Linkage analysis of candidate genes as susceptibility loci for osteoarthritis—suggestive linkage of COL9A1 to female hip osteoarthritis

Z. Mustafa, K. Chapman, C. Irven, A. J. Carr[1], K. Clipsham[1], J. Chitnavis[1], J. S. Sinsheimer[2], V. A. Bloomfield[1], M. McCartney[3], O. Cox[1], B. Sykes and J. Loughlin

University of Oxford, Institute of Molecular Medicine, Oxford, [1]Nuffield Orthopaedic Centre, Oxford, UK, [2]Departments of Biomathematics and Biostatistics, University of California, Los Angeles, USA and [3]Musgrave Park Hospital, Belfast, UK

Abstract

Objective. To examine 11 candidate genes as susceptibility loci for osteoarthritis (OA).

Methods. A total of 481 families have been ascertained in which at least two siblings have had joint replacement surgery of the hip, or knee, or hip and knee for idiopathic OA. Each candidate gene was targeted using one or more intragenic or closely linked microsatellite marker. The linkage data were analysed unstratified and following stratification by sex and by joint replaced (hip or knee).

Results. The analyses revealed suggestive linkage of the type IX collagen gene COL9A1 (6q12-q13) to a subset of 132 families that contained affected females who were concordant for hip OA (female-hip) with a *P*-value of 0.00053 and logarithm of the odds (LOD) score of 2.33 [corrected *P*-value of 0.0016, corrected LOD score of 1.85].

Conclusions. COL9A1 may therefore be a susceptibility locus for female hip OA. In addition, there was weak evidence of linkage to HLA/COL11A2 (6p21.3) in female hip OA with a corrected *P*-value of 0.016.

KEY WORDS: Osteoarthritis, Linkage, Affected sibling pairs, COL9A1.

Idiopathic osteoarthritis (OA) is a debilitating disease involving degeneration of the articular cartilage of synovial joints. The histology of OA varies from minor erosion of cartilage to total cartilage loss requiring joint replacement surgery [1]. OA has long been considered an inevitable consequence of ageing. However, it has become increasingly apparent that OA does have a genetic component. Twin studies have estimated that the heritability of radiographic OA of the hand, knee and hip ranges from 36 to 68% [2], whilst relative risk calculations have revealed that first-degree relatives of individuals who have had hip or knee joint replacement surgery for OA have up to a 2.3-fold increased risk of developing end-stage disease [3]. These studies also highlight the fact that environmental factors are important in the aetiology of OA, placing idiopathic OA into the common, multifactorial class of genetic diseases.

It has often been noted in epidemiological studies that there is a female preponderance for OA. This may be accounted for by differential effects on the two sexes of environmental factors. However, a Finnish twin study has suggested that genetic susceptibility may be greater in women than men [4] and this result has been supported by a segregation analysis [5]. These studies have tended to concentrate on patients with hand, knee, or hand and knee OA. Some studies have attempted to quantify the risks of developing OA for specific joints and differences in the heritability values between joint groups have been reported [3, 6, 7]. These results suggest that there may be genetic heterogeneity of OA between the sexes and between different joint groups.

A number of candidate genes have been proposed as potential susceptibility loci for idiopathic OA. These genes predominantly encode structural proteins of the extracellular matrix. Particular emphasis has been placed on the gene for type II collagen (COL2A1) which is the major cartilage collagen. There are two principal reasons for this: (i) COL2A1 encodes the most abundant protein of the cartilage extracellular matrix and (ii) mutations in COL2A1 have been found in severa osteochondrodysplasias. Osteo-chondrodysplasias are rare monogenic diseases of cartilage and bone which often have OA as one of their principal phenotypic

components. The involvement of COL2A1 in the osteo-chondrodysplasia ocular-Stickler syndrome is particularly interesting as this disease involves preco-cious, early-onset OA [8]. Less severe mutations of the COL2A1 gene could account for some cases of idiopathic OA. Indeed, linkage has been established to COL2A1 in a couple of rare families in which early-onset OA segregates as a Mendelian trait [9]. Furthermore, null alleles of COL2A1 have been reported in some individuals with idiopathic OA [10], whilst two studies have suggested association of polymorphisms in the vitamin D receptor gene (which is syntenic with COL2A1) with OA [11, 12]. However, other studies have excluded COL2A1 as a high impact susceptibility locus [13].

Other candidates proposed as potential susceptibility loci include genes that encode additional structural proteins of the cartilage extracellular matrix, particularly those that, like COL2A1, have been implicated in osteo-chondrodysplasias. These include the type XI collagen genes COL11A1 and COL11A2 (non-ocular Stickler disease) [14, 15], the type IX collagen gene COL9A2 (multiple epiphyseal dysplasia) [16] and the cartilage oligomeric matrix protein gene, COMP (multiple epiphy-seal dysplasia and pseudoachondroplasia) [17].

