White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/95302/

Version: Published Version

Article:
Panoutsopoulou, K., Southam, L., Elliott, K.S. et al. (2010) Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. Annals of the Rheumatic Diseases (ARD), 70 (5). pp. 864-867. ISSN: 0003-4967

https://doi.org/10.1136/ard.2010.141473

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study

K Panoutsopoulou,[1] L Southam,[2] K S Elliott,[3] N Wrayner,[3,4] G Zhai,[5] C Beazley,[1] G Thorleifsson,[6] N K Arden,[7,8] A Carr,[2] K Chapman,[2] P Deloukas,[1] M Doherty,[9] A McCaskie,[10,11] W E R Ollier,[12] S H Ralston,[13] T D Spector,[5] A M Valdes,[5] G A Wallis,[14] J M Wilkinson,[15,16] E Arden,[17] K Battley,[18] H Blackburn,[1] F J Blanco,[18] S Bumpstead,[1] L A Cupples,[19] A G Day-Williams,[1] K Dixon,[12] S A Doherty,[9] T Esko,[20–22] E Evangelou,[23] D Felson,[24] J J Gomez-Reino,[25,26] A Gonzalez,[27] A Gordon,[15] R Gwilliam,[1] B V Halldorsson,[6,28] V B Hauksson,[6] A Hofman,[29,30] S E Hunt,[1] J P A Ioannidis,[23,31–33] T Ingvarsson,[34,35] I Jonsdottir,[6,35] H Jonsson,[35,36] R Keen,[37] H J M Kerkhof,[29,30] M G Kloppenburg,[38] N Koller,[13] N Lakenberg,[39] N E Lane,[40] A T Lee,[41] A Metspalu,[20–22] I Meulenbelt,[30,39] M C Nevitt,[42] F O'Neill,[5] N Parimi,[43] S C Potter,[1] I Rego-Perez,[18] J A Riancho,[44] K Sherburn,[12] P E Slagboom,[30,39] K Stefansson,[6,35] U Styrkarsdottir,[6] M Sumillera,[45] D Swift,[15,16] U Thorsteinsdottir,[6,35] A Tsezou,[46] A G Uitterlinden,[29,30] J B J van Meurs,[29,30] B Watkins,[2] M Wheeler,[9] S Mitchell,[11] Y Zhu,[24] J M Zmuda;[47] arcOGEN Consortium, E Zeggini,[1,3] J Loughlin[10]

► Supplementary materials and methods, results and figures and tables are published online only. To view these files please visit the journal online at (http://ard.bmj.com).

For numbered affiliations see end of article

Correspondence to
Dr J Loughlin, Newcastle University, Institute of Cellular Medicine, 4th Floor Catherine Cookson Building, Newcastle University, Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, UK; john.loughlin@ncl.ac.uk and E Zeggini, Wellcome Trust Sanger Institute, The Morgan Building, Wellcome Trust Genome Campus, Cambridge CB10 1HH, UK, eleftheria@sanger.ac.uk

KP, LS, EZ and JL contributed equally to the study.

Accepted 11 November 2010
Published Online First 21 December 2010

UNLOCKED
This paper is freely available online under the BMJ Journals unlocked scheme, see http://ard.bmj.com/info/unlocked.dtl

ABSTRACT

Objectives The genetic aetiology of osteoarthritis has not yet been elucidated. To enable a well-powered genome-wide association study (GWAS) for osteoarthritis, the authors have formed the arcOGEN Consortium, a UK-wide collaborative effort aiming to scan genome-wide over 7500 osteoarthritis cases in a two-stage genome-wide association scan. Here the authors report the findings of the stage 1 interim analysis.

Methods The authors have performed a genome-wide association scan for knee and hip osteoarthritis in 3177 cases and 4894 population-based controls from the UK. Replication of promising signals was carried out in silico in five further scans (44 449 individuals), and de novo in 14 534 independent samples, all of European descent.

