



## The additive effect of individual genes in predicting risk of knee osteoarthritis

A M Valdes, M Doherty and T D Spector

*Ann Rheum Dis* 2008;67;124-127; originally published online 17 Aug 2007;  
doi:10.1136/ard.2007.075838

---

Updated information and services can be found at:  
<http://ard.bmj.com/cgi/content/full/67/1/124>

---

*These include:*

### References

This article cites 13 articles, 3 of which can be accessed free at:  
<http://ard.bmj.com/cgi/content/full/67/1/124#BIBL>

### Rapid responses

You can respond to this article at:  
<http://ard.bmj.com/cgi/eletter-submit/67/1/124>

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

---

### Notes

---

To order reprints of this article go to:  
<http://journals.bmj.com/cgi/reprintform>

To subscribe to *Annals of the Rheumatic Diseases* go to:  
<http://journals.bmj.com/subscriptions/>

# The additive effect of individual genes in predicting risk of knee osteoarthritis

A M Valdes,<sup>1</sup> M Doherty,<sup>2</sup> T D Spector<sup>1</sup>

<sup>1</sup> St. Thomas' Hospital Campus, Kings College London School of Medicine London, UK;  
<sup>2</sup> Academic Rheumatology Unit, University of Nottingham, City Hospital Nottingham, UK

Correspondence to:  
Ana M Valdes, Twin Research Unit, St. Thomas' Hospital Campus, Kings College London, Lambeth Palace Road, London SE1 7EH, UK

Accepted 12 August 2007  
Published Online First  
16 August 2007

## ABSTRACT

**Objective:** Genetic factors are important determinants of osteoarthritis (OA) but most individual genetic associations appear relatively modest. We aimed to answer whether carrying several genetic variants associated with knee OA could result in a greater risk of OA

**Methods:** Genotypes for 36 single nucleotide polymorphisms (SNPs) in 17 candidate genes previously associated with OA were analysed in 298 men and 305 women diagnosed with knee OA who met American College of Rheumatology (ACR) criteria, and in 297 male and 299 female age- and ethnicity-matched controls. The S-sum statistic method was used to select SNPs that contributed to knee OA, separately for men and women, and the coefficients from a logistic regression were used to add the genotypes in a new genetic risk variable.

**Results:** The odds ratio for individuals in the top quartile of the "genetic risk" variable compared to those in the bottom quartile was found to be 8.68 (95% CI 5.20–14.49,  $p < 2 \times 10^{-16}$ ) for women and 5.06 (95% CI 3.10–8.27,  $p < 1 \times 10^{-10}$ ) for men.

**Conclusions:** Our data suggest that the additive information from a number of genetic variants can predict a substantial proportion of risk of knee OA.

Osteoarthritis (OA) of the knee is a common complex disorder resulting in joint disability with known constitutional and environmental risk factors for development and progression, such as age, obesity, hormonal status, bone density, physical activity and past history of trauma.<sup>1</sup> Knee OA, also has an important genetic component,<sup>2</sup> and several studies have investigated the role of candidate genes in the risk of hip and knee OA. Several genes with common polymorphisms consistently affecting risk of OA have been reported to date.<sup>3–8</sup> However, the genetic variants involved are not mutants with large attributable risks. Rather, the increased risks for carrying a predisposing genetic variant appear to be fairly modest, with most of them having odds ratios between 1.3 and 2.0. If an individual carries risk variants at several genes, does his/her risk of OA increase in proportion, or do genetic risks remain modest when compared with environmental/constitutional risk factors such as obesity and past history of trauma?

To answer this question we have combined the previously published genotypes of 36 variants on 17 genes in a multi-centre case-control study of knee OA. Variation at all the genes included in this study have been reported to be associated with risk of knee or hip OA in at least one independent study.<sup>5–6</sup> We assessed the risk of OA for individuals carrying several gene variants associated with disease susceptibility compared to individuals

who carried few or none. Because several of the genes included (eg, FRZB, BMP2)<sup>5–6</sup> have been reported to be associated only or predominantly in women, and the prevalence of OA is higher in females, all analyses have been performed separately in each gender.

