A Variant in MCF2L Is Associated with Osteoarthritis

Aaron G. Day-Williams,1 Lorraine Southam,1,2 Kalliопе Panoutsopoulou,1 Nigel W. Rayner,2 Tonu Esko,3,4,5 Karol Estrada,6,7 Hafdis T. Helgadottir,8 Albert Hofman,9 Throvaldur Ingvarsson,10,11 Helgi Jonsson,11,12 Aime Keis,13,14 Hanneke J.M. Kerkhof,6,7 Gudmar Thorleifsson,8 Nigel K. Arden,15,16 Andrew Carr,17 Kay Chapman,17 Panos Deloukas,1 John Loughlin,18 Andrew McCaskie,18,19 William E.R. Ollier,20 Stuart H. Ralston,21 Timothy D. Spector,22 Gillian A. Wallis,23 J. Mark Wilkinson,24,25 Nadim Aslam,26 Fraser Birell,18,27 Ian Carluke,27 John Joseph,28 Ashok Rai,29 Mike Reed,27 Kirsten Walker,27 arcOGEN Consortium, Sally A. Doherty,30 Ingileif Jonsdottir,8,11 Rose A. Maciewicz,31 Kenneth M. Muir,32 Andres Metspalu,3,4,5 Fernando Rivadeneira,6,7,9 Kari Stefansson,5,8,11 Unnur Styrkarsdottir,8 Andre G. Uitterlinden,6,7,9 Joyce B.J. van Meurs,6,7 Weiya Zhang,30 Ana M. Valdes,22 Michael Doherty,30 and Eleftheria Zeggini,1

Osteoarthritis (OA) is a prevalent, heritable degenerative joint disease with a substantial public health impact. We used a 1000-Genomes-Project-based imputation in a genome-wide association scan for osteoarthritis (3177 OA cases and 4894 controls) to detect a previously unidentified risk locus. We discovered a small disease-associated set of variants on chromosome 13. Through large-scale replication, we establish a robust association with SNPs in MCF2L (rs11842874, combined odds ratio [95% confidence interval] 1.17 [1.11–1.23], p = 2.1 × 10−12) across a total of 19,041 OA cases and 24,504 controls of European descent. This risk locus represents the third established signal for OA overall. MCF2L regulates a nerve growth factor (NGF), and treatment with a humanized monoclonal antibody against NGF is associated with reduction in pain and improvement in function for knee OA patients.

Osteoarthritis (OA) is the most common form of arthritis and is associated with a large health economic burden.1 The sibling recurrence risk (p) for OA has been estimated to be approximately 5 in the UK.1 Two loci (GDF5 [MIM 601146] on chromosome 20 and a signal on chromosomal region 7q22, both with allelic odds ratios of ~1.15) have reached genome-wide significance in European populations.2–5 This paucity of established risk loci could be ascribed to limitations caused by insufficient sample sizes, phenotype heterogeneity, resolution of known variation, associations with low-frequency and/or rare variants, interaction effects, or structural variation.6–7 We recently carried out a large genome-wide association scan (GWAS) restricted to knee and/or hip OA and detected no replicating signals (arcOGEN GWAS).8 Imputation based on the 1000 Genomes Project (1KGP) has been proposed as an approach that will increase power and resolution in genetic association studies,9 and researchers have already applied the technique to fine map known association signals.10,11 In this work, we applied a 1KGP-based imputation and identify a genome-wide significant locus for OA within a gene previously unlinked to the disease.

We used 1KGP pilot 1 data of 60 CEU individuals as a reference set and imputed 1KGP-identified variants into the arcOGEN GWAS of 3177 cases and 4894 UK controls12–14 (Figure 1). The set of 3177 OA cases are unrelated individuals of European ancestry collected in the UK on the basis of two criteria: (1) radiographic evidence of disease (defined as a Kellgren-Lawrence [KL] grade ≥ 2)5 and/or (2) clinical evidence of disease requiring joint replacement (TJR). The 4894 UK-population-based controls were unrelated individuals from the 1958 British

