CONCISE REPORT

No evidence of an association between mitochondrial DNA variants and osteoarthritis in 7393 cases and 5122 controls

Gavin Hudson,[1] Kalliope Panoutsopoulou,[2] Ian Wilson,[1] Lorraine Southam,[2,3] Nigel W Rayner,[3,4] Nigel Arden,[5,6] Fraser Birrell,[7,8] Ian Carluke,[8] Andrew Carr,[6] Kay Chapman,[6] Panos Deloukas,[2] Michael Doherty,[9] Andrew McCaskie,[7,10] William E R Ollier,[11] Stuart H Ralston,[12] Mike R Reed,[8] Tim D Spector,[13] Ana M Valdes,[13] Gillian A Wallis,[14] J Mark Wilkinson,[15,16] Eleftheria Zeggini,[2] David C Samuels,[17] John Loughlin,[7] Patrick F Chinnery,[1] arcOGEN Consortium

▶ Additional data are published online only. To view this file please visit the journal online (http://dx.doi.org/10.1136/annrheumdis-2012-201932).

For numbered affiliations see end of article.

Correspondence to
Professor John Loughlin, Institute of Cellular Medicine, Newcastle University, The Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, UK; John.loughlin@ncl.ac.uk and Professor Patrick F Chinnery, Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK; P.F.Chinnery@ncl.ac.uk

Received 25 April 2012
Accepted 19 July 2012
Published Online First
14 September 2012

ABSTRACT

Objectives Osteoarthritis (OA) has a complex aetiology with a strong genetic component. Genome-wide association studies implicate several nuclear genes in the aetiology, but a major component of the heritability has yet to be defined at the molecular level. Initial studies implicate maternally inherited variants of mitochondrial DNA (mtDNA) in subgroups of patients with OA based on gender and specific joint involvement, but these findings have not been replicated.

Methods The authors studied 138 maternally inherited mtDNA variants genotyped in a two cohort genetic association study across a total of 7393 OA cases from the arcOGEN consortium and 5122 controls genotyped in the Wellcome Trust Case Control consortium 2 study.

Results Following data quality control we examined 48 mtDNA variants that were common in cohort 1 and cohort 2, and found no association with OA. None of the phenotypic subgroups previously associated with mtDNA haplogroups were associated in this study.

Conclusions We were not able to replicate previously published findings in the largest mtDNA association study to date. The evidence linking OA to mtDNA is not compelling at present.

INTRODUCTION

Osteoarthritis (OA) is the most common age-related joint disease. The pathogenesis of OA is complex, with several environmental and genetic factors implicated in the aetiology.[1] Genome-wide association studies implicate several nuclear genes in the risk of developing OA, but a major component of the heritability remains unexplained.[2]

Mitochondria are intracellular organelles involved in the synthesis of ATP, the principal source of energy within cells. Several lines of evidence implicate mitochondrial dysfunction in the pathogenesis of OA, including the disruption of respiratory chain activity in chondrocytes,[3] the increased production of reactive oxygen species which compromise chondrocyte function[4 5] and a central role in one apoptotic pathway.[6]

Mitochondria contain multiple copies of the 16.5 kb mitochondrial genome (mitochondrial DNA (mtDNA)). MtDNA codes for 13 essential respiratory chain proteins and the RNAs required for intramitochondrial protein synthesis. Point mutations of mtDNA compromise oxidative phosphorylation and are a major cause of human disease.[7] This raises the possibility that more subtle polymorphic variants contribute to genetic susceptibility of common complex traits, including OA. Being strictly maternally inherited, mtDNA undergoes negligible intermolecular recombination. Specific sequence variants that occurred 10–40 000 years ago define clades of the European mtDNA phylogeny called mtDNA haplogroups, each containing both distinct and shared genetic variants.[8] Two studies have reported an association between different mtDNA haplogroups and OA, with one describing a reduced risk of knee OA with haplogroups J (defined by m.4216T>C and m.10398A>G) and JT (m.4216T>C), and another describing a reduced risk of hip OA with haplogroups J (defined by m.4216T>C and m.10398A>G) and J1c (m.14798T>C).[9 10] However, each study lacked the power and resolution to reliably detect an association with all but the most common haplogroups. Moreover, despite being from the same geographic location, the studies reported different haplogroup associations with different joint involvement in OA for specific genders.[9 10] While these data support the potential role of mtDNA in determining OA risk, the absence of direct replication means that the role of mtDNA in OA is thus uncertain, and the joint-specific risk is difficult to explain based on the proposed mechanism. To address this issue, we carried out a two phase study of 138 mtDNA variants in 7393 OA cases and 5122 controls as part of the arcOGEN consortium.

METHODS

We performed a two stage genetic association study. The cases were part of the arcOGEN study, and the controls were part of the Wellcome Trust Case Control Consortium 2 (WTCCC2) study,

