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

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.5 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.10598A>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.10598A>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 7395 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. In cohort 1, 3095 cases (ARC1) were compared with 2640 controls from the 1958 Birth Cohort (WTCCC-58C). In cohort 2, 4500 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). We excluded 50 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%. 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 =<10⁻⁵) reducing the final number of experimental SNPs to 48. Statistical significance was defined as p =<0.05 in both phases. χ² And missingness were computed using PLINK v 1.05 (http://pngu.mgh.harvard.edu/purcell/plink/).

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) was performed on the X:Y ratios of raw intensities for all SNPs for the combined dataset of cases (n=7595) and controls (n=5122) using R. The k-means function of the R cluster package was used to cluster dataset of cases (n=7393) and controls (n=5122) using R. The previously reported haplogroup associations were with specific subgroups of OA patients, subcategorised based on gender and/or specific joint involvement. 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 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. 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 homoplasys. Homoplasys is the recurrence of mutations on different branches of the mtDNA phylogeny in different parts of the world. Homoplasys accounts for up to 20% of mtDNA variation, and often involves non-synonymous substitutions. 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

### RESULTS

Using established QC criteria 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 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. 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 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. 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 homoplasys. Homoplasys is the recurrence of mutations on different branches of the mtDNA phylogeny in different parts of the world. Homoplasys accounts for up to 20% of mtDNA variation, and often involves non-synonymous substitutions. 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>
<thead>
<tr>
<th>Hap</th>
<th>F AC1</th>
<th>F WTCCC-58C</th>
<th>P</th>
<th>F AC2</th>
<th>F WTCCC-NBS</th>
<th>P</th>
<th>F cases</th>
<th>F control</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>1375 (44.5)</td>
<td>1160 (43.9)</td>
<td>5.97e-01</td>
<td>1887 (43.9)</td>
<td>1080 (43.5)</td>
<td>6.36e-01</td>
<td>3622 (44.1)</td>
<td>2240 (43.7)</td>
<td>6.67e-01</td>
</tr>
<tr>
<td>V</td>
<td>107 (3.5)</td>
<td>84 (3.2)</td>
<td>5.40e-01</td>
<td>133 (3.1)</td>
<td>86 (3.5)</td>
<td>4.25e-01</td>
<td>240 (3.2)</td>
<td>170 (3.3)</td>
<td>8.42e-01</td>
</tr>
<tr>
<td>J</td>
<td>335 (10.8)</td>
<td>311 (11.8)</td>
<td>2.81e-01</td>
<td>527 (12.3)</td>
<td>292 (11.8)</td>
<td>5.02e-01</td>
<td>862 (11.7)</td>
<td>603 (11.8)</td>
<td>8.25e-01</td>
</tr>
<tr>
<td>T</td>
<td>309 (10.0)</td>
<td>266 (10.1)</td>
<td>9.56e-01</td>
<td>449 (10.4)</td>
<td>227 (9.1)</td>
<td>7.50e-02</td>
<td>758 (10.3)</td>
<td>493 (9.6)</td>
<td>5.39e-01</td>
</tr>
<tr>
<td>U</td>
<td>440 (14.2)</td>
<td>334 (12.7)</td>
<td>7.20e-02</td>
<td>549 (12.8)</td>
<td>366 (14.7)</td>
<td>2.70e-02</td>
<td>989 (13.4)</td>
<td>700 (13.7)</td>
<td>7.35e-01</td>
</tr>
<tr>
<td>K</td>
<td>280 (9.1)</td>
<td>241 (8.1)</td>
<td>9.59e-01</td>
<td>362 (8.4)</td>
<td>216 (8.7)</td>
<td>7.32e-01</td>
<td>642 (8.7)</td>
<td>457 (8.9)</td>
<td>8.20e-01</td>
</tr>
<tr>
<td>W</td>
<td>56 (1.8)</td>
<td>46 (1.7)</td>
<td>8.30e-01</td>
<td>91 (2.1)</td>
<td>56 (2.3)</td>
<td>7.26e-01</td>
<td>147 (2.0)</td>
<td>102 (2.0)</td>
<td>3.38e-01</td>
</tr>
<tr>
<td>X</td>
<td>50 (1.6)</td>
<td>46 (1.7)</td>
<td>7.26e-01</td>
<td>76 (1.8)</td>
<td>35 (1.4)</td>
<td>2.53e-01</td>
<td>126 (1.7)</td>
<td>81 (1.6)</td>
<td>8.27e-01</td>
</tr>
<tr>
<td>I</td>
<td>88 (2.8)</td>
<td>99 (3.8)</td>
<td>5.80e-02</td>
<td>131 (3.0)</td>
<td>80 (3.2)</td>
<td>7.13e-01</td>
<td>219 (3.0)</td>
<td>179 (3.5)</td>
<td>1.19e-01</td>
</tr>
<tr>
<td>O</td>
<td>53 (1.7)</td>
<td>53 (2.0)</td>
<td>4.22e-01</td>
<td>95 (2.2)</td>
<td>44 (1.8)</td>
<td>2.11e-01</td>
<td>148 (2.0)</td>
<td>97 (1.9)</td>
<td>6.31e-01</td>
</tr>
</tbody>
</table>

Haplogroups determined manually, as described in the Methods section and online supplementary table S1. Uncorrected χ² 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 χ² probability. For study and control group acronyms see the Methods section.
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. 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
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 Sarah Cartwright 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 Sharks 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 Swale and Adam Tal for cohort 2 patient ascertainment. For Southampton we acknowledge the Wellcome Trust Clinical Research Facility at Southampton General Hospital, we thank Philippa-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 VIDE0 study. For Newcastle we acknowledge Sheryl Mitchell and the Freeman Hospital orthopastry 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 DonnaGray for organising cohort 2 patient ascertainment, and we acknowledge the Northumbria NHS Foundation Trust orthopastry consultants Mr Partrington, 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 Kiiler, 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, FD, MO, MAcC, WERO, SHR, MRR, TDS, AMW, 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 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.

**Competing interest**

None.
REFERENCES