1Wellcome Trust Sanger Institute, Hinxton CB10 1HH, UK; 2Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; 3Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 4Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia; 5Estonian Biocenter, Tartu 51010, Estonia; 6Department of Internal Medicine, Erasmus Medical Centre, Rotterdam 3000 DR, The Netherlands; 7The Netherlands Genomics Initiative–Sponsored Netherlands Consortium for Healthy Aging, Leiden 2300 RC, The Netherlands; 8deCODE Genetics, Reykjavik IS-101, Iceland; 9Department of Epidemiology, Erasmus Medical Center, Rotterdam 3000 CA, The Netherlands; 10Institute of Health Sciences, Akureyri 600, Iceland; 11University of Iceland, Faculty of Medicine, Reykjavik 101, Iceland; 12Landspitali University Hospital, Department of Medicine, Reykjavik 101, Iceland; 13Department of Public Health, University of Tartu, Tartu 51010, Estonia; 14Orthopedic Surgeons, Elva Hospital, Elva 61501, Estonia; 15National Institute of Health Research Biomedical Research Unit, University of Oxford, Oxford OX3 7LD, UK; 16Medical Research Council Epidemiology Research Centre, University of Southampton, Southampton SO16 7LB, UK; 17Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK; 18Institute of Cellular Medicine, Musculoskeletal Research Group, Newcastle University, Newcastle upon Tyne NE4 5DD, UK; 19The Newcastle upon Tyne Hospitals NHS Trust Foundation, The Freeman Hospital, Newcastle upon Tyne NE7 7DN, UK; 20Centre for Integrated Genomic Medical Research, University of Manchester, Manchester M13 9PT, UK; 21Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2UX, UK; 22Department of Twin Research and Genetic Epidemiology, King’s College London, London SE1 7EH, UK; 23Wellcome Trust Centre for Cell Matrix Research, University of Manchester, Manchester M13 9PT, UK; 24Academic Unit of Bone Metabolism, Department of Human Metabolism, University of Sheffield, Sheffield S5 7AU, UK; 25Sheffield NIHR Bone Biomedical Research Unit, Centre for Biomedical Research, Northern General Hospital, Sheffield S5 7AU, UK; 26Orthopaedic Department, Worcestershire Acute Hospitals NHS Trust, Worcester WR5 1DD, UK; 27Wansbeck General Hospital, Northumbria Healthcare NHS Foundation Trust, Ashington NE63 9JF, UK; 28Worcestershire Acute Hospitals NHS Trust, Worcester WR5 1DD, UK; 29Rheumatology Department, Worcestershire Acute Hospitals NHS Trust, Worcester WR5 1DD, UK; 30Academic Rheumatology, University of Nottingham, Nottingham NG5 1PB, UK; 31Respiratory and Inflammation Research Area, AstraZeneca, Charnwood R&D, Loughborough LE11 5RH, UK; 32Health Sciences Research Institute, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK

*Correspondence: eleftheria@sanger.ac.uk

DOI 10.1016/j.ajhg.2011.08.001. ©2011 by The American Society of Human Genetics. All rights reserved.

446  The American Journal of Human Genetics 89, 446–450, September 9, 2011
Birth Cohort (58BC) and the UK National Blood Donor Service (UKBS) and were obtained from an early release of the Wellcome Trust Case Control Consortium 2 (WTCCC2) data. The genotyping and quality control (QC) of these individuals and their genotype data were described previously in the initial arcOGEN GWAS.8 Our primary 1KGP imputation was based on the April 2009 release of haplotypes for 57 individuals. After removing rare variants (with minor allele frequency [MAF] < 0.01) and SNPs with low imputation quality (r² < 0.3), 7,258,070 variants were tested for association with OA.

Further quality control was applied by closely examining signals with p<10^{-5} and removing low quality variants. Seven out of the eight SNPs were successfully typed with a Sequenom MassArray iPLEX Gold assay (Table S1) and one SNP, rs11842874 on 13q34, replicated with p = 2.60 × 10^{-3} (allelic odds ratio [OR] 1.17 [1.06–1.30]) and with consistent effect direction as the original scan (Figure 2, Table 1). We subsequently took this signal forward to de novo genotyping in two further sample sets from the UK: the Genetics of Osteoarthritis and Lifestyle (GOAL) study16,17 (1686 total joint replacement cases, 743 non-OA controls) and an additional independent set of 2409 newly recruited arcOGEN cases and 2319 population-based controls from the 58BC and UKBS cohorts. The combined UK meta-analysis (n = 12,437 cases, 14,111 controls) allelic OR was 1.22 [1.14–1.30], p = 2.24 × 10^{-8}. We further investigated association with this variant in four non-UK OA sample sets: two from the Netherlands (Rotterdam Study I [RSI], 1950 cases and 3243 controls, and Rotterdam Study II [RSII], 485 cases and 1460 controls, both in silico),18–20 one sample set from Estonia (Estonian Genome Center, University of Tartu [EGCUT], 2617 cases and 2619 controls, de novo genotyping),21,22 and one from Iceland (deCODE, 1552 cases and 3071 controls, de novo genotyping). We used a meta-analysis framework to combine results across the follow-up studies only and across all data. We obtained the combined estimates of ORs for reference alleles by weighting the logORs of each study by the inverse of their variance via a fixed effects model. We investigated evidence of heterogeneity of ORs by using the Cochran’s Q and I² statistics. The meta-analysis was performed with the GWAMA software package.23 In all seven follow-up datasets combined, rs11842874 was associated with OA with p = 3.0 × 10^{-5} (allelic OR 1.13 [1.07–1.20]). Combined with the discovery sample set, the overall fixed effects meta-analysis (across 19,041 cases and 24,504 controls) established association at this variant with p = 2.07 × 10^{-8} (allelic OR 1.17 [1.11–1.23]; Figure 2, Table 1). The variant appears to be more strongly associated with knee OA (allelic OR 1.17 [1.10–1.25], p = 2.52 × 10^{-6}, effective sample size of 28,987) than with hip OA (allelic OR 1.11 [1.03–1.19], p = 3.54 × 10^{-3}, effective sample size of 27,452). Studies contributing data to this manuscript acquired informed consent from all participants and were approved by the appropriate ethics committee(s) for the respective institutions and countries.