Results None of the association signals the authors identified reach genome-wide levels of statistical significance, therefore stressing the need for corroboration in sample sets of a larger size. Application of analytical approaches to examine the allelic architecture of disease to the stage 1 genome-wide association scan data suggests that osteoarthritis is a highly polygenic disease with multiple risk variants conferring small effects.

Conclusions Identifying loci conferring susceptibility to osteoarthritis will require large-scale sample sizes and well-defined phenotypes to minimise heterogeneity.

Osteoarthritis is the most common form of arthritis affecting 40% of people over the age of 70 years, and is associated with a substantial health economic burden. Osteoarthritis is thought to be caused by a complex interplay between environmental and genetic factors.[1] As with many common complex disorders, the genetic architecture of osteoarthritis has not yet been characterised. Over the past decade, candidate gene association studies and genome-wide linkage scans have failed to identify robustly replicating osteoarthritis loci, with the notable exception of rs143383 in the *GDF5* gene.[2] A genome-wide association study (GWAS) recently reported a single novel locus associated with radiographically defined knee and/or hand osteoarthritis on chromosome 7q22 (rs3815148 in the *COG5* gene),[3] subsequently corroborated by a genome-wide association scan across four studies including the discovery set.[4] Both established that osteoarthritis loci are represented by common variants (>0.2 minor-allele frequency), have small effect sizes (allelic OR ~1.15), and reach genome-wide significance ($p<5\times10^{-8}$).[4 5] These associations have characteristics typical of common complex traits and require large sample sizes for their detection. To enable a well-powered GWAS for osteoarthritis, we formed the arcOGEN Consortium, a UK-wide collaborative effort aiming to scan genome-wide over 7500 osteoarthritis cases in a two-stage genome-wide association scan. Here we report the findings of our GWAS stage 1 analysis and replication studies, and describe the outcomes of statistical analyses designed to model the genetic architecture of osteoarthritis.

METHODS

An expanded description of the methods is provided in supplementary methods (available online only). The stage 1 genome-wide association scan included 3177 knee and/or hip osteoarthritis cases from the UK ascertained based on radiographic evidence of disease (Kellgren–Lawrence grade ≥2)[6] or clinical evidence of disease to a level requiring joint replacement. Cases were genotyped using the Illumina Human610 platform (Illumina, San Diego, California, USA). We used 4894 early-access publicly

available population-based UK controls from the Wellcome Trust Case Control Consortium 2 (WTCCC2) study, genotyped on the Illumina 1.2M Duo platform (Illumina). We compared allele frequencies across 514 898 autosomal single-nucleotide polymorphisms (SNP) passing quality control criteria. We also carried out joint-specific stratified analyses (for hip and knee osteoarthritis). We carried out sensitivity analyses by comparing genome-wide case genotypes against different control sets (one overlapping but typed on a different platform and one non-overlapping set of osteoarthritis-free 'supercontrols'). We took forward 102 independent (r^2<0.4) SNP with p<0.0001 to in-silico replication in three further osteoarthritis genome-wide association scan (from the deCODE, Framingham and Rotterdam studies) and a subset of 52 SNP in a UK-based genome-wide association scan (TwinsUK; supplementary table 1, available online only) (across 4124 cases and 37 581 controls in total). Based on meta-analysis results across arcOGEN and these in-silico replication datasets (supplementary results, available online only), we prioritised 36 SNP for de-novo replication in a further set of 6188 osteoarthritis cases and 8280 controls of European descent and for in-silico replication in 213 cases and 2531 controls from Estonia (supplementary table 1, available online only). To enhance our understanding of the genetic architecture of osteoarthritis, we applied analytical approaches[7] to stage 1 arcOGEN data to test the theory of polygenic inheritance. We used the arcOGEN stage 1 GWAS to derive a set of independent (r^2<0.05) associated SNP (~62 000) from a subset of the data and then used this score allele set to evaluate the proportion of case–control status accounted for in the remaining samples.

