Trials. 1990;11:116-28.
43. OMIM entry for COL11A1. 2012 11/12/2012 [cited; Available from: http://omim.org/entry/120280
44. Stempel J, Fritsch H, Pfaller K, Blumer MJ. Development of articular cartilage and the metaphyseal growth plate: the localization of TRAP cells, VEGF, and endostatin. J Anat. 2011;218:608-18.
45. Ludin A, Sela JJ, Schroeder A, Samuni Y, Nitzan DW, Amir G. Injection of vascular endothelial growth factor into knee joints induces osteoarthritis in mice. Osteoarthritis Cartilage. 2012.
46. Nakamura T, Shi D, Tzetis M, Rodriguez-Lopez J, Miyamoto Y, Tsezou A, et al. Meta-analysis of association between the ASPN D-repeat and osteoarthritis. Hum Mol Genet. 2007;16:1676-81.
47. Wilkins JM, Southam L, Mustafa Z, Chapman K, Loughlin J. Association of a functional microsatellite within intron 1 of the BMP5 gene with susceptibility to osteoarthritis. BMC Med Genet. 2009;10:141.
48. Rodriguez-Fontenla C, Carr A, Gomez-Reino JJ, Tsezou A, Loughlin J, Gonzalez A. Association of a BMP5 microsatellite with knee osteoarthritis: case-control study. Arthritis Res Ther. 2012;14:R257.
49. Stefansson SE, Jonsson H, Ingvarsson T, Manolescu I, Jonsson HH, Olafsdottir G, et al. Genomewide scan for hand osteoarthritis: a novel mutation in matrilin-3. Am J Hum Genet. 2003;72:1448-59.
50. Meulenbelt I, Min JL, Bos S, Riyazi N, Houwing-Duistermaat JJ, van der Wijk HJ, et al. Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. Hum Mol Genet. 2008;17:1867-75.

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

Meta-analysis	Knee OA		Hip OA		OA ascertainment
	Cases	Controls	Cases	Controls	
arcOGEN	1643	4896	1728	4896	TJR and ROA[a]
deCODE*	1312	2318	1423	2318	TJR [b] and clinical
Framingham Study	417	1667	NA		ROA[a]
GARP	154	1671	106	1671	ROA[a] and clinical
Rotterdam Study I	1476	3233	760	3233	ROA[a]
Rotterdam Study II	369	1472	159	1472	ROA[a]
Rotterdam Study III	152	1487	41	1487	ROA[a]
TwinsUK	113	228	68	228	ROA[a]
EGCUT			64	2531	ROA[a]
Total GWAS	5636	16972	4349	17836	
Extension study					
Spain [c]	-	-	695	784	TJR [b]
Greece	-	-	92	358	TJR [b]
Notthingham[d]	-	-	1246	713	TJR [b]
GOAL[d]	-	-	1270	763	TJR and ROA[a]
Total extension			3303	2618	

ROA = Radiographic OA; [b] TJR = total joint replacement because of OA
*deCODE effective sample size
Samples from three collections in the North of Spain: Santiago, Santander and A Corunna
These samples were genotyped only for rs833058*(VEGF)*

Accepted Article

Accepted Article

Table 2: List of OA candidate genes according to our selection from HuGe Navigator