rs11842874 is one of several highly correlated SNPs at 13q34 and constitutes the observed association signal, which spans 12.7 kb (Figure 3). The surrounding 1 Mb region is characterized by low levels of linkage disequilibrium and contains only nine SNPs correlated with rs11842874 at r² > 0.7. rs11842874 was selected for replication because it is the only variant present on some GWAS platforms. The OA risk-increasing allele is the major allele with population frequency of 0.927 (mean over the 446–450, September 9, 2011
human cells *MCF2L* regulates neurotrophin-3-induced cell migration in Schwann cells.\textsuperscript{26} Neurotrophin-3 is a member of the nerve growth factor (NGF) family. Treatment of knee OA patients with a humanized monoclonal antibody that inhibits NGF was found to be associated with joint pain reduction and an improvement in function.\textsuperscript{27,28}

The *MCF2L* OA locus was taken forward on the basis of evidence accrued through 1KGP-based imputation. Direct typing of \~600,000 SNPs through GWAS resulted in modest ($p > 10^{-5}$) evidence for the association of a single variant, rs11842874, with OA in this region and HapMap-based imputation left the picture unchanged (Figures 3A and 3B). Prioritization strategies for follow-up in our published GWAS\textsuperscript{8} down-weighted lone variants with no corroboration of association from neighboring SNPs. The denser 1KGP reference set empowered the association of multiple additional correlated (i.e., non-independent), imputed variants, several of which showed stronger evidence for association with OA, thus highlighting this region for validation and replication genotyping (Figure 3C, Table S1).

The associated variants are common, but their minor allele frequencies are toward the lower end of the frequency spectrum (at \~0.07). The identification of similar variants with modest effect sizes (OR 1.17) at genome-wide significance levels will require sample sizes in the order of \~23,000 cases and an equal number of controls. Through several rounds of cluster-plot inspection, removal of poor quality SNPs, and re-imputation, we observed that the

<table>
<thead>
<tr>
<th>Study</th>
<th>OR (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>arcOGEN GWAS</td>
<td>1.32 (1.16-1.50)</td>
<td>1.67e-05</td>
</tr>
<tr>
<td>arcOGEN Follow-up Set 1</td>
<td>1.17 (1.06-1.30)</td>
<td>2.60e-03</td>
</tr>
<tr>
<td>GOAL</td>
<td>1.24 (0.99-1.56)</td>
<td>7.20e-02</td>
</tr>
<tr>
<td>arcOGEN Follow-up Set 2</td>
<td>1.16 (0.98-1.37)</td>
<td>7.86e-02</td>
</tr>
<tr>
<td>RSI</td>
<td>0.98 (0.83-1.17)</td>
<td>9.61e-01</td>
</tr>
<tr>
<td>RSII</td>
<td>1.46 (1.07-2.00)</td>
<td>1.60e-02</td>
</tr>
<tr>
<td>EGCUT</td>
<td>1.16 (1.01-1.34)</td>
<td>4.01e-02</td>
</tr>
<tr>
<td>deCODE</td>
<td>1.03 (0.98-1.20)</td>
<td>7.31e-01</td>
</tr>
<tr>
<td>Meta Analysis</td>
<td>1.17 (1.11-1.23)</td>
<td>2.07e-08</td>
</tr>
</tbody>
</table>

Figure 2. Forest Plot of OA Association and Meta-Analysis Results for rs11842874
The purple squares designate the estimated Odds Ratio (OR) for each individual study and the error bars extending out on both sides show the 95% confidence interval for the OR estimate. The pink diamond designates the estimated OR of the fixed effects meta-analysis of all the studies.