both ascertained as described previously.[2 11] In cohort 1, 3093 cases (ARC1) were compared with 2640 controls from the 1958 Birth Cohort (WTCCC-58C). In cohort 2, 4300 cases (ARC2) were compared with 2482 UK Blood Service controls (WTCCC-NBS). All cases were genotyped using the Illumina Human610 platform (Illumina, San Diego, California, USA). All controls were genotyped on the Illumina 1.2M Duo platform (Illumina). Individuals were excluded from subsequent analysis if data were absent in >10% of single nucleotide polymorphism (SNP) (cohort 1 cases=0 and WTCCC-58C=44 and cohort 2 cases=0 and WTCCC-NBS=11) and SNPs were excluded if >10% of genotypes were absent (cohort 1=5 and cohort 2=4).[11] We excluded 30 SNPs (cohort 1=15 and cohort 2=21) with a study-wide missing data rate >5% or >1% for SNPs with a study-wide minor allele frequency (MAF) <5%.[11] Finally, 64 SNPs with MAF<1% were removed (cohort 1=53 and cohort 2=64). Subsequent analysis was restricted to a concordant dataset of 62 mtDNA variants passing quality control (QC) in both phases. Differential missingness tests between cases and controls revealed significant differences in 14 SNPs ($p=<10^{-4}$) reducing the final number of experimental SNPs to 48. Statistical significance was defined as $p<0.05$ in both phases. χ^2 And missingness were computed using *PLINK v 2.050* (http://pngu.mgh.harvard.edu/purcell/plink/).[12]

Array genotypes were used to identify mtDNA haplogroup-specific sequence motifs, when compared with the mtDNA reference sequence (see online supplementary materials and table S1), allowing 98.48% of subjects to be successfully assigned to a European haplogroup.

A principal components analysis (PCA)[13] was performed on the X : Y ratios of raw intensities for all SNPs for the combined dataset of cases (n=7393) and controls (n=5122) using R. The k-means function of the R cluster package was used to cluster individuals into 5, 10, 15 and 20 groups using these PCA scores.

RESULTS

Using established QC criteria[14] we rejected 90 SNPs, leaving 48 in both cohorts. There was no association between any mtDNA variant passing QC and OA. Stratifying by gender, joint involvement (knee or hip) and method of case ascertainment (radiograph or arthroplasty) failed to reveal any significant associations (see online supplementary table S2).

We then studied mtDNA haplogroups using two independent approaches: (1) manual haplogroup calling based on phylogenetics (for methods see online supplementary table 1) and (2) hypothesis-free approach based on raw fluorescence intensities in a principal components analysis. Analysis of mtDNA haplogroups between cases and controls failed to replicate previous studies (table 1). Stratifying by gender, joint involvement (knee or hip) and method of case ascertainment (radiograph or arthroplasty) failed to reveal any significant associations with mtDNA haplogroups (see online supplementary table S3). PCA analysis identified the main mtDNA haplogroups previously identified through phylogenetic analysis. These appear as clusters in figure 1, where each cluster represents a major European mtDNA haplogroup. However, there was no significant difference in cluster membership for any of the disease or control cohorts, or when the cases and controls were pooled (figure 1, online supplementary table S4). The data were only pooled after confirming that the allele frequencies were not significantly different between the two case groups or between the two control groups.

DISCUSSION

The previously reported haplogroup associations were with specific subgroups of OA patients, subcategorised based on gender and/or specific joint involvement.[9 10] Our study had >99% power to detect these associations, and there was no significant difference in the frequency of the previously associated SNPs between controls in the published studies[9 10] and the controls used in our study. The previously published results were based on substantially smaller study groups than the ones we describe here, raising the possibility that the published findings are a false positive finding. On the other hand, the relative contribution of specific mtDNA variants could vary in different ethnic groups, possibly through an interaction with environmental factors and different nuclear genes.[15] In practice, this means that the specific mtDNA variants which fail to show an association with disease in this study could be associated with disease in a different ethnic population. Geographic variation in allelic association could also arise through homoplasy. Homoplasy is the recurrence of mutations on different branches of the mtDNA phylogeny in different parts of the world. Homoplasy accounts for up to 20% of mtDNA variation, and often involves non-synonymous substitutions.[16] This raises the possibility that haplogroup markers tag different homoplastic functional variants in different populations. If the homoplasies are having a functional effect, then this would lead to different haplogroup associations in different studies across the globe. Finally, it is possible that geographic differences in the fine detail of the sub-haplogroup structure of mtDNA could account for inconsistencies between studies, as

Table 1 Frequency of mitochondrial DNA (mtDNA) haplogroups in osteoarthritis and control subjects

Hap	F ACR1	F WTCCC-58C	P	F ARC2	F WTCCC-NBS	P	F cases	F control	p Value
H	1375 (44.5)	1160 (43.9)	5.97E−01	1887 (43.9)	1080 (43.5)	6.36E−01	3262 (44.1)	2240 (43.7)	6.67E−01
V	107 (3.5)	84 (3.2)	5.40E−01	133 (3.1)	86 (3.5)	4.25E−01	240 (3.2)	170 (3.3)	8.42E−01
J	335 (10.8)	311 (11.8)	2.81E−01	527 (12.3)	292 (11.8)	5.02E−01	862 (11.7)	603 (11.8)	8.25E−01
T	309 (10.0)	266 (10.1)	9.56E−01	449 (10.4)	227 (9.1)	7.50E−02	758 (10.3)	493 (9.6)	5.39E−01
U	440 (14.2)	334 (12.7)	7.20E−02	549 (12.8)	366 (14.7)	2.70E−02	989 (13.4)	700 (13.7)	7.35E−01
K	280 (9.1)	241 (9.1)	9.59E−01	362 (8.4)	216 (8.7)	7.32E−01	642 (8.7)	457 (8.9)	8.20E−01
W	56 (1.8)	46 (1.7)	8.30E−01	91 (2.1)	56 (2.3)	7.26E−01	147 (2.0)	102 (2.0)	3.38E−01
X	50 (1.6)	46 (1.7)	7.26E−01	76 (1.8)	35 (1.4)	2.53E-01	126 (1.7)	81 (1.6)	8.27E−01
I	88 (2.8)	99 (3.8)	5.80E−02	131 (3.0)	80 (3.2)	7.13E−01	219 (3.0)	179 (3.5)	1.19E−01
0	53 (1.7)	53 (2.0)	4.22E−01	95 (2.2)	44 (1.8)	2.11E−01	148 (2.0)	97 (1.9)	6.31E−01
Missing	6	59		11	19		17	78	