Two loci which do not encode cartilage structural proteins but which have been considered as potential candidates are the HLA gene cluster on chromosome 6p and the α1-antitrypsin gene on chromosome 14q. These loci have been proposed on the basis that a subset of OA which is characterized by the presence of multiple Heberden's nodes on the distal interphalangeal joints (nodal generalized OA) may have an inflammatory component. Tentative association was established in a small group of unrelated affected individuals [18].

Since idiopathic OA rarely segregates as a Mendelian trait, and as both genetic and environmental factors play a role in the aetiology of the disease, its genetic dissection is not readily amenable to conventional family-based parametric linkage analysis. Instead, it is more suitable to the affected sibling pair linkage approach which does not specify the mode of inheritance.

There are a number of candidates which merit analysis as susceptibility loci for OA and in this paper we describe the results of an analysis of 11 such candidate genes. To test for linkage we used an affected sibling pair approach and recruited affected pairs using joint replacement surgery for idiopathic OA as our ascertain-ment criterion. Our aim was to use families whose idiopathic OA was severe and is therefore more likely to have involved a genetic component. We used a two-stage approach. In the first stage all candidate micro-satellites were genotyped in 297 of our 481 families. Any candidate that demonstrated evidence for linkage at a *P*-value $\leqslant 0.05$ in the 297 families or in a stratified subset of the families (stratified by sex and by hip or knee joint replacement) was then analysed in the remaining 184 families.

Patients and methods

OA families

Families with at least two siblings who had undergone one or more replacements of the total hip, or of the total knee, or both, for primary idiopathic OA were recruited. A detailed description of these families has been published elsewhere [19].

Candidate genes

The candidate genes examined are shown in Table 1, which also lists the polymorphic markers used to target each candidate. Where available, intragenic markers were used. Other markers were chosen that were 5 cM or less from the candidate gene. For the COL2A1/VDR locus, the variable number of tandem repeat (VNTR) marker located immediately downstream of the COL2A1 gene [28] and two additional markers were used to cover chromosome 12q13-q14 (D12S368 and D12S87).

Polymerase chain reaction (PCR), gel electrophoresis and genotyping

Either the forward or the reverse primer in a PCR pair was fluorescently labelled and amplification products were electrophoresed through Applied Biosystems 373 or 377 Automated DNA Sequencers as described by Reed *et al.* [29]. Alleles were sized using Applied Biosystems Genescan version 2.0.2 and Genotyper version 1.1 software.

Linkage analysis strategy

The linkage analysis strategy was to genotype all markers in a first stage containing 297 of our 481 families. Any marker that had a *P*-value $\leqslant 0.05$ in stage 1 would then be examined in the remaining 184 families (stage 2). The aim of this strategy was to take through to stage 2 only those markers that demonstrated reasonable evidence of linkage. In stage 2 we were not necessarily expecting to repeat any positive linkage results of stage 1 but instead were looking for further evidence of linkage, even if only moderate: if a marker's *P*-value for stages 1 and 2 combined was less than the *P*-value for stage 1 only then this would support linkage at that marker. There was no difference in the ascertain-ment criteria between the stage 1 and stage 2 families— once a reasonably large number of families had been collected we began stage 1 of our linkage strategy.

As well as analysing all the stage 1 families together, these families were also stratified into those containing affected females-only (132 families), affected males-only (60 families), hips-only (male and/or female) (194 famil-ies), knees-only (male and/or female) (34 families), affected females-only who had undergone hip replace-ment but not knee replacement (female-hip) (85 families) and affected males-only who had undergone hip replace-ment but not knee replacement (male-hip) (44 families). We did not stratify by female-knee or male-knee as the number of families was too low (16 and four, respect-ively) to allow reliable inference of linkage. Again, any

TABLE 1. The candidate genes examined, their chromosomal location, the respective polypeptides and the polymorphic markers used to target each gene

Gene/locus	Chromosome	Polypeptide	Marker
COL2A1	12q13-q14	α1 chain type II collagen	COL2A1 3′VNTR [20], D12S368[a], D12S87[a]
VDR	12q13-q14	Vitamin D receptor	As for COL2A1[b]
COL9A1	6q12-q13	α1 chain type IX collagen	509–8B2, 509–12B1 [21]
COL9A2	1p32	α2 chain type IX collagen	L-myc poly(A) [22], D1S2722[a], D1S463[a]
COL11A1	1p21	α1 chain type XI collagen	Intragenic $(CA)_n$ [23], D1S248[a], D1S239[a]
COL11A2	6p21.3	α2 chain type XI collagen	D6S265[a], D6S273[a]
HLA	6p21.3	Human leucocyte antigen cluster	As for COL11A2[b]
COMP	19p13.1	Cartilage oligomeric matrix protein	Intragenic $(GAAA)_n$ [24], D19S212[a], D19S215[a]
CRTL1	5q13–14.1	Cartilage link protein	Intragenic $(CA)_n$ [25]
α1AT (PI locus)	14q32.1	α1-antitrypsin	Intragenic $(CA)_n$ [26]
DTDST	5q31-q34	Diastrophic dysplasia sulphate transporter	CSF1R $(CA)_n$, CSF1R $(TAGA)_n$ [27]

[a]http://www.hgmp.mrc.ac.uk/gdb/gdbtop.html.
[b]COL2A1 and VDR are genetically tightly linked as are COL11A2 and the HLA.

marker that had a P-value $\leqslant 0.05$ following stratification in stage 1 would then be examined in the stage 2 families.