RESULTS

In our stage 1 genome-wide association scan analysis, we observed a slight excess of associations compared with the null distribution (supplementary figure 2, available online only) and similar patterns of association for the joint and gender-stratified analyses (supplementary results; supplementary figure 3, available online only). The genomic control inflation factor λ was 1.077, in keeping with other UK-based GWAS.[8] Although there was no signal exceeding genome-wide significance ($p<5\times10^{-8}$), 89 SNP reached p values of less than 10^{-4} (as opposed to 51 expected under the null, binomial $p=10^{-6}$).

The strongest statistical evidence for association with osteoarthritis was obtained for rs4512391 on chr8 (OR for allele C 1.17; 95% CI 1.10 to 1.25; $p=1.8\times10^{-6}$) approximately 66 kb upstream of the *TRIB1* gene. For knee osteoarthritis the most significant finding was also observed at rs4512391 (OR for allele C 1.23, 95% CI 1.13 to 1.33; $p=1.1\times10^{-6}$); for hip osteoarthritis, the strongest signal was rs4977469 (OR for allele A 1.30, 95% CI 1.17 to 1.45; $p=1.2\times10^{-6}$), within intron 3 of the predicted *FAM154A* gene (supplementary figure 1; supplementary figure 4; supplementary table 2, all available online only). Following replication studies in up to 58 917 independent European-ancestry samples, the overall statistical evidence for association was reduced, with the strongest signal for knee and/or hip osteoarthritis observed at rs2277831 on chr22 (OR for allele G 1.07, 95% CI 1.1.04 to 1.11; combined $p=2.3\times10^{-5}$), within intron 32 of the *MICAL3* gene. For knee osteoarthritis the most significant finding post-replication was observed at rs11280 (OR for allele C 1.10, 95% CI 1.05 to 1.16; $p=3.2\times10^{-5}$), within *C6orf130*; for hip osteoarthritis, the strongest signal was at rs2615977 (OR for allele A 1.10, 95% CI 1.05 to 1.15; $p=1.1\times10^{-5}$) within intron 31 of the *COL11A1* gene (supplementary figure 1; supplementary figure 4; supplementary table 2, all available online only). Out of the 65 independent signals (composed of 89 SNP) with $p<10^{-4}$ in all the osteoarthritis analysis, 17 are present with $p<10^{-4}$ in the hip osteoarthritis analysis and nine are present with $p<10^{-4}$ in the knee osteoarthritis analysis, but none reach $p<10^{-4}$ in both.

Additional analyses, including testing for pairwise interactions, examining overlap between genome-wide association scan and linkage scan signals, association with body mass index, deviation from the additive model, genome-wide HapMap-based imputation, chromosome X association and genome-wide low frequency/rare variant analysis did not identify any additional osteoarthritis signals (supplementary methods; supplementary results, available online only).

The arcOGEN stage 1 GWAS is well powered to detect association with common variants of modest effect at the genome-wide significance level (eg, 90% power to detect an allelic OR of 1.25 at a SNP with frequency 0.35). However, it is poorly powered to detect effects at the established osteoarthritis variants (8% and 4% power for *GDF5* and 7q22, respectively), as the index SNP (rs143383, risk allele frequency 0.67; rs3815148, risk allele frequency 0.23) both have small effect sizes (OR ~1.15).[2 3] Retaining the arcOGEN case–control ratio of 1.54, 7774 cases would be required to achieve 80% power to detect an allelic OR of 1.15 at a SNP with risk allele frequency 0.67, and 9084 cases would be required to achieve the same power to detect the same effect size at a SNP with risk allele frequency 0.23.