Haplogroups determined manually, as described in the Methods section and online supplementary table S1.
Uncorrected χ^2 comparison of haplogroup frequencies showing: (i) cohort 1 (*ARC1,* n=3093 and controls *WTCCC-58C*, n=2640); (ii) cohort 2 (*ARC2,* n=4300 and controls *WTCCC-NBS*, n=2482); and (iii) pooled arcOGEN cases (n=7393) and controls (n=5122). Hap=mtDNA haplogroup, *F* = frequency (percentage in brackets) and p=uncorrected Pearson's χ^2 probability. For study and control group acronyms see the Methods section.

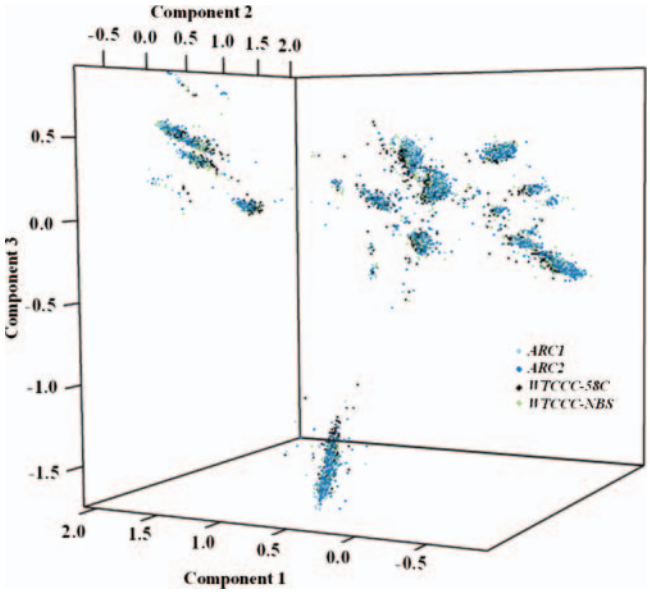

Figure 1 Principal components analysis (PCA) performed on raw fluorescent intensities in the pooled data set of 7393 osteoarthritis (OA) cases and 5122 controls. The PCA provided independent confirmation of the European haplogroup structure. There was no significant difference in the cluster distribution between OA cases and controls (online supplementary table S4). 58C, WTCCC MRC 1958 birth cohort control genotypes; ARC1, arcOGEN cohort 1; ARC2, arcOGEN cohort 2; NBS, WTCCC national blood transfusion service control genotypes; WTCCC, Wellcome Trust Case Control Consortium. This figure is only reproduced in colour in the online version.

described for Leber's hereditary optic neuropathy, where sub-branches of haplogroup J are associated with either an increased or decreased risk of visual failure in different populations, largely due to specific differences in the cytochrome B protein sequence.[17] However, in the largest mtDNA association study to date, we found no association between OA and the major European mtDNA haplogroups in either cohort. If an association does exist between OA and mtDNA, it is likely that this will only be resolved through extensive genotyping, ideally at the whole mtDNA level, in a much larger cohort of cases and controls and should be consistent in more than one geographic region.

Author affiliations
[1]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK
[2]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
[3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
[4]The Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK
[5]The MRC Epidemiology Resource Centre, University of Southampton, Southampton, UK
[6]The Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[7]Institute of Cellular Medicine, The Medical School, Framlington Place, Newcastle upon Tyne, UK
[8]Northumbria Healthcare NHS Foundation Trust, Ashington, UK
[9]Academic Rheumatology, University of Nottingham, Nottingham, UK
[10]The Freeman Hospital, Newcastle upon Tyne, UK
[11]The Centre for Integrated Genomic Medical Research, The University of Manchester, Manchester, UK
[12]The Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[13]TwinsUK Unit, King's College London, London, UK
[14]The Wellcome Trust Centre for Cell-Matrix Research, University of Manchester, Manchester, UK
[15]The Academic Unit of Bone Metabolism, Department of Human Metabolism, University of Sheffield, Sheffield, UK
[16]The Sheffield NIHR Musculoskeletal Biomedical Research Unit, Northern General Hospital, Sheffield, UK
[17]Center for Human Genetics Research, Vanderbilt University Medical Centre, Nashville, Tennessee USA