## SUBJECTS AND METHODS

Subjects for this study have been described elsewhere.<sup>6</sup> Briefly, 603 knee OA cases (298 men and 305 women) were recruited in Nottingham, UK. Osteoarthritis was assessed both clinically and radiographically, patients had standardised extended weight-bearing anteroposterior radiographs and met American College of Rheumatology (ACR) criteria.<sup>10</sup> In addition, 596 Caucasian age-matched controls aged 50–80 without clinical signs or symptoms typical of osteoarthritis were recruited from two centres: Nottingham (111 women and 50 men) and Oxford (185 women and 250 men). No radiographs were performed on controls. Controls were only characterised by clinical criteria. The mean age was: female cases, 73.5 years (SD = 7.16); female controls, 72.1 (SD = 8.5); male cases, 72.1 (SD = 6.9); and male controls, 71.0 years (SD = 7.8). All study subjects gave informed consent to participate and the Oxford and Nottingham Research Ethics Committees approved the protocol.

## Genotypes

The genotyping data included in this study have been previously reported. For each SNP we identified the allele that was more common among cases than among controls and assigned a value of 1 to the homozygote carrying two copies of such allele, 0.5 to the heterozygote and 0 for the lower risk homozygote. The genes and SNPs included are shown in table 1.

## Set association

This method is used to find a set of SNPs that are jointly associated with disease by computing an association statistic for each SNP ie, a  $\chi^2$  for a 2×3 table, where the two rows correspond to cases and controls, and the three columns refer to SNP genotypes. Markers are then ordered by the size of their test statistics, and sums are formed sequentially, starting with the largest test statistic and gradually adding one after another SNP and sums,  $S_i$ , are formed sequentially, starting with the largest test statistic and gradually adding one after another SNP.  $S_i$  is the sum of the  $i$  largest test statistics for the SNPs. For each  $S_i$ , an associated significance level is computed with permutation

**Table 1** Genes and single nucleotide polymorphisms (SNPs) used to assess the combined risk of knee osteoarthritis (OA)

Gene symbol	Gene name	SNP alias	SNP rs no.	Chromosome	Position	Higher risk/ lower risk allele	Freq of higher risk allele in controls (%)
AACT	Alpha1 antiproteinase antitrypsin	aact	rs4934	14	941 505 556	(G/A)	50.9
ADAM12	A disintegrin and metalloproteinase domain 12 (meltrin alpha)	adam_48	rs3740199	10	128 009 015	(G/C)	47.6
		adam_504	rs1278279	10	127 743 468	(A/G)	21.6
		adam_825	rs1044122	10	127 714 768	(C/T)	73.1
ASPN	Asporin	adam_int	rs1871054	10	127 772 399	(A/G)	43.9
		aspn_5p	rs7022562	9	92 295 895	(C/T)	69.7
		aspn_int1	rs7033979	9	92 303 535	(C/T)	73.2
		aspn_int2	rs13301537	9	92 308 602	(T/C)	26.2
		aspn_3p1	rs3739606	9	92 316 777	(C/A)	30.4
		aspn_3p2	rs331377	9	92 323 104	(C/T)	50.1
BMP2	Bone morphogenetic protein 2	bmp2_87	rs1049007	20	6 699 034	(C/T)	35.7
		bmp2_190	rs235768	20	6 707 115	(A/T)	37.5
CALM1	calmodulin 1	calm1_int	rs3213718	14	89 939 666	(T/C)	60.6
CD36	CD36 antigen (collagen type I receptor, thrombospondin receptor)	cd36_5p	rs1049654	7	80 113 391	(C/A)	54.9
		cd36_int	rs3211822	7	80 116 562	(G/A)	59.4
CILP	Cartilage intermediate layer protein, nucleotide pyrophosphohydrolase	cilp_3p	rs1561888	15	63 289 664	(T/C)	29.4
		cilp_395	rs2073711	15	63 281 265	(A/G)	57.9
COL2A1	Collagen, type II, alpha 1	col2_int	rs1635560	12	46 654 096	(C/T)	74.5
		col2_1405	rs2070739	12	46 654 243	(T/C)	90.3
COX2	Prostaglandin-endoperoxide synthase 2	cox2_102	rs5277	1	184 914 820	(C/G)	18.4
		cox2_3p	rs689470	1	184 907 681	(A/G)	96.3
ESR1	Oestrogen receptor alpha	esr_325	rs1801132	6	152 307 215	(C/G)	78.4
		esr_594	rs2228480	6	152 461 788	(A/G)	16.8
		esr_int1	rs2234693	6	152 205 028	(C/T)	53.9
		esr_int2	rs827421	6	152 198 815	(A/G)	49.6
		frzb_200	rs288326	2	183 528 842	(C/G)	10.9
FRZB	Frizzled-related protein	frzb_324	rs7775	2	183 525 090	(T/C)	7.3
		lrch1_int	rs912428	13	46 065 904	(T/C)	20.4
LRCH1	Leucine-rich repeats and calponin homology (CH) domain containing						
NCOR2	Nuclear receptor co-repressor 2	ncor_1699	rs2229840	12	123 392 415	(G/A)	82.0
OPG	Tumour necrosis factor receptor superfamily, member 11b (osteoprotegerin)	opg_5p	rs1564858	8	120 014 347	(A/G)	11.7
		opg_3	rs2073618	8	120 033 233	(C/G)	55.1
TNA	Tetranectin (plasminogen binding protein)	tna_106	rs13963	3	45 052 127	(A/G)	31.1
		tna_int	rs939309	3	45 047 587	(C/T)	41.9
TNFAIP6	Tumour necrosis factor, alpha-induced protein	tnfaip	rs1046668	2	151 934 816	(G/A)	87.9
VDR1	Vitamin D receptor	vdr_1	rs10735810	12	46 559 162	(A/G)	63.7
		vdr_365	rs731236	12	46 525 024	(C/T)	37.4

