Identification of new susceptibility loci for osteoarthritis
(arcOGEN): a genome-wide association study

arcOGEN Consortium and arcOGEN Collaborators*  

Summary

Background Osteoarthritis is the most common form of arthritis worldwide and is a major cause of pain and disability in elderly people. The health economic burden of osteoarthritis is increasing commensurate with obesity prevalence and longevity. Osteoarthritis has a strong genetic component but the success of previous genetic studies has been restricted due to insufficient sample sizes and phenotype heterogeneity.

Methods We undertook a large genome-wide association study (GWAS) in 7410 unrelated and retrospectively and prospectively selected patients with severe osteoarthritis in the arcOGEN study, 80% of whom had undergone total joint replacement, and 11 099 unrelated controls from the UK. We replicated the most promising signals in an independent set of up to 7473 cases and 42 938 controls, from studies in Iceland, Estonia, the Netherlands, and the UK. All patients and controls were of European descent.

Findings We identified five genome-wide significant loci (binomial test p≤5·0x10⁻⁸) for association with osteoarthritis and three loci just below this threshold. The strongest association was on chromosome 3 with rs6976 (odds ratio 1·12 [95% CI 1·08–1·16]; p=7·24×10⁻¹¹), which is in perfect linkage disequilibrium with rs11177. This SNP encodes a missense polymorphism within the nucleostemin-encoding gene GNL3. Levels of nucleostemin were raised in chondrocytes from patients with osteoarthritis in functional studies. Other significant loci were on chromosome 9 close to ASTN2, chromosome 6 between FILIP1 and SENP6, chromosome 12 close to KLHDC5 and PTHLH, and in another region of chromosome 12 close to CHST11. One of the signals close to genome-wide significance was within the FTO gene, which is involved in regulation of bodyweight—a strong risk factor for osteoarthritis. All risk variants were common in frequency and exerted small effects.

Interpretation Our findings provide insight into the genetics of arthritis and identify new pathways that might be amenable to future therapeutic intervention.

Funding arcOGEN was funded by a special purpose grant from Arthritis Research UK.

Introduction

Osteoarthritis is the most common form of arthritis worldwide, affecting about 40% of people older than 70 years.¹ It is a complex disease of the musculoskeletal system with both genetic and environmental risk factors.² From the results of heritability studies in twins, sibling pairs, and families, genetic factors are estimated to account for about 50% of the risk of developing osteoarthritis in the hip or knee, although precise estimates vary according to sex, affected site, and severity of disease.³ Despite extensive efforts, only three loci (GDF5, chromosome 7q22, and MCF2L)³–⁶ have thus far been associated with osteoarthritis at genome-wide significance levels (p≤5·0x10⁻⁸) in European populations. This lack of success in osteoarthritis might be attributed to several factors such as insufficient sample sizes in previous studies and disease heterogeneity that might result from different underlying causes, both genetic and environmental, depending on which joints are affected.⁸ The three established osteoarthritis loci have fairly small effect sizes (allele-wise odds ratios [OR] of about 1·15) and we have previously shown that the genetic basis of osteoarthritis is likely to consist of several signals of similar or smaller magnitude that necessitate large sample sizes for their detection at genome-wide significance levels.⁹ Phenotype definition in cases and controls is also a likely factor contributing to the dilution of power to detect strong signals. Improved definition of phenotype and reduced misclassification in such a prevalent disorder can enhance power.

To identify additional loci that confer susceptibility to osteoarthritis, we undertook a large well powered genome-wide association study (GWAS) of osteoarthritis within the context of the Arthritis Research UK Osteoarthritis Genetics (arcOGEN) Consortium. In 2011, we reported the results of an interim analysis of 43·8% of the full GWAS sample size, detecting no novel loci for osteoarthritis, but showing the polygenic nature of the disease’s underlying genetics.¹⁰ Here, we report the results of the full-scale arcOGEN GWAS.