A2BP1	*COL1A1*	*HMGA2*	*KIR3DL3*	*SCN9A*
ACAN	*COL1A2*	*HOXB9*	*KIR3DS1*	*SELS*
ACE	*COL2A1*	*HTRA1*	*KL*	*SEPP1*
ADAM12	*COL3A1*	*ICAM1*	*LECT1*	*SERPINE1*
ADAMTS14	*COL5A1*	*IBSP*	*LEP*	*SERPINA3*
ADAMTS3	*COL5A2*	*IFNG*	*LEPR*	*SLC26A2*
ADAMTS5	*COL6A1*	*IGF1*	*LOC344875*	*SLC6A4*
AIF1	*COL9A1*	*IGFBP7*	*LPAR1/EDG2*	*SMAD3*
ANP32A	*COL9A2*	*IHH*	*LRCH1*	*SOCS1*
APOE	*COL9A3*	*IL10*	*LRP5*	*SOST*
APOM	*COMP*	*IL12A*	*LRP6*	*SPP1*
ASPN	*COMT*	*IL13*	*LTA*	*STAT4*
B2M	*CRP*	*IL18*	*LY6G5C*	*TBX4*
BAT2	*CSNK1D*	*IL18R1*	*MATN1*	*TGFB1*
BAT3	*CSNK2B*	*IL18RAP*	*MATN3*	*THBS2*
BAT4	*CYP19A1*	*IL1A*	*MMP1*	*THRA*
BAT5	*CYP2D6*	*IL1B*	*MMP13*	*TIMP2*
BDNF	*DCT*	*IL1R1*	*MMP2*	*TIMP4*
BMP2	*DIO2*	*IL1R2*	*MMP3*	*TLR2*
BMP5	*DIO3*	*IL1RL1*	*MMP8*	*TLR3*
BTNL2	*ENG*	*IL1RL2*	*MMP9*	*TLR4*
C5	*ENPP1*	*IL1RN*	*MTHFR*	*TLR9*
C6orf10	*EPAS1*	*IL2*	*NCOR2*	*TNF*
C6orf47	*ESR1*	*IL4*	*NFKB1*	*TNFAIP3*
CALCA	*ESR2*	*IL4R*	*NOS2*	*TNFAIP6*
CALCRL	*FRZB*	*IL6*	*NOS3*	*TNFRSF11B*
CALM1	*GHR*	*IL8*	*OLIG3*	*TNFRSF1A*
CALM2	*GNB3*	*KIAA1217*	*OPRM1*	*TNFRSF1B*
CAT	*GPX1*	*KIR2DL1*	*PAPPA2*	*TNFSF11*
CASP9	*GSTM1*	*KIR2DL2*	*PAPSS2*	*TP53*
CCL2	*GSTP1*	*KIR2DL3*	*PCSK6*	*TRAF1*
CCR5	*GSTT1*	*KIR2DL4*	*PITX1*	*TRPV1*
CD36	*HAPLN1*	*KIR2DL5A*	*PLA2G4A*	*TXNDC3*
CILP	*HFE*	*KIR2DP1*	*PPARG*	*TXNRD2*
CLEC3B	*HIF1A*	*KIR2DS1*	*PTGS2*	*USP33*
CNR1	*HLA-B*	*KIR2DS2*	*PTH*	*VDR*
CNTF	*HLA-DQA1*	*KIR2DS3*	*PTPN22*	*VEGF*
COL10A1	*HLA-DQB1*	*KIR2DS4*	*RAMP2*	*WISP1*
COL11A1	*HLA-DRB1*	*KIR3DL1*	*RHOB*	*WRN*
COL11A2	*HLA-DRB5*	*KIR3DL2*	*ROR2*	

Accepted Article

Table 3: SNPs independently associated with hip OA in all samples or in the female or male stratified analyses. OR and their 95 % CI, are relative to the first listed allele. SNPs below the threshold of significance p = 1.58 x 10^{-5} are shown.

	Locus	SNP	Chr	Position	Alleles	OR (95 % CI)	*P*-value	I^2	P_{het}-value
All subjects	*COL11A1*	rs4907986	1	103322221	T/C	1.12 (1.06-1.17)	1.29 x 10^{-5}	0	0.63
	COL11A1	rs1241164	1	103129065	T/C	0.82 (0.74-0.89)	1.47 x 10^{-5}	0	0.92
Women	*COL11A1*	rs4908291	1	103345324	A/T	0.87 (0.82-0.92)	1.29 x 10^{-5}	33.2	0.16
Men	*VEGF*	rs833058	6	43839832	T/C	0.85 (0.79-0.91)	1.35 x 10^{-5}	18.6	0.29

Table 4: Association of the two independent SNPs in *COL11A1* and the top SNP in *VEGF* with hip OA. Results of the GWAS meta-analysis, the extension study and their summary are shown for all, females and males. OR and their 95 % CI, are relative to the same SNP as in table 3. No significant heterogeneity ($p < 0.05$) was detected in any of the combined analyses.