The control frequency refers to the combined male and female population of the present study.

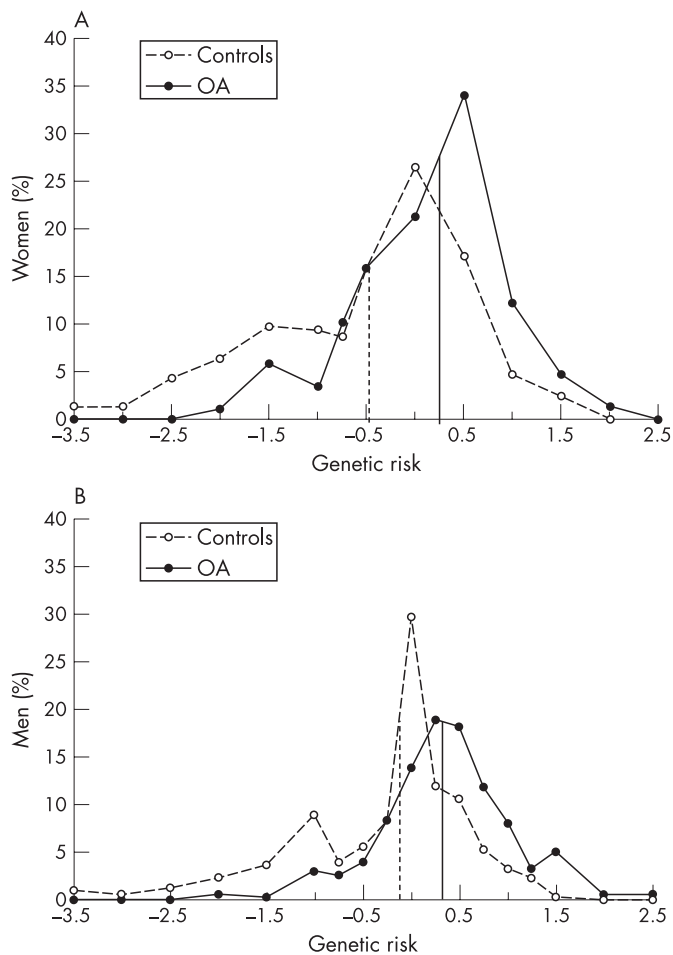
testing (10 000 replicates). Only those SNPs that are significantly associated as part of a set (having taken into account any lack of independence between markers) result in a p value <0.05 after correcting for multiple testing. The S statistic software available at <http://www.genemapping.cn/sumstat.html> was used.<sup>9</sup> This method has been shown to have consistently higher statistical power than individual marker associations under a range of simulated models.<sup>11</sup> Although set association can incorporate multiplicative effects, no interactions have been

reported in the literature for these genes and no such exploratory analyses were undertaken.