Methods

Study population

We undertook a large case–control genome-wide association study (GWAS) in which 7410 unrelated patients were retrospectively and prospectively selected from hospitals and clinics, with about 80% of patients ascertained for the severe endpoint of total joint
replacement [TJR], and publicly available data were obtained from 11009 unrelated controls from the UK’s Wellcome Trust Case Control Consortium 2 (WTCCC2), Avon Longitudinal Study of Parents and Children, Type 1 Diabetes Genetics Consortium, and People of the British Isles studies (table 1; appendix p 2). We replicated the most promising signals in an independent set of up to 7473 cases and 42938 controls from studies in Iceland (deCODE), Estonia (Estonian Genome Centre of University of Tartu [EGCUT]), the Netherlands (Genetics OsteoArthritis and Progression [GARP], Rotterdam study RSI and RSII cohorts), and the UK (TwinsUK). All patients and controls were of European descent. For further details, see appendix pp 4–5.

Procedures

The arcOGEN samples were genotyped by use of Illumina Human 610-Quad BeadChips (Illumina, San Diego, CA, USA; appendix p 2). Most cases had primary osteoarthritis requiring joint replacement of the hip or knee (table 1). We did quality-control checks both at the sample and single nucleotide polymorphism (SNP) levels (appendix p 2). We did a case–control analysis for an overlapping set of 485491 autosomal SNPs in 7410 arcOGEN cases of osteoarthritis and 11009 controls (table 1; appendix p 2, p 73). We imputed genotypes for autosomal SNPs by use of the directly typed data and phased genotype data from all HapMap III populations but did not see any additional signals (p<10–⁵) arising from imputed SNPs that were not captured by the directly genotyped SNP analyses (appendix p 3, p 6). To assess the effect of phenotype misclassification in controls, we also analysed genome-wide genotypes in 4476 female cases of osteoarthritis from arcOGEN compared with 1828 osteoarthritis-free female controls from the TwinsUK dataset. We used a multiplicative model for the analysis of all data and also stratified by site of osteoarthritis, sex, and osteoarthritis severity—ie, only cases with TJR were assessed (for a total of 24 specific phenotypes). We have presented the results for all analyses in the appendix pp 81–98, and for the most significant specific phenotype in the main text. 129 prioritised SNPs (p<10–⁵) were followed up by in-silico replication in the studies deCODE, EGCUT, GARP, Rotterdam study RSI and RSII cohorts, and TwinsUK (total in-silico replication sample size of 5064 cases of osteoarthritis and 40619 controls). We subsequently undertook a de-novo replication of the 26 most significant SNPs in an additional independent set of 2409 arcOGEN cases of osteoarthritis and 2319 WTCCC2 population-based controls (UK replication). For our primary analysis, we used a meta-analysis framework to combine results for all data (discovery and replication) by use of a fixed-effects model (appendix p 5).

We used RT-PCR to assess the expression of at least one gene per signal in joint tissues from patients with osteoarthritis who were undergoing TJR, and in cartilage from individuals without clinical osteoarthritis. We investigated the expression patterns of nucleostemin, encoded by the GNL3 gene, through immunohistochemical staining of normal and osteoarthritic

<table>
<thead>
<tr>
<th>Study description</th>
<th>Country of origin</th>
<th>Number of SNPs</th>
<th>Cases</th>
<th>Controls</th>
<th>Effective sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>arcOGEN Discovery</td>
<td>UK</td>
<td>485491</td>
<td>7410</td>
<td>Number</td>
<td>Women</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4476 (60.4%)</td>
<td>1606 (21.7%)</td>
</tr>
<tr>
<td></td>
<td>deCODE In silico</td>
<td>532234</td>
<td>4476</td>
<td>Number</td>
<td>Women</td>
</tr>
<tr>
<td></td>
<td>Iceland</td>
<td>128</td>
<td>2021</td>
<td>1179 (58.1%)</td>
<td>0 (100%)</td>
</tr>
<tr>
<td></td>
<td>GARP In silico</td>
<td>128</td>
<td>215</td>
<td>170 (79.1%)</td>
<td>169 (78.6%)</td>
</tr>
<tr>
<td></td>
<td>Netherlands</td>
<td>128</td>
<td>1950</td>
<td>1353 (69.4%)</td>
<td>1628 (83.5%)</td>
</tr>
<tr>
<td>RSI In silico</td>
<td>Netherlands</td>
<td>129</td>
<td>485</td>
<td>306 (63.1%)</td>
<td>398 (82.1%)</td>
</tr>
<tr>
<td></td>
<td>TwinsUK In silico</td>
<td>129</td>
<td>170</td>
<td>170 (100%)</td>
<td>170 (100%)</td>
</tr>
<tr>
<td>UK replication</td>
<td>De novo</td>
<td>24</td>
<td>2409</td>
<td>1453 (60.3%)</td>
<td>392 (16.3%)</td>
</tr>
</tbody>
</table>