Acknowledgements The UK samples used in arcOGEN derive from eight centres: Nottingham, London, Oxford, Sheffield, Southampton, Newcastle, Wansbeck and Edinburgh. For Nottingham we acknowledge Arthritis Research UK for funding the collection of the majority of extant cases (cohort 1) and we thank Sally Doherty and Maggie Wheeler for organising cohort 2 patient ascertainment. For London we thank the staff from the TwinsUK unit and the Chingford Study for cohort 1 patient ascertainment, we thank Fiona O'Neill for organising cohort 2 patient ascertainment, and we acknowledge financial support from Arthritis Research UK, from the Wellcome Trust, from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre (BRC) award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London, and from the European Commission framework 7 programme TREAT-OA (grant 200800). For Oxford we acknowledge funding support from the Collisson Foundation, the Botnar Foundation and the Jean Shanks Foundation for cohort 1 patient ascertainment, we acknowledge the NIHR for supporting the Biomedical Research Unit (BRU) at the University of Oxford, and we thank Bridget Watkins, Anne Hanley and Kim Clipsham for cohorts 1 and 2 patient ascertainment. For Sheffield we acknowledge the NIHR for supporting the Sheffield Musculoskeletal BRU collection site, the South Yorkshire Clinical Research Network for part-funding the Sheffield research nurses and for clerical support, the Royal College of Surgeons of England and the Cavendish Hip Foundation, Andrew Gordon for cohort 1 patient ascertainment, and Diane Swift, Ellen Paling and Adam Taylor for cohort 2 patient ascertainment. For Southampton we acknowledge the Wellcome Trust Clinical Research Facility at Southampton General Hospital, we thank Phillippa-Kate Battley and Elizabeth Arden for organising cohort 2 patient ascertainment, and we thank Richard Keen and Anna Bara for facilitating access to the Arthritis Research UK-funded VIDEO study. For Newcastle we acknowledge Sheryl Mitchell and the Freeman Hospital arthroplasty surgeons Nigel Brewster, David Deehan, Craig Gerrand, Andy Gray, Munawar Hashmi, Jim Holland and David Weir for assistance in cohort 2 patient ascertainment, we thank Karen Bettinson for her assistance, we acknowledge the support of the UK NIHR BRC for Ageing and Age-related disease award to the Newcastle upon Tyne Hospitals NHS Foundation Trust and the support of the Northumberland, Tyne and Wear CLRN. For Wansbeck we thank Kirsten Walker and Donna Gray for organising cohort 2 patient ascertainment, and we acknowledge the Northumbria NHS Foundation Trust arthroplasty consultants Mr Partington, Mr Muller, Mr Emmerson, Mr Prasad, Mr Kramer, Mr Jones, Mr Murty, Mr Asaad, Mr Leitch, Mr Pratt, Mr Townshend and Miss Van Kampen for accrual. For Edinburgh we acknowledge the contribution of Nichola Koller, Gayle Scott and Helen Watters for recruitment and ascertainment of cohort 2 cases. We acknowledge cohorts 1 and 2 sample management undertaken by the UK DNA Banking Network funded by the Medical Research Council at CIGMR—the Centre for Integrated Genomic Medical Research, University of Manchester, with special thanks to Kate Sherburn and Kate Dixon. Genotyping was performed at the Wellcome Trust Sanger Institute and we thank Emma Gray, Sarah Edkins and Cordelia Langford for their assistance. We thank Vanessa Cox of the MRC Epidemiology Resource Centre, Southampton, for database and computing services for the arcOGEN cohort 2 patient questionnaires.

Contributions GH and IW carried out the primary analysis, supervised by PFC, DCS and JL. KP, LS, NWR, NA, FB, IC, AC, KC, PD, MD, AMcC, WERO, SHR, MRR, TDS, AMV, GAW, JMW, EZ and JL generated the primary data. PFC, JL and GH wrote the manuscript, which was modified following comments from the other authors . The authors declare no conflict of interest.

Funding arcOGEN (http://www.arcogen.org.uk/) was funded by a special purpose grant from Arthritis Research UK (grant 18030). This study used genotype data from population controls that were generated by the Wellcome Trust Case Control Consortium 2 (a full list of the investigators who contributed to the generation of the data is available from http://www.wtccc.org.uk, funded by The Wellcome Trust (grant 083948). The population controls were from the 1958 British Birth Cohort collection funded by the Medical Research Council (grant G0000934) and The Wellcome Trust (grant 068545) and from the UK Blood Services Collection of Common Controls funded by The Wellcome Trust. KP, LS and EZ are supported by the Wellcome Trust (098051), KC is supported by a Botnar Fellowship, NWR is supported by the Wellcome Trust (WT079557MA) and JMW is supported by the Higher Education Funding Council for England. The funding sources were not involved in the study design; in the collection, analysis, and interpretation of data, in the writing of the report, or in the decision to submit the paper for publication. Funding for the project was provided by the Wellcome Trust under award 076113 and 085475. PFC is a Wellcome Trust Senior Fellow in Clinical Science (WT084980/Z/08/Z) and an NIHR Senior Investigator, who is also supported through the Wellcome Trust centre for Mitochondrial Research (WT096919Z/11/Z), the Medical Research Council (UK), the UK Parkinson's Disease Society, and the UK NIHR Biomedical Research Centre for Ageing and Age-related disease award to the Newcastle upon Tyne Foundation Hospitals NHS Trust.

Competing interest None.