### Logistic regression

In order to combine all the significant SNPs resulting from the set association analysis into a single variable (not part of the set association method), we then fitted a logistic regression model. As not all significant markers are expected to contribute equally to disease risk, in order not to dilute the information on those

## Concise report



**Figure 1** Frequency distributions of the genetic risk variable in cases and control in men (A) and women (B). The median of each distribution is indicated by a line.

genes that have a larger influence on risk of OA, the coefficients from the logistic regression were used as weights for the significant SNPs to generate a new “genetic risk” variable. The distribution was standardised to have mean = 0 and SD = 1. Odds ratios and 95% CI were derived from the logistic regression results. The genetic risk distribution between cases and controls was also compared using a Wilcoxon-rank test with a continuity correction. Analyses were carried out in S-Plus 6.0 (Insightful Corp, Seattle, Washington, USA).

## RESULTS

The set association method was used to identify the SNPs significantly contributing to knee OA risk while adjusting for multiple testing.<sup>9</sup> The set of markers were then included in a logistic regression and the coefficients from the model fitted to assign a weight to each genotype.

The final models fitted, standardised to have a mean of 0 and a variance of 1, were:

$$\text{Genetic risk (women)} = (1/0.9829) \times ((0.3898 \times \text{adam\_int}) + (0.5113 \times \text{adam\_504}) + (0.4825 \times \text{aact}) + (1.6926 \times \text{cilp\_3p}) + (1.4256 \times \text{tna\_106}) + (1.2562 \times \text{frzb\_200}) + (0.7132 \times \text{frzb\_324}) + (1.7504 \times \text{cilp\_395}) + (1.0828 \times \text{aspn\_5p}) + (1.3776 \times \text{aspn\_3p1}) + (0.9611 \times \text{bmp2\_87}) + (0.8695 \times \text{esr\_594}) - 4.59064).$$

$$\text{Genetic risk (men)} = (1/0.8338) \times ((1.9538 \times \text{cilp\_3p}) + (0.5236 \times \text{adam\_int}) + (0.3414 \times \text{vdr\_365}) + (0.3894 \times \text{esr1\_325})$$

$$+ (2.8437 \times \text{cilp\_395}) + (0.7996 \times \text{col2\_int}) + (0.7187 \times \text{lrch1}) + (0.4273 \times \text{aspn\_5p}) - 3.5814).$$

Where *adam\_int*, *adam\_504* etc denote the genotype (0, 0.5, 1) of the SNP listed in table 1 coded as described under the methods section.

Figure 1 shows the frequency distribution of the genetic risk. The odds ratios between individuals in the top and bottom half of the distribution was 3.32 (95% CI 2.38–4.64,  $p < 1 \times 10^{-17}$ ) in women and 3.12 (95% CI 2.24–4.36,  $p < 1 \times 10^{-10}$ ) in men. A Wilcoxon-rank sum test revealed that the genetic risk distribution was significantly different between cases and controls in females with  $p < 1 \times 10^{-16}$  and in males with  $p < 1 \times 10^{-13}$ .

When the top and bottom quartiles were used the odds ratios became 8.68 (95% CI 5.20–14.49,  $p < 2 \times 10^{-16}$ ) for women and 5.06,  $2 \times 10^{-16}$  (95% CI 3.10–8.27,  $p < 1 \times 10^{-10}$ ) for men.

## DISCUSSION

The data shown here indicate that it is possible to identify individuals at high risk of knee OA using genotype data. Although family and epidemiological studies have consistently indicated an important genetic contribution to OA, to date no single large genetic effect has been found. From a biological point of view these results confirm that the genetic risk to knee OA is likely to be due to the sum of many loci making a small contribution each. Consistent with the gender differences previously reported<sup>5</sup> these data suggest that by using gender-specific genetic factors it may be possible to predict a high risk of knee OA both in men and women.

All the genes included in this model have been reported to be significantly associated with clinical or radiographic features of knee or hip OA in at least one independent study. The odds ratios obtained using the genetic risk variable are comparable than those reported for obesity or knee injury by some studies.<sup>12–14</sup> For example the odds ratios for obesity/BMI as a risk factor range from 3.0<sup>12 13</sup> to 18.0<sup>12 14</sup> and the odds ratios for knee injury range from 3.0 for bilateral disease<sup>12</sup> or 4.8<sup>14</sup> to 16.0<sup>12</sup> for unilateral disease.