Data are number (%), unless otherwise indicated. SNP= single nucleotide polymorphism. arcOGEN=Arthritis Research UK Osteoarthritis Genetics. EGCUT = Estonian Genome Centre of University of Tartu. See Online appendix. The table includes the following columns: Study description, Country of origin, Number of SNPs, Cases, and Controls. The table also includes the following sub-columns: Number, Women, Radiography, Joint replacement, Hip, Knee, Hip and knee, Definition, Number, Women.
cartilage samples and analysed protein expression in cultured chondrocytes from samples of normal and osteoarthritic cartilage by use of western blotting. Full details are available in the appendix (p 6).

**Statistical analysis**

Statistical and imputational methods and Hardy-Weinberg equilibrium are described in the appendix pp 2–5.

**Role of the funding source**

The sponsors of the study had no role in the study design, data collection, analysis or interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

**Results**

Genome-wide analysis showed an excess of signals compared with the null expectation of no association (genomic control inflation factor \( \lambda_{1000}=1.009 \); appendix p 74). In the main analysis (all cases of osteoarthritis vs population-based controls), 28 SNPs (representing 12 independent signals) had a p value of less than \( 1.0 \times 10^{-5} \) compared with five SNPs according to the null expectation (binomial test \( p=5.0 \times 10^{-13} \); appendix p 12, p 36). We checked for false positives that were attributable to population stratification by adjusting our analyses with the first ten ancestry-informative principal components: our results did not change qualitatively (appendix p 6, p 75). New signals were not detected with HapMap-based imputation across the genome (appendix p 3, p 6, pp 60–61).

After large-scale replication, 71 (55%) of 129 signals taken forward had effects in the same direction as the discovery analysis on the basis of which they had been prioritised (appendix pp 76–81). All 26 SNPs selected for further follow-up had effects in the same direction (binomial \( p=2.98 \times 10^{-9} \)). Five of 26 loci had genome-wide significance (\( p=5.0 \times 10^{-9} \)) and three were very close to this (table 2). Effect size estimates for seven of eight loci were lower in the replication analysis than in the discovery analysis (table 2), with the exception of one signal (ASTN2).

The most significant signal was on chromosome 3 and was followed up by two SNPs in perfect linkage disequilibrium (LD) with each other: rs11177 (allele A; figure A), a missense polymorphism within exon three of GNL3, coding for nucleostemin, and rs6976 (allele T; table 2), situated in the 3′ untranslated region (UTR) of the GLT8D1 gene. It was the only genome-wide significant signal in the discovery GWAS before replication (table 2), with the exception of one signal (ASTN2).