Ethics approval Ethics approval was provided by Newcastle REC.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Valdes AM,** Spector TD. The contribution of genes to osteoarthritis. *Rheum Dis Clin North Am* 2008;**34**:581–603.
2. **Panoutsopoulou K,** Southam L, Elliott KS, *et al*. Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Ann Rheum Dis* 2011;**70**:864–7.
3. **Blanco FJ,** Lopez-Armada MJ, Maneiro E. Mitochondrial dysfunction in osteoarthritis. *Mitochondrion* 2004;**4**:715–28.
4. **Maneiro E,** Martin MA, de Andres MC, *et al*. Mitochondrial respiratory activity is altered in osteoarthritic human articular chondrocytes. *Arthritis Rheum* 2003;**48**:700–8.
5. **Fernandez-Moreno M,** Soto-Hermida A, Pertega S, *et al*. Mitochondrial DNA (mtDNA) haplogroups and serum levels of anti-oxidant enzymes in patients with osteoarthritis. *BMC Musculoskelet Disord* 2011;**12**:264.
6. **Grishko V,** Xu M, Wilson G, Pearsall AWT. Apoptosis and mitochondrial dysfunction in human chondrocytes following exposure to lidocaine, bupivacaine, and ropivacaine. *J Bone Joint Surg Am* 2010;**92**:609–18.
7. **Taylor RW,** Turnbull DM. Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 2005;**6**:389–402.
8. **Torroni A,** Huoponen K, Francalacci P, *et al*. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 1996;**144**:1835–50.
9. **Rego I,** Fernandez-Moreno M, Fernandez-Lopez C, *et al*. Role of European mitochondrial DNA haplogroups in the prevalence of hip osteoarthritis in Galicia, Northern Spain. *Ann Rheum Dis* 2010;**69**:210–13.
10. **Rego-Perez I,** Fernandez-Moreno M, Fernandez-Lopez C, *et al*. Mitochondrial DNA haplogroups: role in the prevalence and severity of knee osteoarthritis. *Arthritis Rheum* 2008;**58**:2387–96.
11. **Saxena R,** Bakker PIW, Singer K, *et al*. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am J Hum Genet* 2006;**79**:54–71.
12. **Purcell S,** Neale B, Todd-Brown K, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75.
13. **Jolliffe IT.** *Principal component analysis*. Springer-Verlag, 1986.
14. **Wellcome-Trust-Case-Control-Consortium**. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007;**447**:661–78.
15. **Ruiz-Pesini E,** Mishmar D, Brandon M, *et al*. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 2004;**303**:223–6.
16. **Herrnstadt C,** Elson JL, Fahy E, *et al*. Reduced median network analysis of complete mtDNA coding region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 2002;**70**:1152–71.
17. **Carelli V,** Achilli A, Valentino ML, *et al*. Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber hereditary optic neuropathy pedigrees. *Am J Hum Genet* 2006;**78**:564–74.

Online Supplementary Material

Materials and Methods

Study subjects. Case ascertainment has been described previously[2]. In brief, cases that were subjected to the GWAS and used in this study were collected from eight locations in the arcOGEN consortium (Edinburgh, London, Nottingham, Newcastle, Oxford, Sheffield, Southampton and Wansbeck (Northumbria NHS Foundation Trust)). All cases were collected in the UK, were unrelated and of European origin. The phenotype was determined by two criteria; radiographic evidence of disease and clinical evidence of disease to a level requiring joint replacement. Each investigator defined radiographic disease as a Kellgren-Lawrence (KL)[18] grade of ≥2. We used 5122 publicly available population-based UK controls from the Wellcome Trust Case Control Consortium 2 (WTCC2) study: the 1958 Birth Cohort (WTCCC-58C) and the UK Blood Service control group (WTCCC-NBS) forming part of the Wellcome Trust Case Control Consortium study and genotyped in WTCCC2.

mtDNA Haplogroup determination. The hierarchical relationship amongst mtDNA variants is represented at www.phylotree.org. Each array variant (n=36, post QC), in each subject (n=12515, post QC), was compared to the rCRS and used to identify haplogroup specific motifs. Assignment of haplogroup was performed according to published criteria (Torroni 1996), for example haplogroup J is defined by variants m.11251 *and* [m.295+m.489+m.10398] (Supplementary Table 1). If a no-call was detected in a major-haplogroup defining SNP, then clade specific sub-type variants were used to identify mtDNA haplogroup. Based on available genotyping data 98.48% of subjects were assigned to a European haplogroup.

Table S1. European haplogroup definitions for Illumina array genotyping data. Table showing the major European haplogroups (H, V, J, T, U, K, W, X, I,) and corresponding sub-haplogroups. Shown are the non-rCRS (thus, non-haplogroup-H) branch SNPs, array specific 'super-haplogroup' SNPs (i.e. JT=11251), and array specific haplogroup defining SNPs (i.e. J=295+489+10398).

Major European haplogroups	***rCRS Branch SNPs***	***'Super-haplogroup' SNPs***	***Array haplogroup SNPs***
H	-	-	-
V	750, 1438, 2706, 4769	15904	15904
J	750, 1438, 2706, 4769, 14766	11251	295, 489, 10398
T	750, 1438, 2706, 4769, 14766	11251	4917, 10463, 15928
U	750, 1438, 2706, 4769, 14766	-	11467, 12308, 12372
K (aka – U8b)	750, 1438, 2706, 4769, 14766	11467,12308, 12372, 9698, 3480	10550, 14798
W	750, 1438, 2706, 4769, 14766, 12705	5046, 12414	5460
X	750, 1438, 2706, 4769, 14766, 12705	-	6221, 16278
I	750, 1438, 2706, 4769, 14766, 12705	1719, 10238, 13780, 10398, 15043, 250, 15294	10034, 16129, 16391

European sub-haplogroups	***rCRS Branch SNPs***	***'Super-haplogroup' SNPs***	***Array haplogroup SNPs***
I1	750, 1438, 2706, 4769, 14766, 12705	1719, 10238, 13780, 10398, 15043, 250, 15924, 10034, 16129, 16391	6734
I2	750, 1438, 2706, 4769, 14766, 12705	1719, 10238, 13780, 10398, 15043, 250, 15924, 10034, 16129, 16391	15758
I5	750, 1438, 2706, 4769, 14766, 12705	1719, 10238, 13780, 10398, 15043, 250, 15924, 10034, 16129, 16391	14233
X1	750, 1438, 2706, 4769, 14766, 12705	6221, 16278	16278
X2	750, 1438, 2706, 4769, 14766, 12705	6221, 16278	1719
J2	750, 1438, 2706, 4769, 14766	11251, 295, 489, 10398	15257
T1	750, 1438, 2706, 4769, 14766	11251, 4917, 10463, 15928	16163
T2	750, 1438, 2706, 4769, 14766	11251, 4917, 10463, 15928	14233
U5	750, 1438, 2706, 4769, 14766	11467, 12308, 12372	16270