To evaluate the robustness of association signals, we performed sensitivity analyses using different control sets. We first used 'supercontrols' (hip and knee osteoarthritis-free individuals) and found high correlation between association effect estimates (r=0.88) and 93% concordance in the direction of effects (supplementary results, available online only). The two established osteoarthritis loci demonstrated stronger evidence for association compared with the main analysis, even though the 'supercontrol' sample size was smaller. We also used a subset of the population-based controls, genotyped on a different platform, to assess robustness to the typing method. We observed 100% concordance in the direction of effect and high correlation between estimates of the OR (r=0.94) (supplementary results, available online only). Therefore, the choice of controls may have affected association strength but not the direction of effect.

The genetic architecture of osteoarthritis is likely to be polygenic with multiple variants along the spectrum of allele frequencies contributing modest and small effects. Our polygene analyses support a model of osteoarthritis in which there is a substantial genetic component comprising multiple contributing variants with small effect sizes (figure 1). SNP with p values as high as 0.25 appear to contribute to the genetic component of osteoarthritis (empirical $p=3\times10^{-5}$, based on ~85 million permutations), with the bulk of the contribution seen in the 0–0.1 p value range (empirical $p=3.5\times10^{-5}$, ~101 million permutations). Evaluation of discreet p value bins corroborates this observation and supports a role for SNP with p values up to 0.25 (0.10>p>0.15 bin empirical p=0.06; 0.15>p>0.20 bin p=0.0072; 0.20>p>0.25 bin p=0.022), but with the major contribution coming from SNP with p<0.10 (0>p>0.05 bin $p=3\times10^{-5}$; 0.05>p>0.10 bin p=0.0092). The estimated proportion of variance in disease state explained by the osteoarthritis score alleles is 3.05% ($p=3.3\times10^{-4}$ based on nine million permutations). This is in keeping with findings in other complex common diseases.[7]

DISCUSSION

Our stage 1 arcOGEN genome-wide association scan analysis results are in agreement with other large-scale genetic studies,[3 4]

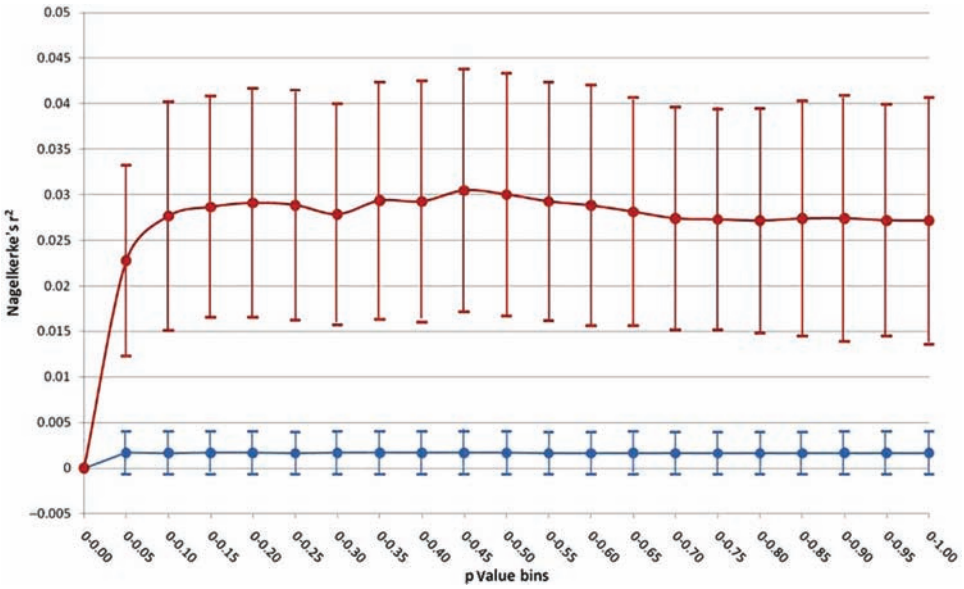

Figure 1 Summary of polygene analysis results evaluating the genetic architecture of osteoarthritis. Lines in red (real data) and blue (permuted data) show mean±1 SD of the Nagelkerke's pseudo r^2 statistic, which indicates the proportion of case–control status accounted for by score alleles in the single-nucleotide polymorphism (SNP) sets. The blue line represents expectation under the null hypothesis of no genetic component to the disease. The red line represents stage 1 genome-wide association study data and shows that SNP at the tail of the p value distribution (up to $p<0.25$, but primarily $p<0.10$) have a significantly higher case–control discriminatory capacity.