**Table 2:** Association summary statistics for the eight replicating signals

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Position</th>
<th>Nearest gene(s)</th>
<th>Effect allele</th>
<th>Discovery</th>
<th>Replication</th>
<th>Discovery and replication</th>
<th>Discovery and replication for analysis of all cases of osteoarthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>118306516</td>
<td>ASTN2</td>
<td>C</td>
<td>THR-female</td>
<td>1.19 1.34</td>
<td>1.23 1.34</td>
<td>1.18 1.25</td>
</tr>
<tr>
<td>12</td>
<td>27966237</td>
<td>RHD5, TP4H</td>
<td>A</td>
<td>THR</td>
<td>1.26 1.54</td>
<td>1.14 1.28</td>
<td>1.12 1.25</td>
</tr>
<tr>
<td>12</td>
<td>103584897</td>
<td>GSTT1</td>
<td>G</td>
<td>THR</td>
<td>1.26 1.48</td>
<td>1.14 1.28</td>
<td>1.12 1.25</td>
</tr>
<tr>
<td>3</td>
<td>191082854</td>
<td>TP63</td>
<td>G</td>
<td>THR-female</td>
<td>1.30 1.42</td>
<td>1.14 1.28</td>
<td>1.12 1.25</td>
</tr>
<tr>
<td>16</td>
<td>52396636</td>
<td>FTO</td>
<td>C</td>
<td>Female</td>
<td>1.30 1.50</td>
<td>1.14 1.28</td>
<td>1.12 1.25</td>
</tr>
<tr>
<td>6</td>
<td>44885669</td>
<td>SUPT3H, CDCS1</td>
<td>Male</td>
<td>1.30 1.50</td>
<td>1.14 1.28</td>
<td>1.12 1.25</td>
<td>1.12 1.25</td>
</tr>
</tbody>
</table>

The details of the individual discovery and replication results of analyses of all cases of osteoarthritis are shown in the appendix p 99. TJR=total joint replacement. THR=total hip replacement. TKR=total knee replacement. arcOGEN=Arthritis Research UK Osteoarthritis Genetics. ∗Represent the same signal, †both were prioritised and showed similar p values and effect sizes before and after replication. ‡Analyses in which a subset of samples with osteoarthritis at the hip and knee are also included in the arcOGEN discovery set and in the UK replication set. †‡This signal was attenuated after adjustment for body-mass index, suggesting that the FTO locus exerts its effect on osteoarthritis through obesity.
Figure: Regional association plots of replicating signals
Case-control association results (−log_10[p value]) for genotyped SNPs in the discovery set are plotted against genomic position (National Center for Biotechnology Information build 36) for the stratum in which the most significant meta-analysis p value occurred. The index SNP is denoted by a purple diamond in the discovery set and by a purple square in the final meta-analysis. The circles indicate association results of genotyped SNPs in the region; the colour reflects the correlation coefficient (r^2) of each genotyped SNP with the index SNP estimated with the CEU HapMap II panel. Estimated recombination rates (in cM/Mb) are plotted in red. The region shown in the plots extends to either 500 kb upstream and downstream of the index SNP or until the next recombination hotspot if this lies further than a distance of 500 kb.

(A) Chromosome 3 signal centred on rs6976 in total joint replacement. (B) Chromosome 9 signal centred on rs4836732 in female total hip replacement. (C) Chromosome 6 signal centred on rs9350591 in hip osteoarthritis. (D) Chromosome 12 signal centred on rs10492367 in hip osteoarthritis. (E) Chromosome 12 signal centred on rs835487 in total hip replacement. SNP=single nucleotide polymorphism. CEU=Utah residents with ancestry from northern and western Europe.
All the remaining four genome-wide significant signals emanated from hip-specific strata (table 2): rs4836732 (allele C; female total hip replacement (THR)) located within intron 18 of the ASTN2 gene (figure B; table 2); rs9350591 (allele T; hip osteoarthritis) located 38 kb upstream of FILIP1 and 70 kb upstream of SENP6 (figure C; table 2); rs10492367 (allele T; hip osteoarthritis) 59 kb downstream of KILDC5 and 96 kb downstream of PTHLH (figure D; table 2); and rs835487 (allele G; THR) located within intron two of CHST11 (figure E; table 2).

Three additional signals lacked genome-wide significance and therefore need to be interpreted with caution. Rs12107036 was associated with osteoarthritis in women with total knee replacement (TKR; allele G) and is located within intron 12 of TP63 (appendix p 70; table 2). Association with rs8044769 was strongest in the female osteoarthritis stratum (allele C), within intron 1 of FTO (appendix p 71; table 2). The final signal was seen in the male osteoarthritis stratum at rs10948172 (allele G), situated in the vicinity of the SUPT3H gene (appendix p 72; table 2).