U6	750, 1438, 2706, 4769, 14766	11467, 12308, 12372	3348
K1	750, 1438, 2706, 4769, 14766	11467, 12308, 12372, 9698, 3480, 10550, 14798	10398
K2	750, 1438, 2706, 4769, 14766	11467, 12308, 12372, 9698, 3480, 10550, 14798	9716
V1	750, 1438, 2706, 4769	15904	8869
V2	750, 1438, 2706, 4769	15904	13105
V6	750, 1438, 2706, 4769	15904	16162
H1	750, 4769, 1438	-	3010
H4	750, 4769, 1438	-	5004

Table S2. Comparison of mtDNA variant frequencies between cases and controls, stratified by gender, joint involvement, and method of ascertainment. Ascertainment groups: hip cases only = isolated hip OA on diagnosis; knee cases only = isolated knee OA on diagnosis; both cases only = both hip and knee OA on diagnosis; X-ray = diagnosis radiologically. Uncorrected chi-squared comparison of mtDNA variant frequencies in cohort 1 (*ARC1*, n = 3093 and *WTCCC-58C*, n = 2640) and cohort 2 (*ARC2*, n = 4300 and *WTCCC-NBS*, n = 2482), stratified by i) gender, ii) affected joint type (hip, knee and both) and iii) diagnosis ascertainment (arthro=arthroplasty and X-ray). The number of association tests reaching a nominal significance level are shown below the number of cases and controls contributing to each specific stratification. Uncorrected = uncorrected P-value; corrected (x48) = corrected for 48 different SNPs; and corrected (x336) = corrected for 99 different SNPs and each one of the seven phenotypic groups in two cohorts. n.a.= not available.

		Cohort 1							***Cohort 2***						
		Males v Males	*Females v Females*	*Hip cases only*	*Knee cases only*	*Both cases only*	*Arthro cases only*	*X-ray cases only*	*Males v Males*	*Females v Females*	*Hip cases only*	*Knee cases only*	*Both cases only*	*Arthro cases only*	*X-ray cases only*
	Cases n=	1136	1936	1503	1398	192	676	1270	1793	2507	1782	2071	447	3969	331
	Controls n=	1363	1277	2640	2640	2640	2640	2640	1228	1254	2482	2482	2482	2482	2482
Uncorrected	*P=<0.05*	1	7	3	3	1	3	2	2	4	4	6	4	8	1
Corrected (x48)	*P=<0.001*	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Corrected (x336)	*P=<0.0001*	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Array ID	***rCRS***	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***	***P***
MitoC295T	295	0.600	0.724	0.393	0.922	0.512	0.586	0.701	0.112	0.790	0.117	0.265	0.597	0.260	0.160
MitoT491C	489	0.854	0.322	0.771	0.268	0.606	0.468	0.461	0.315	0.611	0.375	0.825	0.285	0.851	0.356
MitoG752A	750	0.311	0.710	0.039	0.846	0.644	0.306	0.610	0.338	0.688	0.328	0.925	0.836	0.707	0.509
MitoG1440A	1438	0.220	0.938	0.072	0.400	0.324	0.123	0.568	0.513	0.639	0.539	0.511	0.672	0.387	0.791
MitoG1721A	1719	0.842	0.217	0.579	0.487	0.914	0.585	0.535	0.451	0.337	0.825	0.894	0.848	0.750	0.687
MitoT2160C	2158	0.954	0.088	0.345	0.391	0.228	0.310	0.274	0.427	0.705	0.149	0.851	0.809	0.565	0.224
MitoG2708A	2706	0.728	0.986	0.211	0.297	0.315	0.507	0.383	0.991	0.789	0.527	0.889	0.079	0.814	0.625
MitoG3012A	3010	0.563	0.399	0.773	0.199	0.198	0.495	0.207	0.319	0.463	0.278	0.511	0.830	0.951	0.326