Stratification criteria

We stratified by sex, by joint replaced (hip or knee) and by sex combined with joint replaced. The stratification criteria have been published elsewhere [19].

Statistical analysis

Linkage analysis was performed using the SIBPAIR module of the ANALYZE package (ftp://linkage.cpmc. columbia.edu). This module is able to use data from siblings to determine identity by descent (IBD) allele transmittance. In the linkage analysis siblings who had not undergone joint replacement were given a clinical status of unknown. The SIBPAIR module produces a single-point logarithm of the odds (LOD) score and its asymptotic P-value. Allele frequencies were calculated from the input data using GAS (http://users.ox.ac.uk/ayoung/gas.html). As well as the unstratified data we tested six strata in stage 1, three strata in stage 2 and three strata when stages 1 and 2 were combined. We therefore adjusted the stratified LOD scores and P-values to correct for the number of strata tested [30]. For stage 1 we deducted $\log 6 = 0.78$ from original LOD scores and multiplied original P-values by 6, whilst for stage 2, and for stage 1 and 2 combined, we deducted $\log 3 = 0.48$ from original LOD scores and multiplied original P-values by 3.

Results

Stage 1

Unstratified linkage analysis. The unstratified LOD scores and P-values obtained for each marker in the 297 families of stage 1 are given in Table 2. Also included are the mean IBD allele sharing and the polymorphism information content (PIC) values of the markers calculated from stage 1. The L-myc poly(A) repeat polymorphism is reported to have alleles that differ by 2 bp [22]. However, in the current analysis several alleles differed by only 1 bp and it was not possible to assign

TABLE 2. Linkage results from stage 1 (*$P \leqslant 0.05$)

Gene	Marker	LOD	P	IBD[a]	PIC[b]
COL2A1 & VDR	VNTR	0.01	0.40	0.51	0.57
	D12S368	0.00	0.50	0.50	0.79
	D12S87	0.25	0.14	0.53	0.74
COL9A1	509–8B2	0.49	0.068	0.53	0.72
	509–12B1	0.24	0.15	0.52	0.76
COL9A2	L-myc	–	–	–	–
	D1S2722	0.26	0.14	0.52	0.83
	D1S463	0.00	0.50	0.50	0.70
COL11A1	Intragenic	0.05	0.32	0.51	0.63
	D1S248	0.02	0.37	0.51	0.77
	D1S239	0.08	0.27	0.52	0.72
HLA & COL11A2	D6S265	0.47	0.07	0.54	0.74
	D6S273	0.98	0.017*	0.55	0.73
COMP	Intragenic	0.07	0.28	0.52	0.82
	D19S212	0.00	0.50	0.50	0.68
	D19S215	0.07	0.29	0.52	0.80
CRTL1	Intragenic	0.03	0.35	0.51	0.72
α1-antitrypsin	Intragenic	0.01	0.43	0.51	0.86
DTDST	$(CA)_n$	1.20	0.0093*	0.56	0.78
	$(TAGA)_n$	0.00	0.50	0.50	0.68

[a]Mean identity by descent estimates.
[b]Polymorphism information content of marker.

these alleles unambiguously. In consequence this marker was abandoned.

Two of the 19 markers had a P-value $\leqslant 0.05$: the HLA/COL11A2 marker D6S273 ($P = 0.017$) and the diastrophic dysplasia sulphate transporter gene (DTDST) marker CSF1R$(CA)_n$ ($P = 0.0093$).

Stratified linkage analysis. Table 3 lists the uncorrected P-values and LOD scores for the stratified data obtained for each marker in the 297 families of stage 1. Twenty of the 114 P-values were $\leqslant 0.05$. Following correction for multiple testing [30], five of these 20 P-values were still $\leqslant 0.05$, representing three of the 11 candidates: COL9A1 marker 509–8B2 for hips-only ($P = 0.042$) and for female-hips ($P = 0.028$), HLA/COL11A2 marker D6S265 for hips-only ($P = 0.024$) and HLA/COL11A2 marker D6S273 for females-only ($P = 0.018$) and for female-hips ($P = 0.014$) (Table 4).