and clearly indicate that common SNP with large effect sizes are not likely to underpin the aetiology of osteoarthritis. The only two established osteoarthritis loci to date have small OR and our polygene analyses on common SNP suggest that the genetic architecture of osteoarthritis is likely to consist of numerous signals of similar magnitude. The significant increase in sample size with stage 2 of the arcOGEN genome-wide association scan (~2.4 times as many cases) will increase the power to detect osteoarthritis associations. In addition, large-scale international meta-analysis efforts are underway and will ensure maximal GWAS sample size. A further important parameter in enhancing power and increasing the chances of success involves the improved definition of phenotype. One of the reasons why replication of findings has been difficult to achieve in osteoarthritis could be the inherent heterogeneity of the different osteoarthritis diagnostic and study inclusion criteria used. The field is currently active in evaluating phenotype definition differences and their effects on study power.[4]

The Wellcome Trust Case Control Consortium first demonstrated the utility of population-based, rather than disease-free, controls in GWAS.[8] Our 'supercontrol'-based sensitivity analyses suggest that for a highly prevalent and heterogeneous disorder such as osteoarthritis, with multiple smaller effects contributing to overall susceptibility, the brute force approach of maximising sample size to balance misclassification in controls is not as successful as it has been for other common diseases, in which 'low-hanging fruit' discoveries (of loci with substantial effects) were robust to subtle allele frequency fluctuations.

Osteoarthritis is a heterogeneous disease characterised by variable clinical features with conceivably different genetic aetiologies.[9] Although the allele score was associated with osteoarthritis, the proportion of disease variance explained cannot be highly accurately quantified, primarily because of signal attenuation (risk alleles are unlikely to be at the actual causal locus) and sampling variation. The genetic architecture of osteoarthritis is emerging as complex. Large-scale sample sizes and well-defined phenotypes will be required to gain a better understanding of this genetic component, possibly leading to optimising treatment, developing efficacious disease-modifying interventions, improving prognosis and tailoring intervention to the individual.

Funding This study was financially supported by Arthritis Research UK.

Competing interests None.

Ethics approval This study was conducted with the approval of the Oxfordshire Research Ethics Committee C reference 07/H0606/150.

Provenance and peer review Not commissioned; externally peer reviewed

Author affiliations [1]Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK
[2]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
[4]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK
[5]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
[6]deCODE Genetics, Reykjavik, Iceland
[7]NIHR Biomedical Research Unit, University of Oxford, Oxford, UK
[8]MRC Epidemiology Resource Centre, University of Southampton, Southampton, UK
[9]Academic Rheumatology, University of Nottingham, Nottingham, UK
[10]Institute of Cellular Medicine, Musculoskeletal Research Group, Newcastle University, Newcastle upon Tyne, UK
[11]The Newcastle upon Tyne Hospitals NHS Trust Foundation Trust, The Freeman Hospital, Byker, Newcastle upon Tyne, UK
[12]Centre for Integrated Genomic Medical Research, University of Manchester, Manchester, UK
[13]Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[14]Wellcome Trust Centre for Cell Matrix Research, University of Manchester, Manchester, UK
[15]Academic Unit of Bone Metabolism, Department of Human Metabolism, University of Sheffield, Sheffield, UK
[16]Sheffield NIHR Bone Biomedical Research Unit, Centre for Biomedical Research, Northern General Hospital, Sheffield, UK
[17]Wellcome Trust Clinical Research Facility, Southampton General Hospital, Southampton, UK
[18]INIBIC-Hospital Universitario A Coruña, Osteoarticular and Aging Research Laboratory, A Coruña, Spain
[19]Boston University School of Public Health, Boston, Massachusetts, USA
[20]Estonian Genome Center, University of Tartu, Tartu, Estonia
[21]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia
[22]Estonian Biocenter, Tartu, Estonia