MitoA3349G	3348	0.063	0.064	0.068	0.040	0.046	0.026	0.060	0.086	0.064	0.041	0.026	0.048	0.070	0.526
MitoT3395C	3394	0.622	0.451	0.442	0.892	0.273	0.893	0.825	0.503	0.864	0.348	0.933	0.375	0.724	0.893
MitoA3481G	3480	0.812	0.629	0.596	0.545	0.904	0.412	0.246	0.380	0.230	0.491	0.254	0.736	0.621	0.583
MitoT4337C	4336	0.326	0.765	0.859	0.864	0.127	0.439	0.152	0.047	0.253	*n.a.*	0.944	0.253	0.032	0.490
MitoT4562C	4561	0.322	0.387	0.542	0.820	0.763	0.421	0.602	0.069	0.140	0.070	0.130	0.335	0.017	0.659
MitoA4918G	4917	0.699	0.583	0.420	0.400	0.773	0.771	0.662	0.269	0.162	0.259	0.243	0.010	0.500	0.657
MitoT5005C	5004	0.433	0.523	0.273	0.348	0.428	0.473	0.308	0.688	0.377	0.809	0.235	0.367	0.240	0.372
MitoG5047A	5046	0.089	0.235	0.801	0.931	0.744	0.688	0.824	0.212	0.455	0.951	0.827	0.831	0.616	0.286
MitoG5461A	5460	0.168	0.931	0.769	0.159	0.964	0.542	0.279	0.219	0.216	0.898	0.573	0.866	0.933	0.307
MitoA5657G	5656	0.923	0.224	0.516	0.462	0.689	0.820	0.068	0.236	0.834	0.319	0.669	0.106	0.417	0.409
MitoT6153C	6152	0.607	0.420	0.478	0.660	0.784	0.761	0.430	0.156	0.501	0.830	0.576	0.378	0.730	0.709
MitoT6222C	6221	0.975	0.370	0.386	0.945	0.920	0.345	0.742	0.949	0.483	0.603	0.487	0.085	0.544	0.801
MitoG6261A	6260	0.970	0.093	0.602	0.274	0.248	0.250	0.499	0.212	0.111	0.189	0.930	0.200	0.026	0.634
MitoG6735A	6734	0.931	0.143	0.636	0.392	0.379	0.451	0.436	0.611	0.032	0.054	0.387	1.000	0.171	0.658
MitoA7769G	7768	0.623	0.166	0.381	0.930	0.929	0.676	0.561	0.428	0.627	0.704	0.887	0.333	0.893	0.215
MitoA8870G	8869	0.038	0.014	0.010	0.060	0.550	0.010	0.130	0.117	0.064	0.120	0.020	0.394	0.012	0.018
MitoA9668G	9667	0.240	0.302	0.582	0.902	0.592	0.392	0.560	0.866	0.261	0.693	0.179	0.479	0.497	0.232
MitoT9699C	9698	0.376	0.053	0.331	0.026	0.379	0.158	0.053	0.867	0.764	0.698	0.256	*n.a.*	*n.a.*	0.260
MitoT9717C	9716	0.656	0.063	0.151	0.569	0.869	0.065	0.686	0.128	0.315	0.016	0.431	0.650	0.056	0.756
MitoT9900C	9899	0.070	0.116	0.687	0.813	0.181	0.810	0.585	0.656	0.319	0.305	0.312	0.751	0.641	*n.a.*
MitoT10035C	10034	0.987	0.016	0.151	0.097	0.995	0.101	0.202	0.323	0.733	0.293	0.280	0.355	0.892	0.136
MitoT10239C	10238	0.846	0.031	0.237	0.171	0.928	0.253	0.178	0.397	0.727	0.420	0.400	0.360	0.874	0.587
MitoG10399A	10398	0.928	0.020	0.173	0.011	0.391	0.083	0.022	0.132	*n.a.*	0.463	0.365	*n.a.*	*n.a.*	0.165
MitoT10464C	10463	0.667	0.537	0.481	0.509	0.469	0.818	0.667	0.497	0.183	0.428	0.307	0.030	0.096	0.441
MitoA10551G	10550	0.761	0.624	0.571	0.569	0.916	0.390	0.260	0.393	0.319	0.374	0.262	0.795	0.715	0.535
MitoA11252G	11251	0.724	0.467	0.793	0.083	0.451	0.569	0.294	0.739	0.173	0.886	0.358	*n.a.*	0.147	0.382

MitoA11468G	11467	0.882	0.049	0.079	0.368	0.821	0.111	0.332	0.950	0.356	0.948	0.381	0.482	0.396	0.538
MitoA12309G	12308	0.943	0.083	0.103	0.369	0.975	0.159	0.600	0.358	0.039	0.167	0.038	0.183	0.017	0.843
MitoG12373A	12372	0.421	0.169	0.010	0.118	0.657	0.012	0.137	0.208	0.014	0.075	0.013	0.083	0.040	0.982
MitoA13106G	13105	0.119	0.267	0.158	0.417	0.845	0.091	0.118	0.786	0.892	0.015	0.042	0.235	0.020	0.837
MitoA13781G	13780	0.699	0.011	0.073	0.078	0.638	0.059	0.103	0.204	0.690	0.460	0.280	0.351	0.796	0.349
MitoA14234G	14233	0.677	0.949	0.368	0.345	0.920	0.828	0.784	0.037	0.033	0.049	0.025	*n.a.*	0.010	0.933
MitoG15044A	15043	0.679	0.186	0.208	0.416	0.947	0.489	0.158	0.225	0.943	0.897	0.394	0.479	0.593	0.705
MitoG15258A	15257	0.834	0.139	0.644	0.130	0.556	0.226	0.440	0.754	0.638	0.735	0.290	0.669	0.753	0.162
MitoA15759G	15758	0.868	0.820	0.703	0.689	0.497	0.937	0.603	0.469	0.211	0.509	0.236	0.177	0.294	0.067
MitoC15905T	15904	0.089	0.526	0.624	0.424	0.662	0.319	0.977	0.271	0.944	0.058	0.828	0.881	0.274	0.340
MitoA15925G	15924	0.680	0.010	0.154	0.108	0.319	0.217	0.038	0.714	0.943	0.485	0.741	0.630	0.633	0.978
MitoG15929A	15928	0.768	0.617	0.374	0.326	0.606	0.764	0.562	0.460	0.234	0.451	0.379	0.019	0.125	0.653
MitoG16130A	16129	0.943	0.055	0.234	0.213	0.860	0.159	0.249	0.526	0.437	0.172	0.439	0.774	0.723	0.502
MitoA16163G	16162	0.990	0.659	0.302	0.896	0.211	0.402	0.834	0.619	0.684	0.522	0.981	0.127	0.975	0.894