Table 1. OA Association and Meta-Analysis Results for rs11842874

<table>
<thead>
<tr>
<th>Study</th>
<th>Number Cases</th>
<th>Number Controls</th>
<th>Effect Allele</th>
<th>MAF</th>
<th>OR (95% Confidence Interval)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>arcOGEN GWAS</td>
<td>3177</td>
<td>4894</td>
<td>A</td>
<td>0.0718</td>
<td>1.32 (1.16-1.50)</td>
<td>1.67e-05</td>
</tr>
<tr>
<td>arcOGEN replication set 1</td>
<td>5165</td>
<td>6155</td>
<td>A</td>
<td>0.0694</td>
<td>1.17 (1.06-1.30)</td>
<td>2.60e-03</td>
</tr>
<tr>
<td>GOAL</td>
<td>1686</td>
<td>743</td>
<td>A</td>
<td>0.0720</td>
<td>1.23 (0.99-1.56)</td>
<td>7.20e-02</td>
</tr>
<tr>
<td>arcOGEN replication set 2</td>
<td>2409</td>
<td>2319</td>
<td>A</td>
<td>0.0636</td>
<td>1.16 (0.98-1.37)</td>
<td>7.86e-02</td>
</tr>
<tr>
<td>UK meta-analysis</td>
<td>12437</td>
<td>14111</td>
<td>A</td>
<td>0.0917</td>
<td>1.42 (1.14-1.80)</td>
<td>2.24e-08</td>
</tr>
<tr>
<td>deCODE</td>
<td>1552</td>
<td>3071</td>
<td>A</td>
<td>0.0917</td>
<td>1.03 (0.88-1.20)</td>
<td>7.31e-01</td>
</tr>
<tr>
<td>EGCUT</td>
<td>2617</td>
<td>2619</td>
<td>A</td>
<td>0.0769</td>
<td>1.16 (1.01-1.34)</td>
<td>4.01e-02</td>
</tr>
<tr>
<td>RSI</td>
<td>1950</td>
<td>3243</td>
<td>G</td>
<td>0.0608</td>
<td>1.01 (0.86-1.20)</td>
<td>8.61e-01</td>
</tr>
<tr>
<td>RSII</td>
<td>2409</td>
<td>2319</td>
<td>A</td>
<td>0.0715</td>
<td>1.46 (1.07-2.00)</td>
<td>1.68e-02</td>
</tr>
<tr>
<td>Non-UK meta-analysis</td>
<td>6604</td>
<td>10393</td>
<td>A</td>
<td>1.09</td>
<td>1.00 (1.00-1.19)</td>
<td>4.70e-10</td>
</tr>
<tr>
<td>Combined meta-analysis</td>
<td>19041</td>
<td>24504</td>
<td>A</td>
<td>1.17</td>
<td>1.11 (1.11-1.23)</td>
<td>2.07e-08</td>
</tr>
</tbody>
</table>

Meta-analysis results are denoted in bold.
The majority of signals were caused by genotyping and imputation artifacts. This fact highlights the need for imputed signals to be scrutinized postimputation before follow-up studies are deployed. As the field of complex trait association studies shifts its focus toward low-frequency and rare variants, thorough quality control of signals becomes highly relevant. In this study, we have restricted 1KGP-based imputation to variants with an MAF > 0.01. As reference panel sizes become larger, imputation of lower frequency variants will become more feasible, empowering the examination of rare variation in next generation association studies.

The genetic architecture of OA has not been elucidated yet. By identifying this susceptibility locus, the third one discovered for OA, our study now provides a foundation on which functional studies can be based. New additions to the genetic study toolset, including larger sample sets, well-characterized phenotypes, and resequenced reference-panel-based imputation approaches hold the promise of providing insights into the etiology of this common degenerative joint disease.

Supplemental Data

Supplemental Data include one table and the Acknowledgments and can be found with this article online at [http://www.cell.com/AJHG/](http://www.cell.com/AJHG/).

Acknowledgments

The collection of GOAL samples was funded by AstraZeneca UK. The author affiliated with AstraZeneca is an employee of AstraZeneca, a global research-based biopharmaceutical company focused on discovering, developing and marketing medicines for some of the world’s most serious illnesses, and owns stocks or stock options and has a pending patent application in the company. The authors that are affiliated with deCODE genetics are all employees of deCODE, a biotechnology company that provides genetic testing services, and some own stocks or stock options in the company. Additional acknowledgments are in the Supplemental Data.

Received: June 15, 2011
Revised: July 28, 2011
Accepted: August 1, 2011
Published online: August 25, 2011
Web Resources
The URLs for data presented herein are as follows:

1000 Genomes Project, http://www.1000genomes.org
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
Wellcome Trust Case Control Consortium 2, http://www.wtccc.org.uk

References