[23]Department of Hygiene and Epidemiology, University of Ioannina, Ioannina, Greece
[24]Clinical Epidemiology Unit, Boston University School of Medicine, Boston, Massachusetts, USA
[25]Rheumatology Unit, Hospital Clinico Universitario de Santiago, Santiago de Compostela, Spain
[26]Department of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain
[27]Laboratorio de Investigacion 10, Hospital Clinico Universitario de Santiago, Santiago de Compostela, Spain
[28]Reykjavic University, Reykjavik, Iceland
[29]Erasmus Medical Centre, Rotterdam, The Netherlands
[30]The Netherlands Genomics Initiative-Sponsored Netherlands Consortium for Healthy Aging, Rotterdam and Leiden, The Netherlands
[31]Tufts Clinical and Translational Science Institute and Tufts University School of Medicine, Boston, Massachusetts, USA
[32]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA
[33]Stanford Prevention Research Center, Stanford University, Stanford, California, USA
[34]Department of Health Sciences, University of Akureyri, Norðurslóð, Akureyri, Iceland
[35]Department of Medicine, University of Iceland, Reykjavik, Iceland
[36]Department of Medicine, Landspitali University Hospital, Reykjavik, Iceland
[37]Royal National Orthopaedic Hospital, Brockley Hill, Stanmore, UK
[38]Departments of Rheumatology and Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands
[39]Section of Molecular Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands
[40]Aging Center, Medicine and Rheumatology, University of California at Davis Medical Center, Sacramento, California, USA
[41]The Robert S Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, Manhasset, New York, USA
[42]Department of Epidemiology and Biostatistics, San Francisco Coordinating Center, University of California, San Francisco, California, USA
[43]California Pacific Medical Center, San Francisco, California, USA
[44]Department of Internal Medicine, Hospital U M Valdecilla, University of Cantabria, Santander, Spain
[45]Department of Orthopedic Surgery and Traumatology, Hospital U M Valdecilla, Santander, Spain
[46]University of Thessaly, Larissa, Greece
[47]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

REFERENCES

1. **Valdes AM,** Spector TD. The contribution of genes to osteoarthritis. *Med Clin North Am* 2009;**93**:45–66; x.
2. **Evangelou E,** Chapman K, Meulenbelt I, *et al.* Large-scale analysis of association between GDF5 and FRZB variants and osteoarthritis of the hip, knee, and hand. *Arthritis Rheum* 2009;**60**:1710–21.
3. **Kerkhof HJ,** Lories RJ, Meulenbelt I, *et al.* A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. *Arthritis Rheum* 2010;**62**:499–510.
4. **Evangelou E,** Valdes AM, Kerkhof HJ, *et al.* Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis.* Published Online First: 10 November 2010. doi: 10.1136/ard.2010.132787.
5. **Valdes AM,** Evangelou E, Kerkhof HJ, *et al.* The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. *Ann Rheum Dis* 2010; (in Press).
6. **Kellgren JH,** Lawrence JS, Bier F. Genetic factors in generalized osteo-arthrosis. *Ann Rheum Dis* 1963;**22**:237–55.
7. **Purcell SM,** Wray NR, Stone JL, *et al.* International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;**460**:748–52.
8. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661–78.
9. **Valdes AM,** McWilliams D, Arden NK, *et al.* Involvement of different risk factors in clinically severe large joint osteoarthritis according to the presence of hand interphalangeal nodes. *Arthritis Rheum* 2010;**62**:2688–95.