We investigated the effect of the definition of disease phenotype on the strength of association of the replicating signals (table 2) by comparing the results of analyses of TJR-only cases (severe endpoint of disease) and all cases (TJR and radiography combined) for the stratum that had the lowest meta-analysis p value. We noted that, after large-scale replication, four signals (rs6976, rs4836732, rs835487, and rs12107036) were stronger in the TJR meta-analysis, one signal (rs9350591) was stronger in the meta-analysis of TJR and radiographic cases, and three signals (rs10492367, rs8044769, and rs10948172) were little changed (appendix p 100).

In the discovery set analyses of two types of controls—population-based controls unselected for osteoarthritis (n=11 009) and osteoarthritis-free controls (n=1828)—we compared the strength of association of the three previously established osteoarthritis loci (appendix p 7, p 101) and the eight replicating signals (table 2). We noted that one signal (GDF5) was stronger in the analysis of osteoarthritis-free controls, whereas the other ten were stronger or similar in the analyses of population-based controls (appendix p 7, p 101, p 103). In the meta-analysis of discovery and replication sets, effect size estimates for the replicating loci were the same or smaller for all variants except one (CHST11) when we used disease-free controls, and all signals were statistically stronger in the analyses of population-based controls (appendix p 7, p 104).

Because obesity is an established risk factor for osteoarthritis, we investigated whether any of the 129 prioritised signals were attenuated after adjustment for body-mass index (BMI) with arcOGEn cases and disease-free TwinsUK control data (BMI data were not available for the population-based controls). Two SNPs (rs11107957 and rs2626053) showed greater than 100-times increase in their p value (ie, became less significant) after adjustment for BMI (appendix p 5, p 105). Rs804476 in FTO was the only replicating signal to be attenuated after BMI adjustment in the discovery GWAS, but we cannot exclude the possibility that BMI adjustment might have had an effect on the other variants in the replication cohorts. However, in the discovery scan, the association p value at the four remaining loci that had genome-wide significance and the three signals just below this threshold were not affected by more than one order of magnitude after adjustment for BMI.

All the studied genes were expressed within osteoarthritic or control fracture neck-of-femur joint tissues (appendix p 6, p 108).

Immunohistochemical staining showed strong nucleolar expression of nucleostemin in cytosins of primary cultured chondrocytes from human articular cartilage derived from patients with osteoarthritis (appendix p 69), and a similar pattern of staining was seen in chondrocytes from frozen sections of osteoarthritic cartilage (appendix p 69). We measured levels of nucleostemin protein expression in cultured articular chondrocytes from five controls and five patients with osteoarthritis. Nucleostemin was barely detectable in cultured chondrocytes from controls but was clearly detectable in cultured chondrocytes from patients with osteoarthritis. Nucleostemin expression was corrected for expression of the housekeeping gene β-actin, the upregulation of nucleostemin protein levels in osteoarthritis chondrocytes versus control chondrocytes was significant (p=0.003; appendix p 69).

Discussion

We have identified eight novel loci that are associated with the risk of developing osteoarthritis, five of which surpassed genome-wide significance and three that were just less than the threshold (panel). With the addition of previously established associations on chromosome 7q22, and in the GDF5 and MCF2L genes, the total number of osteoarthritis susceptibility loci in European populations is 11.

Our most significant finding was in a gene-rich region of chromosome 3p21.1. Rs6976 lies within the 3’ UTR of GLT8D1 (glucosyltransferase 8 domain containing 1; figure A). Alterations in the expression of GLT8D1 could affect the glycosylation of cartilage proteins. Rs11177, in perfect LD with rs6976, encodes a missense variant (Arg→Gln) in the third exon of GLN3 (guanine nucleotide binding protein-like 3, or nucleostemin; figure A). GLN3 is expressed in mesenchymal stem cells, from which chondrocytes are derived, and regulates the G1–S phase transition in stem cells. 13–15 The function of GLN3 in bone and cartilage is not known because mice homozygous for deletion of GLN3 die on embryonic day 4.16 However, of interest is that nucleostemin protein levels were...
substantially increased in cultured chondrocytes from patients with osteoarthritis compared with controls, raising the possibility that this gene might be functionally important in the pathogenesis of osteoarthritis.