Overall, microsatellite markers for four of our 11 candidate genes had evidence for linkage at $P \leqslant 0.05$,

TABLE 3. Stratification of stage 1 (LOD/*P*-value). LOD scores and *P*-values are not corrected for multiple tests. When LOD score is zero - has been inserted (*$P \leqslant 0.05$)

Gene	Marker	Females-only ($n = 132$)	Males-only ($n = 60$)	Hips-only ($n = 194$)	Knees-only ($n = 34$)	Female-hips ($n = 85$)	Male-hips ($n = 44$)
COL2A1 & VDR	VNTR	–	0.15/0.20	–	0.02/0.38	–	0.08/0.28
	D12S368	–	–	–	0.32/0.11	–	–
	D12S87	0.63/0.044*	–	–	0.82/0.026*	0.02/0.38	–
COL9A1	509–8B2	0.90/0.02*	–	1.31/0.007*	–	1.47/0.0047*	–
	509–12B1	0.11/0.24	–	0.95/0.018*	–	1.05/0.014*	–
COL9A2	D1S2722	0.43/0.079	0.18/0.18	0.55/0.055	–	0.64/0.043*	0.39/0.09
	D1S463	0.08/0.28	0.03/0.35	–	–	0.25/0.14	0.04/0.33
COL11A1	Intragenic	0.08/0.27	–	–	0.25/0.14	–	–
	D1S248	0.09/0.26	–	–	–	0.04/0.34	–
	D1S239	0.75/0.03*	–	0.01/0.42	–	0.48/0.069	–
HLA & COL11A2	D6S265	0.48/0.069	–	1.50/0.004*	–	1.11/0.011*	0.10/0.25
	D6S273	1.62/0.003*	–	0.62/0.046*	0.65/0.041*	1.73/0.0024*	–
COMP	Intragenic	–	–	0.03/0.36	–	–	0.07/0.28
	D19S212	–	–	–	–	–	0.03/0.35
	D19S215	0.09/0.26	–	–	0.99/0.016*	–	–
CRTL1	Intragenic	–	–	–	0.76/0.030*	–	–
α1-antitrypsin	Intragenic	–	–	–	–	–	–
DTDST	(CA)$_n$	0.54/0.058	–	0.87/0.02*	0.24/0.15	0.60/0.049*	–
	(TAGA)$_n$	0.12/0.23	–	–	0.65/0.042*	0.06/0.30	–

either in the unstratified data (HLA/COL11A2 and DTDST) or in the stratified data (COL9A1 and HLA/COL11A2). The six markers targeting these four genes were therefore genotyped in the 184 families of stage 2.

Stage 2

Unstratified linkage analysis. None of the six markers had a *P*-value $\leqslant 0.05$ in the unstratified families of stage 2 (Table 5). After combining data for screens 1 and 2, only DTDST marker (CA)$_n$ had a *P*-value $\leqslant 0.05$ ($P = 0.022$). This compares with a *P*-value of 0.0093 for stage 1 only—increasing the number of families studied had reduced the evidence for linkage at this marker.

Stratified linkage analysis. In the corrected results from the stage 1 stratification COL9A1 and HLA/COL11A2 markers were significant at $P \leqslant 0.05$ in

one or more of the following strata: females-only, hips-only or female-hips. We therefore restricted our stratification analysis of COL9A1 and HLA/COL11A2 in stage 2, and stages 1 and 2 combined, to these three strata.

In stage 2, the COL9A1 marker 509–8B2 had uncorrected *P*-values $\leqslant 0.05$ for females-only ($P = 0.034$) and for female-hips ($P = 0.024$) (Table 6). However, following correction for the three tests performed, none of the four markers targeting COL9A1 and HLA/COL11A2 had a *P*-value $\leqslant 0.05$ in any of the three strata, although 509–8B2 did approach significance in females-only ($P = 0.10$) and female-hips ($P = 0.072$). When stages 1 and 2 were combined (Table 7) 509–8B2 had a corrected *P*-value $\leqslant 0.05$ in females-only ($P = 0.01$) and female-hips ($P = 0.0016$) and approached significance in hips-only ($P = 0.051$). 509–12B1 was also significant for

TABLE 4. Stratification of stage 1 with LOD scores and *P*-values corrected for six multiple tests (LOD/*P*-value). The results for COL9A1 and HLA/COL11A2 only are listed as these were the only loci whose markers had corrected $P \leqslant 0.05$ (*$P \leqslant 0.05$)

Gene	Marker	Females-only ($n = 132$)	Males-only ($n = 60$)	Hips-only ($n = 194$)	Knees-only ($n = 34$)	Female-hips ($n = 85$)	Male-hips ($n = 44$)
COL9A1	509–8B2	0.12/0.12	0.00/0.5	0.53/0.042*	0.00/0.5	0.69/0.028*	0.00/0.5
	509–12B1	0.00/0.50	0.00/0.5	0.17/0.11	0.00/0.5	0.27/0.084	0.00/0.5
HLA & COL11A2	D6S265	0.00/0.41	0.00/0.5	0.72/0.024*	0.00/0.5	0.33/0.066	0.00/0.5
	D6S273	0.84/0.018*	0.00/0.5	0.00/0.28	0.00/0.25	0.95/0.014*	0.00/0.5

TABLE 5. Linkage results for stage 2 and for stages 1 and 2 combined (*$P \leqslant 0.05$)

Gene	Marker	Stage 2				Stages 1 & 2 combined			
		LOD	*P*	IBD	PIC	LOD	*P*	IBD	PIC
COL9A1	509–8B2	0.04	0.33	0.51	0.73	0.46	0.07	0.53	0.72
	509–12B1	0.00	0.50	0.50	0.75	0.07	0.29	0.51	0.76
HLA & COL11A2	D6S265	0.00	0.50	0.50	0.72	0.28	0.13	0.52	0.73
	D6S273	0.00	0.50	0.50	0.76	0.44	0.08	0.53	0.74
DTDST	(CA)$_n$	0.03	0.35	0.51	0.77	0.88	0.022*	0.54	0.78
	(TAGA)$_n$	0.03	0.35	0.51	0.68	0.00	0.46	0.50	0.68