There are some limitations to the present study. First, the genes and all the model parameters were derived from within the same population. Therefore, although all the genes have been independently associated with OA before, we cannot assess the value of these data in predicting knee OA risk without testing this model first in an independent sample. In addition, there are other candidate genes that have been consistently associated with OA that were not genotyped as part of this study (eg, interleukin genes)<sup>7 8</sup> and that could result in improved prediction.

If these data are replicated in independent studies, it should be possible to indeed use genetic information for risk assessment. An important insight from this study is that the inclusion of more genes as genome-wide association scans and new candidates become available is likely to improve prediction of risk of knee OA to a level where it can be clinically useful.

**Acknowledgements:** The authors wish to thank the two anonymous referees for their valuable comments on an earlier version of this manuscript

**Funding:** This work was supported by the Arthritis Research Campaign and by the Wellcome Trust

**Competing interests:** None declared

## REFERENCES

1. Felson DT. An update on the pathogenesis and epidemiology of osteoarthritis. *Radiol Clin North Am* 2004;**42**:1–9.

2. **Valdes AM**, Hart DJ, Jones KA, Surdulescu G, Swarbrick P, Doyle DV, *et al.* Association study of candidate genes for the prevalence and progression of knee osteoarthritis. *Arthritis Rheum* 2004;**50**:2497–507.
3. **Lane NE**, Lian K, Nevitt MC, Zmuda JM, Lui L, Li J, *et al.* Frizzled-related protein variants are risk factors for hip osteoarthritis. *Arthritis Rheum* 2006;**54**:1246–54.
4. **Jiang Q**, Shi D, Yi L, Ikegawa S, Wang Y, Nakamura T, *et al.* Replication of the association of the aspartic acid repeat polymorphism in the asporin gene with knee-osteoarthritis susceptibility in Han Chinese. *J Hum Genet* 2006;**51**:1068–72.
5. **Valdes AM**, Loughlin J, Oene MV, Chapman K, Surdulescu GL, Doherty M, *et al.* Sex and ethnic differences in the association of ASPN, CALM1, COL2A1, COMP, and FRZB with genetic susceptibility to osteoarthritis of the knee. *Arthritis Rheum* 2007;**56**:137–46.
6. **Valdes AM**, Van Oene M, Hart DJ, Surdulescu GL, Loughlin J, Doherty M, *et al.* Reproducible genetic associations between candidate genes and clinical knee osteoarthritis in men and women. *Arthritis Rheum* 2006;**54**:533–9.
7. **Smith AJ**, Keen LJ, Billingham MJ, Perry MJ, Elson CJ, Kirwan JR, *et al.* Extended haplotypes and linkage disequilibrium in the IL1R1-IL1A-IL1B-IL1RN gene cluster: association with knee osteoarthritis. *Genes Immun* 2004;**5**:451–60.
8. **Miyamoto Y**, Mabuchi A, Shi D, Kubo T, Takatori Y, Saito S, *et al.* A functional polymorphism in the 5'UTR of GDF 5 is associated with susceptibility to osteoarthritis. *Nat Genet* 2007;**39**:529–33.
9. **Hoh J**, Wille A, Ott J. Trimming, weighting, and grouping SNPs in human case-control association studies. *Genome Res* 2001;**11**:2115–9.
10. **Altman RD**, Block DA, Brandt KD, Cooke DV, Greenwald RA, Hochberg MC, *et al.* Osteoarthritis: definitions and criteria. *Ann Rheum Dis* 1990;**49**:201.
11. **Kim S**, Zhang K, Sun F. Detecting susceptibility genes in case-control studies using set association. *BMC Genet* 2003 Dec 31;**4**(Suppl 1):S9–14.
12. **Davis MA**, Ettinger WH, Neuhaus JM, Cho SA, Hauck WW. The association of knee injury and obesity with unilateral and bilateral osteoarthritis of the knee. *Am J Epidemiol* 1989;**130**:278–88.
13. **Oliveria SA**, Felson DT, Cirillo PA, Reed JI, Walker AM. Body weight, body mass index, and incident symptomatic osteoarthritis of the hand, hip, and knee. *Epidemiology* 1999;**10**:161–6.
14. **Cooper C**, Snow S, McAlindon TE, Kellingray S, Stuart B, Coggon D, *et al.* Risk factors for the incidence and progression of radiographic knee osteoarthritis. *Arthritis Rheum* 2000;**43**:995–1000.