Rs4836732 lies within the ASTN2 gene (astrotactin 2; figure B), which is highly expressed in the developing and adult brain. ASTN2 is a membrane protein that regulates surface levels of ASTN1 during neuronal migration.7 Rs9350591 lies between the FILIP1 (filamin A interacting protein 1) and SENP6 (sentrin specific peptidase 6) genes (figure C). COL12A1 (collagen, type XII, alpha 1), the product of which is found in articular cartilage,18 lies at a distance of about 326 kb. Rs10492367 is located between the KLHDC5 (Kelch domain containing 5) and PTHLH (parathyroid hormone-like gene) genes (figure D). Parathyroid hormone-related protein (PTHrP) regulates endochondral bone development. Results from studies of Pthrpr–/– mice showed that animals who survived gestation had accelerated differentiation of chondrocytes in bone.19 Rs835487 is located in the CHST11 (carbohydrate sulfotransferase 11) gene (figure E), which encodes a Golgi enzyme that catalyses the transfer of sulphate groups to the 4-O position of chondroitin and dermatan sulphate, important components of cartilage proteoglycans.20 CHST11 has a role in skeletal development, signalling pathways, and cancer progression.21 It has substantially higher expression in osteoarthritic cartilage than in normal articular cartilage22 and mice that are null for this gene have a disorganised growth plate, accelerated chondrocyte differentiation, altered patterns of signalling with transforming growth factor β and bone morphogenetic protein, and a fibrillated cartilage extracellular matrix that is characteristic of osteoarthritic cartilage.23 Chondroitin sulfate is used as a symptomatic, slow-acting drug for osteoarthritis, recommended by the latest Osteoarthritis Research Society International treatment guidelines, but evidence for its effectiveness remains controversial.24 Rs12107036 lies within intron 12 of the TP63 gene (encoding tumour protein p63; appendix p 70). p63 null mice have serious defects in their limb, craniofacial, and epithelial development.25,26 Rs8044769 lies within FTO (appendix p 71), a well established locus that is associated with fat mass and obesity27 and is in partial LD (r²>0·6) with the reported index BMI-associated SNPs. This signal was attenuated after adjustment for BMI, suggesting that the FTO gene exerts its effect on osteoarthritis through obesity. Rs10948172 occurs between CDC5L (CDC5 cell division cycle 5-like) and SUPT3H genes (suppressor of Ty3 homologue; appendix p 72). At a distance of 500 kb, but possibly containing SNPs correlated with rs10948172, is the RUNX2 (runx-related transcription factor 2) gene, which encodes a protein that is essential for osteoblast differentiation28 and skeletal morphogenesis.29 RUNX2 was identified in peripheral blood expression profiles as a possible biomarker for bone metabolism in arthritis.30 The results of our study draw attention to three important issues with respect to design and analysis that are of relevance to GWAS investigations of complex disease: definition of significant association in view of several correlated stratified analyses; effect of phenotype homogeneity and misclassification; and allowing for power constraints when seeking replication of previously identified loci. First, we analysed several specific non-independent phenotypes in addition to the main analysis of all osteoarthritis cases versus controls so as to increase power to detect signals specific for joint, sex, or disease severity. Compelling evidence suggests that joint-specific genetic factors play a part in the pathogenesis of osteoarthritis31 and that the familial concordance for hip and knee osteoarthritis is greater in surgically defined32,33 than in radiographically defined disease.34 The results of these reports, in addition to the significant differences reported in the prevalence of osteoarthritis between skeletal sites, the sexes, and radiographic versus symptomatic disease,35 underpin our decision to stratify our data by site of osteoarthritis (hip or knee), sex, sex with site, and the two modes of osteoarthritis definition (radiographic and joint replacement) that we have used. That most of the loci we have identified are associated with osteoarthritis at a particular joint site provides genetic
GWAS. We detected signals with $p=0.002$ to $p=6\times10^{-4}$ for (appendix pp 101–04). Patient phenotype homogeneity (appendix p 100). All eight loci reported in this study were prioritised on the basis of analyses of osteoarthritis-status-agnostic associations with other common complex trait studies, these findings can be ascribed to a combination of power, stochastic variation in allele frequency estimates, and the absence of the index variant from the genotyping platform (appendix p 7, p 101).