TABLE 6. Stratification of stage 2 for COL9A1 and HLA/COL11A2 with uncorrected LOD scores and *P*-values and with LOD scores and *P*-values corrected for the three multiple tests (*$P \leqslant 0.05$) (LOD/*P*-value)

Gene	Marker	Females-only ($n = 64$)		Hips-only ($n = 117$)		Female-hips ($n = 47$)	
		Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected
COL9A1	509–8B2	0.73/0.034*	0.25/0.10	0.01/0.43	0.00/0.50	0.86/0.024*	0.38/0.072
	509–12B1	0.01/0.40	0.00/0.50	0.00/0.50	0.00/0.50	0.09/0.26	0.00/0.50
HLA & COL11A2	D6S265	0.47/0.071	0.00/0.21	0.00/0.50	0.00/0.50	0.19/0.18	0.00/0.50
	D6S273	0.10/0.24	0.00/0.50	0.00/0.50	0.00/0.50	0.04/0.33	0.00/0.50

TABLE 7. Stratification of stages 1 and 2 combined for COL9A1 and HLA/COL11A2, with uncorrected LOD scores and *P*-values and with LOD scores and *P*-values corrected for the three multiple tests (*$P \leqslant 0.05$) (LOD/*P*-value)

Gene	Marker	Females-only ($n = 196$)		Hips-only ($n = 311$)		Female-hips ($n = 132$)	
		Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected
COL9A1	509–8B2	1.60/0.0033*	1.12/0.01*	0.97/0.017*	0.49/0.051	2.33/0.00053*	1.85/0.0016*
	509–12B1	0.13/0.22	0.00/0.50	0.47/0.07	0.00/0.21	1.03/0.014*	0.55/0.042*
HLA & COL11A2	D6S265	0.96/0.018*	0.48/0.054	0.60/0.049*	0.12/0.15	1.25/0.0083*	0.77/0.025*
	D6S273	1.59/0.0034*	1.11/0.01*	0.18/0.18	0.00/0.50	1.42/0.0053*	0.94/0.016*

female-hips ($P = 0.042$). The HLA/COL11A2 marker D6S273 had a corrected *P*-value $\leqslant 0.05$ in females-only ($P = 0.01$) and female-hips ($P = 0.016$). D6S265 was also significant in female-hips ($P = 0.025$) and approached significance in females-only ($P = 0.054$).

A comparison of the results for stage 1 only with the results of stages 1 and 2 combined can assist in determining whether the positive results for COL9A1 and HLA/COL11A2 are true or false: if the positive results from stage 1 are false, adding more families should reduce their significance. Table 8 is a comparison of the corrected *P*-values between stage 1 and stages 1 and 2 combined. For the COL9A1 marker 509–8B2, the *P*-value in stages 1 and 2 combined was less than the *P*-value in stage 1 for females-only (0.01 *vs* 0.12) and for female-hips (0.0016 *vs* 0.028), and was only slightly increased in hips-only (0.051 *vs* 0.042). For 509–12B1, the *P*-value for the combined was also less than the stage 1 *P*-value for female-hips (0.042 *vs* 0.084). For the HLA/COL11A2 marker D6S265, the *P*-value for the combined was less than the stage 1 *P*-value for females-only (0.054 *vs* 0.41) and for female-hips (0.025 *vs* 0.066) but increased for hips-only (0.15 *vs* 0.024). A similar pattern was observed for D6S273, with a decreased combined *P*-value in females-only (0.01 *vs* 0.018), a slightly increased *P*-value in female-hips (0.016 *vs* 0.014) and an increased *P*-value in hips-only (0.50 *vs* 0.28).

Overall, these comparisons indicate that increasing the number of families studied increases the evidence for linkage at the COL9A1 and HLA/COL11A2 loci in females-only and female-hips. This suggests that our results for these loci in these two strata are real rather than false positives.

There was a greater number of affected females-only than affected males-only in our stage 1 cohort of 297 families (132 female *vs* 60 male), in our stage 2 cohort of 184 families (64 female *vs* 42 male) and in the combined cohort of 481 families (196 female *vs* 102 male). This greater number of females-only would provide this strata with greater power to detect linkage and may account for our ability to detect linkage in females-only but not in males-only. However, the difference in the stage 1 *P*-values for the COL9A1 and HLA/COL11A2 markers between the two strata is striking with no evidence for linkage in males-only (*P*-values = 0.50 and LOD scores = 0.00 for each marker, Table 3). Although the 42 males-only families of stage 2 were not a strata specifically tested in stage 2 (and therefore stages 1 and 2 combined), these families were genotyped by default since a proportion of them constituted the hips-only families. Again, the COL9A1 and HLA/COL11A2 markers each had a *P*-value of 0.50 (LOD = 0.00) in the 102 males-only families of stages 1 and 2 combined. Another factor that could