Table S3. Comparison of mtDNA haplogroup frequencies between cases and controls, stratified by gender, joint involvement, and method of ascertainment Uncorrected chi-squared comparison of haplogroup frequencies in cohort 1 (*ARC1*, n = 3093 and *WTCCC-58C*, n = 4300) and cohort 2 (*ARC2*, n = 2640 and *WTCCC-NBS*, n = 2482), stratified by; a) gender, b) affected joint type and c) diagnosis ascertainment.

a) Stratified by gender

	Cohort 1		Cohort 2	
Haplogroup	*Male*	*Female*	*Male*	*Female*
H	9.09E-01	6.13E-01	8.37E-01	7.53E-01
V	1.24E-01	5.67E-01	2.73E-01	9.46E-01
J	7.82E-01	2.45E-01	9.23E-01	2.90E-01
T	6.41E-01	5.21E-01	2.71E-01	1.79E-01
U	9.81E-01	7.00E-03	7.10E-02	1.76E-01
K	7.75E-01	5.77E-01	3.46E-01	2.49E-01
W	9.60E-02	2.87E-01	2.43E-01	1.45E-01
X	9.86E-01	5.17E-01	9.87E-01	1.24E-01
I	9.06E-01	1.70E-02	4.46E-01	8.99E-01
O	3.36E-01	8.70E-01	1.03E-01	8.72E-01

b) Stratified by joint type

	Cohort 1			Cohort 2		
Haplogroup	*Hip*	*Knee*	*Both*	*Hip*	*Knee*	*Both*
H	4.20E-01	1.53E-01	2.21E-01	3.25E-01	6.60E-01	1.17E-01
V	6.14E-01	4.90E-01	6.59E-01	5.90E-02	8.24E-01	8.78E-01
J	6.33E-01	2.19E-01	6.43E-01	9.56E-01	4.98E-01	1.04E-01
T	4.12E-01	3.71E-01	7.57E-01	2.75E-01	2.57E-01	1.00E-02
U	8.30E-02	1.81E-01	8.80E-01	3.70E-02	1.41E-01	2.07E-01
K	6.04E-01	5.39E-01	9.00E-01	4.51E-01	2.83E-01	7.65E-01
W	7.66E-01	9.66E-01	7.28E-01	8.14E-01	7.96E-01	9.99E-01
X	5.06E-01	9.02E-01	8.54E-01	7.84E-01	1.83E-01	2.40E-02
I	1.07E-01	1.47E-01	6.59E-01	4.67E-01	4.18E-01	2.70E-01
O	6.46E-01	6.43E-01	1.46E-01	3.60E-01	2.52E-01	5.21E-01

c) Stratified by diagnosis ascertainment

	Cohort 1		Cohort 2	
Haplogroup	Arthroplasy	X-ray	Arthroplasy	X-ray
H	8.56E-01	2.17E-01	7.47E-01	5.28E-01
V	3.61E-01	9.81E-01	2.77E-01	3.43E-01
J	3.91E-01	3.54E-01	3.99E-01	6.28E-01
T	7.72E-01	6.29E-01	5.50E-02	6.49E-01
U	2.19E-01	6.40E-02	2.00E-02	8.88E-01
K	4.20E-01	2.43E-01	6.78E-01	5.59E-01
W	6.53E-01	8.58E-01	9.56E-01	2.23E-01
X	4.17E-01	7.28E-01	2.42E-01	8.99E-01
I	1.07E-01	1.50E-01	8.14E-01	6.31E-01
O	5.53E-01	4.60E-01	1.94E-01	9.75E-01

Table S4. Chi-squared tests of association between groups of principal component analysis (PCA) clusters and disease in cohort 1 (ARC1 versus WTCCC 58C), cohort 2 (ARC2 versus WTCCC-NBS) and combined cases and controls. Uncorrected chi-squared comparison of PCA-cluster frequencies in cohort 1 (*ARC1*, n = 3099 and *WTCCC-58C*, n = 2618), cohort 2 (*ARC2*, n = 4311 and *WTCCC-NBS*, n = 2434), and combined cases (*ARC1+ARC2*, *n=7410*) versus combined controls (WTCCC-58c+-NBS, n=5052).

	Cohort 1			***Cohort 2***			***Combined***		
PCA Cluster	***F ARC1***	***F WTCCC-58C***	***P***	***F ARC2***	***WTCCCC-NBS***	***P***	***F Cases***	***F Con.***	***P***
1	400	300	1.0E-01	509	330	4.0E-02	909	630	7.6E-01
2	55	57	3.2E-01	77	51	4.2E-01	132	108	2.0E-01
3	85	96	5.6E-02	122	70	9.7E-01	207	166	1.3E-01
4	534	410	1.2E-01	701	389	7.9E-01	1235	799	2.5E-01
5	1071	926	5.4E-01	1543	850	4.9E-01	2614	1776	9.0E-01
6	330	294	5.1E-01	514	290	9.9E-01	844	584	7.9E-01
7	221	201	4.6E-01	291	178	4.1E-01	512	379	2.2E-01
8	40	32	9.0E-01	40	28	4.5E-01	80	60	6.3E-01
9	59	39	2.7E-01	70	28	1.5E-01	129	67	8.0E-02
10	304	263	8.0E-01	444	220	1.0E-01	748	483	3.4E-01
	3099	2618		4311	2434		7410	5052	