We have studied common variants and therefore the findings of our report cannot be used to estimate the effect that rarer variants might have on causing osteoarthritis. Further, although our discovery sample size of 7410 cases is the largest yet used for osteoarthritis, the sample was stratified; an analysis of larger stratified subsets might show additional associated loci. In our study, we focused on Europeans and it will be of great interest to assess whether these loci have global relevance to osteoarthritis risk through similar studies of non-European cohorts with osteoarthritis.

All eight identified loci are represented by common SNPs (minor allelic frequency 0·11–0·50) and have small effect sizes (allelic ORs from 1·11 to 1·21). In keeping with other common complex trait studies, these findings indicate that common SNPs with large effect sizes are not likely to have a role in the pathogenesis of osteoarthritis. Association at the index SNPs reported here does not imply causality. Functional studies will be necessary to pinpoint the precise functional variants. The eight signals identified in this study have a combined sibling recurrence risk ($\lambda_s$) estimate of 1·028 and account for only a small fraction of familial clustering of osteoarthritis. This outcome leads us to advocate larger-scale sample sizes, typically achieved through GWAS meta-analysis efforts in complex traits. The arcOGEN GWAS is the largest study of osteoarthritis genetics so far and has sufficient power to detect slight effects at common loci at the genome-wide significance level (eg, 80% power to detect an allele with frequency 0·35 and allelic OR of 1·15). We have established novel loci represented by common SNPs that confer a slight risk for osteoarthritis and are associated with the clinically important phenotype of TJR. These results provide a basis for functional studies to identify the underlying causative variants, biological networks, and molecular cause of osteoarthritis.

**Contributors**

EZ, KP, LS, GAW, JWJ, and JL participated in the writing group. EZ, KC, PD, WERO, GAW, Nia, AC, MD, AMcC, JWJ, SHR, AMV, TDS, and JL participated in the arcOGEN study design. EZ, KP, LS, NWR, AGD-W, MCL, VB, CB, KSE, SYS, VGY, ZG, MM, and SaM participated in data analysis and interpretation. Nia, AC, MD, AMcC, JWJ, SHR, TDS, JL, FB, AR, EA, NaA, P-KB, IC, SD, AG, Jj, RR, NCK, SHF, FON, EP, MRR, DI, KW, BW, and MW participated in arcOGEN patient and data collection. WERO, KSH, and KD participated in sample DNA processing. SuB and PD participated in genotyping for the arcOGEN study. HB, SEH, and SCP provided informatics support for the arcOGEN study. TI, HJ, GT, KSt, and US participated in data replication from the deCODE study. TE and AM participated in data replication from the EGCUT study. JJH D, MX, SB, PES, and IM participated in data replication from the GARP study. AH, HjMK, FR, AGU, and JbVM participated in data replication from the Rotterdam study. EE, MM, SaM, JPAI, AMV, and TDS participated in data replication from the TwinsUK study. SHR, EVAR, MaR, MiR, and DoS participated in functional studies.

**arcOGEN Consortium and arcOGEN Collaborators**


**arcOGEN Investigators**

Sanger Institute, Hinxton, UK (E Zeggini PhD, K Panoutsopoulou PhD, L Southam BSc, A G Day-Williams PhD, M C Lopes PhD, V Boraska PhD, C Beazley PhD, H Blackburn MSc, S Bumpstead BSc, S E Hunt BSc, S C Potter PhD, S-Y Shin PhD, V K Yadav PhD, P Deloukas PhD); Welcomes Trust Centre for Human Genetics, University of Oxford, Oxford, UK (L Southam BSc, N W Rayner PhD, M Lopes PhD, K S Elliott PhD); Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK (N W Rayner PhD); University of Split School of Medicine, Split, Croatia (V Boraska PhD);
References


37 Wellcome Trust Case Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; 447: 661–78.