TABLE 8. Corrected *P*-values of stage 1 compared with corrected *P*-values of stage 1 and 2 combined for COL9A1 and HLA/COL11A2 (*P*-value in stage 1 and 2 combined is less than *P*-value in stage 1 only)

Gene	Marker	Females-only		Hips-only		Female-hips	
		Stage 1 ($n = 132$)	Stage 1 + 2 ($n = 196$)	Stage 1 ($n = 194$)	Stage 1 + 2 ($n = 311$)	Stage 1 ($n = 85$)	Stage 1 + 2 ($n = 132$)
COL9A1	509–8B2	0.12	0.01*	0.042	0.051	0.028	0.0016*
	509–12B1	0.50	0.50	0.11	0.21	0.084	0.042*
HLA & COL11A2	D6S265	0.41	0.054*	0.024	0.15	0.066	0.025*
	D6S273	0.018	0.01*	0.28	0.50	0.014	0.016

affect power is the informativeness of the families and the degree to which IBD status can be determined, which will be influenced by the number of siblings typed who have an 'unknown' clinical status. There was, however, no significant difference in these numbers between females-only and males-only (data not shown). Furthermore, the PIC values for the COL9A1 and HLA/COL11A2 markers in females-only and males-only were not significantly different (data not shown). Of our total of 481 families, 185 were affected pairs that contained an affected brother and affected sister and so were not used in the stratification analysis. Linkage to the COL9A1 and HLA/COL11A2 markers was not significant in these affected pairs [$P = 0.14$, PIC = 0.72 for 509–8B2, $P = 0.50$, PIC = 0.76 for 509–12B1, $P = 0.50$, PIC = 0.72 for D6S265, $P = 0.40$, PIC = 0.75 for D6S273 (P-values are uncorrected)]. Overall, these results suggest that the specific nature of the COL9A1 and HLA/COL11A2 linkages with their restriction to females-only is not an artefact of power differences between the strata. Regarding the apparent linkage differences between hips-only and knees-only, there were significantly more hips-only families than knees-only families (194 *vs* 34 in stage 1) and this could account for our inability to detect linkage in the knees-only strata.

Nodal OA

As mentioned in the Introduction, the association that has previously been reported for both the HLA gene cluster and the α1-antitrypsin gene is to nodal generalized OA [18]. We therefore analysed the D6S265, D6S273 and the intragenic α1-antitrypsin markers in those families from stage 1 which contained at least two affected siblings who were concordant for the presence of three or more Heberden's nodes. Of the 297 families in stage 1, 29 were concordant for this phenotype and the P-values obtained for each marker were >0.1 (uncorrected).

Discussion

Linkage analysis has proved to be an essential tool in the genetic dissection of rare monogenic diseases and is now being applied to the identification of loci that predispose to the common, complex disorders. Using highly polymorphic markers and large numbers of families with small numbers of affected relatives it is possible to detect reproducible distortions of allele transmittance that indicate the presence of a nearby susceptibility locus. If the pathology of a disease is reasonably well understood, it is possible to target genes as susceptibility loci. For OA there are several strong candidates that merit targeting. In this study candidates were chosen on the basis that they had previously been implicated in idiopathic OA (or in a rare monogenic disease in which OA is one of the phenotypic components) and that intragenic or closely linked polymorphic markers were available to test for linkage.

Of the 11 genes targeted, two demonstrated evidence of suggestive linkage at a P-value of $\leqslant 0.05$ when the 297 families of stage 1 were analysed without stratification: the HLA/COL11A2 locus and DTDST. Analysis of these two genes in an additional 184 families demonstrated no evidence of linkage and when the data for stages 1 and 2 were combined only the DTDST gene still supported linkage at $P \leqslant 0.05$. However, the P-value for the DTDST $(CA)_n$ marker was greater in the combined stages than in stage 1 only, suggesting that the positive result in stage 1 was probably a false positive.

When we stratified our stage 1 results and corrected for multiple testing, the COL9A1 marker 509–8B2 was significant in hips-only and female-hips, the HLA/COL11A2 marker D6S265 was significant in hips-only and female-hips and the HLA/COL11A2 marker D6S273 was significant in females-only. No markers targeting the remaining eight candidate genes were significant in the six strata tested. Despite correction, these positive stratification results could still be false. The two-stage nature of our linkage analysis strategy enabled us to test this possibility in that we were able to genotype the COL9A1 and HLA/COL11A2 markers in additional families. After correction, none of the four markers genotyped in stage 2 were significant in any of the three strata tested. However, this was not unexpected: to have repeated the significant results in an independent data set (which had lower numbers in each of the strata tested) would have been surprising [31]. Nevertheless, 509–8B2 did approach significance in females-only and female-hips. A more informative analysis was the comparison between the P-values obtained in stage 1 with those of stages 1 and 2 combined: a reduction in P-values when more families were genotyped (i.e. when the power of the study was increased) would support the positive results of stage 1. These comparisons revealed a reduced P-value in the combined *vs* stage 1 for 509–8B2 in females-only and female-hips (with a very moderate increase in P-value for hips-only), a reduced P-value for the second COL9A1 marker 509–12B1 in female-hips, a reduced P-value for the two HLA/COL11A2 markers (D6S265 and D6S273) in females-only with a reduced P-value for D6S265 in female-hips and only a very moderate increase in the P-value for D6S273 in this strata. Overall, therefore, increasing the number of families increased the evidence for linkage of the COL9A1 and the HLA/COL11A2 markers in females-only and female-hips, suggesting that these positive results are probably real.