			GWAS meta-analysis		Extension study		Summary = GWAS meta + extension	
Gene	SNP	Gender	$O.R._{M-H}$ (95%CI)	p-value	$O.R._{M-H}$ (95%CI)	p-value	O.R. (95%CI)	p-value
COL11A1[a]	rs1241164	All	0.82 (0.74, 0.89)	1.47×10^{-5}	0.85 (0.68, 1.06)	0.15	0.82 (0.75, 0.89)	5.3×10^{-6}
		Female	0.79 (0.7, 0.89)	1.27×10^{-4}	0.96 (0.70, 1.32)	0.81	0.81 (0.72,0.91)	2.4×10^{-4}
		Male	0.84 (0.74, 0.97)	0.012	0.77 (0.56, 1.06)	0.11	0.83 (0.73, 0.94)	3.7×10^{-3}
COL11A1[a]	rs4907986	All	1.12 (1.06,1.17)	1.29×10^{-5}	0.98 (0.85, 1.12)	0.77	1.09 (1.05, 1.15)	5.8×10^{-5}
		Female	1.13 (1.07,1.21)	1.05×10^{-4}	1.15 (0.95, 1.39)	0.13	1.13 (1.07, 1.21)	3.2×10^{-5}
		Male	1.1 (1.03,1.19)	0.0069	0.95 (0.77, 1.18)	0.67	1.08 (1.01, 1.16)	0.015
VEGF	rs833058	All	0.92 (0.88, 0.97)	0.0022	0.92 (0.85, 0.99)	0.03	0.92 (0.88, 0.96)	1.9×10^{-4}
		Female	0.99 (0.93, 1.06)	0.75	0.91 (0.81, 1.01)	0.07	0.97 (0.91, 1.02)	0.21
		Male	0.85 (0.79, 0.91)	1.3×10^{-5}	0.94 (0.84, 1.06)	0.31	0.87 (0.82, 0.93)	2.6×10^{-5}

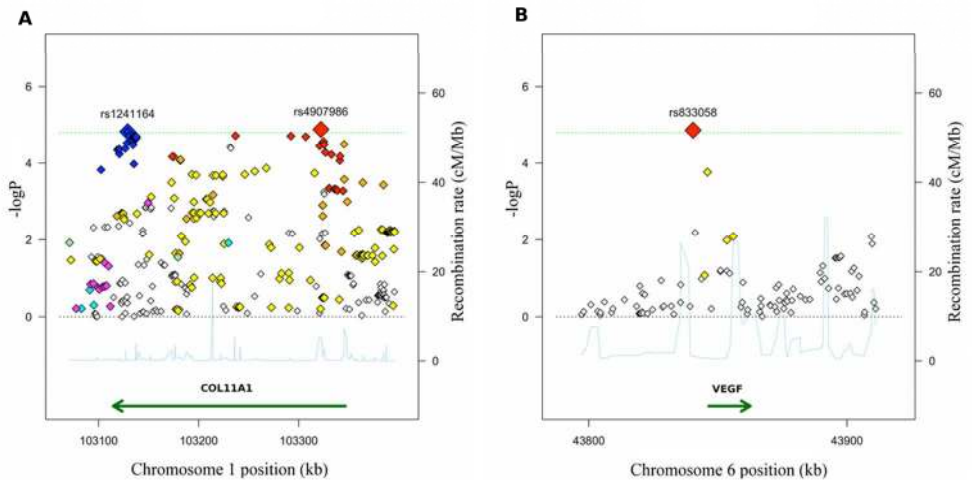
[a] These SNPs were only genotyped in samples from Spain and Greece

Accepted Article

Accepted Article

Figure Legends

Figure 1. Regional association plot of A) the COL11A1 and B) the VEGF loci with hip OA. Each SNP is a diamond located in the genome physical map (X axis) and in the –log 10 scale of its p-value of association (left Y axis). The dotted horizontal line shows the 1.58 x10-5 threshold for significance. The continuous blue line represents the recombination rate (right Y axis). Gene annotations are given over the X axis..In A, pairwise correlation of each SNP with rs4907986 is shown from minimal = white to maximal ($r2 \geq 0.8$) = red, through orange for $r2 \geq 0.5 < 0.8$. In contrast, pairwise correlation of each SNP with rs1241164 is shown from minimal = white to maximal ($r2 \geq 0.8$) = blue through magenta for $r2 \geq 0.5 < 0.8$. In B, pairwise correlation of each SNP with the top associated SNP (rs833058) is shown from minimal = white to maximal ($r2 \geq 0.8$) = red.



96x52mm (300 x 300 DPI)

Accepted