Why the linkage should be restricted to females-only deserves comment, particularly so for COL9A1 where the difference between females-only and males-only is particularly striking. The obvious possibility is that the susceptibility at COL9A1 is mediated through a pathway that involves a female-specific hormonal element. Females are at a greater risk than males of developing OA, and this risk increases after the menopause [1]. Indeed, hormone replacement therapy can affect the development and/or progression of OA in females [32].

Detailed analysis of the regulation of the COL9A1 gene, with particular emphasis on controlling elements that have a hormonal component, may be helpful in dissecting the apparent female bias in susceptibility at this locus.

Type IX collagen is a heterotrimeric protein composed of α1(IX), α2(IX) and α3(IX) polypeptide chains that are encoded by the COL9A1, COL9A2 and COL9A3 genes, respectively. Type IX collagen is a quantitatively minor cartilage collagen that decorates the type II collagen fibrils and this interaction serves to regulate the growth of the type II fibril. Type IX is composed of three collagenous (COL1–3) and four non-collagenous domains (NC1–4) with the NC4 domain encoded by COL9A1 being significantly larger than the NC4 domains encoded by COL9A2 or COL9A3. This large COL9A1 NC4 domain may interact with cartilage proteoglycans and could help stabilize the cartilage extracellular matrix. Two mouse models have shed light on to the function of type IX collagen, and more specifically the COL9A1-encoded α1(IX) chain. In the first model, a truncated form of col9a1 resulted in a mild osteo-chondrodysplasia with OA [33] whilst in the second model, a col9a1 knock-out mouse had no congenital abnormality but developed severe OA that was comparable in timing and pathology to human idiopathic OA [34]. These models highlighted the importance of the α1(IX) chain in stabilizing cartilage and demonstrated that mutations in Col9a1 can result in an OA phenotype.

Overall, we have identified two genetic loci that demonstrate evidence of suggestive linkage to severe hip OA in females. The COL9A1 gene sequence is available in public databases and this can be used for the identification of coding and regulatory single nucleotide DNA variants that would be a starting point for detailed association analysis of this gene with OA.

Stratification increases the level of genetic homogeneity and can therefore assist in the mapping of loci for complex traits. Our analysis highlights the potential utility of this approach for OA.

Our analysis cannot exclude as susceptibility loci all the genes that were negative in this study. It may be that one or more of these genes is of low impact in the severe end-stage form of OA examined and that we lacked the power to detect this effect. Alternatively, they may encode susceptibility for different forms of OA (i.e. hand or disc) that we did not examine.

Acknowledgements

References

1. Creamer P, Hochberg MC. Osteoarthritis. Lancet 1997;350:503–8.
2. MacGregor AJ, Spector TD. Twins and the genetic architecture of osteoarthritis. Rheumatology 1999;38:583–90.
3. Chitnavis J, Sinsheimer JS, Clipsham K et al. Genetic influences in end-stage osteoarthritis. Sibling risks of hip and knee replacement for idiopathic osteoarthritis. J Bone Joint Surg 1997;79B:660–4.
4. Kaprio J, Kujala UM, Peltonen L, Koskenvuo M. Genetic liability to osteoarthritis may be greater in women than men. Br Med J 1996;313:232.
5. Felson DT, Couropmitree NN, Chaisson CE et al. Evidence for a Mendelian gene in a segregation analysis of generalized radiographic osteoarthritis. Arthritis Rheum 1998;41:1064–71.
6. Lindberg H. Prevalence of primary coxarthrosis in siblings of patients with primary coxarthrosis. Clin Orthop 1986;203:273–5.
7. Cooper C, McAlindon T, Snow S et al. Mechanical and constitutional risk factors for symptomatic knee osteoarthritis: differences between medial tibio-femoral and patello-femoral disease. J Rheumatol 1994;21:307–13.
8. Williams CJ, Ganguly A, Considine E et al. A-2→G transition at the 3′ acceptor splice site of IVS17 characterizes the COL2A1 gene mutation in the original Stickler syndrome kindred. Am J Med Genet 1996;63:461–7.
9. Vikkula M, Palotie A, Ritvaniemi P et al. Early-onset osteoarthritis linked to the type II procollagen gene. Arthritis Rheum 1993;36:401–9.
10. Loughlin J, Irven C, Athanosou N, Carr A, Sykes B. Differential allelic expression of the type II collagen gene (COL2A1) in osteoarthritic cartilage. Am J Hum Genet 1995;56:1186–93.
11. Keen RW, Hart DJ, Lanchbury JS, Spector TD. Association of early osteoarthritis of the knee with a Taq I polymorphism of the vitamin D receptor gene. Arthritis Rheum 1997;40:1444–9.
12. Uitterlinden AG, Burger H, Huang Q et al. Vitamin D receptor genotype is associated with radiographic osteoarthritis at the knee. J Clin Invest 1997;100:259–63.
13. Loughlin J, Irven C, Fergusson C, Sykes B. Sibling pair analysis shows no linkage of generalized osteoarthritis to the loci encoding type II collagen, cartilage link protein or cartilage matrix protein. Br J Rheumatol 1994;33:1103–6.
14. Richards AJ, Yates JR, Williams R et al. A family with Stickler syndrome type 2 has a mutation in the COL11A1 gene resulting in the substitution of glycine 97 by valine in α1(XI) collagen. Hum Mol Genet 1996;5:1339–43.
15. Vikkula M, Mariman ECM, Lui VCH et al. Autosomal dominant and recessive osteochondrodysplasias associated with the COL11A2 locus. Cell 1995;80:431–7.
16. Holden P, Canty EG, Mortier GR et al. Identification of novel pro-α2 (IX) collagen gene mutations in two families with distinctive oligo-epiphyseal forms of multiple epiphyseal dysplasia. Am J Hum Genet 1999;65:31–8.
17. Loughlin J, Irven C, Mustafa Z et al. Identification of five novel mutations in the cartilage oligomeric matrix protein gene in pseudoachondroplasia and multiple epiphyseal dysplasia. Hum Mutat 1998;suppl.1:S10–17.
18. Pattrick M, Manhire A, Ward AM, Doherty M. HLA-A, B antigens and α1-antitrypsin phenotypes in nodal generalised osteoarthritis and erosive osteoarthritis. Ann Rheum Dis 1989;48:470–5.

19. Chapman K, Mustafa Z, Irven C *et al.* Osteoarthritis-susceptibility locus on chromosome 11q, detected by linkage. Am J Hum Genet 1999;65:167–74.
20. Priestley L, Kumar D, Sykes B. Amplification of the COL2A1 3′ variable region used for segregation analysis in a family with the Stickler syndrome. Hum Genet 1990;85:525–6.
21. Warman ML, Tiller GE, Polumbo PA *et al.* Physical and linkage mapping of the human and murine genes for the α1 chain of type IX collagen (COL9A1). Genomics 1993;17:694–8.
22. Makela TP, Hellsten E, Vesa J, Alitalo K, Peltonen L. An Alu variable polyA repeat polymorphism upstream of L-myc at 1p32. Hum Mol Genet 1992;1:217.
23. Griffith AJ, Sprunger LK, SirkoOsadsa DA, Tiller GE, Meisler MH, Warman ML. Marshall syndrome associated with a splicing defect at the COL11A1 locus. Am J Hum Genet 1998;62:816–23.
24. Briggs MD, Hoffman SM, King LM *et al.* Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. Nat Genet 1995;10:330–6.
25. Hecht JT, Wang Y, Rhodes C, Yamada Y. GT repeat polymorphism in the human proteoglycan link gene (CRTL1) promoter region. Nucleic Acids Res 1991;19:6666.
26. Byth BC, Cox DW. A (CA)n repeat polymorphism at the 5′ end of the α1-antitrypsin gene (PI). Hum Mol Genet 1993;2:1752.
27. Hastbacka J, de la Chapelle A, Mahtani MM *et al.* The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure disequilibrium mapping. Cell 1994;78:1073–87.
28. Stoker NG, Cheah KSE, Griffin JR, Pope FM, Solomon E. A highly polymorphic region 3′ to the human type II collagen gene. Nucleic Acids Res 1985;13:4613–22.
29. Reed PW, Davies JL, Copeman JB *et al.* Chromosome-specific microsatellite sets for fluorescence-based, semiautomated genome mapping. Nat Genet 1994;7:390–5.
30. Lernmark A, Ott J. Sometimes it's hot, sometimes it's not. Nat Genet 1998;19:213–4.
31. Suarez B, Hampe C, Van Eerdewegh P. In: Gershon ES, Cloninger CR, eds. Genetic approaches to mental disorders. Washington, DC: American Psychiatric Press, 1994:23–46.
32. Dennison EM, Arden NK, Kellingray S, Croft P, Coggon D, Cooper C. Hormone replacement therapy, other reproductive variables and symptomatic hip osteoarthritis in elderly white women: a case–control study. Br J Rheumatol 1998;37:1198–1202.
33. Nakata K, Ono K, Miyazaki J-I *et al.* Osteoarthritis associated with mild chondrodysplasia in transgenic mice expressing α1(IX) collagen chains with a central deletion. Proc Natl Acad Sci USA 1993;90:2870–4.
34. Fassler R, Schnegelsberg PNJ, Dausman J *et al.* Mice lacking α1(IX) collagen develop noninflammatory degenerative joint disease. Proc Natl Acad Sci USA 1994;91:5070–4.