



OPEN ACCESS

EXTENDED REPORT

# Novel genetic variants associated with lumbar disc degeneration in northern Europeans: a meta-analysis of 4600 subjects

Frances M K Williams,<sup>1</sup> Aruna T Bansal,<sup>2</sup> Joyce B van Meurs,<sup>3,4</sup> Jordana T Bell,<sup>1</sup> Ingrid Meulenbelt,<sup>4,5</sup> Pradeep Suri,<sup>6</sup> Fernando Rivadeneira,<sup>3,4</sup> Philip N Sambrook,<sup>7</sup> Albert Hofman,<sup>8</sup> Sita Bierma-Zeinstra,<sup>9</sup> Cristina Menni,<sup>1</sup> Margreet Kloppenburg,<sup>10</sup> P Eline Slagboom,<sup>4,5</sup> David J Hunter,<sup>7</sup> Alex J MacGregor,<sup>1</sup> Andre G Uitterlinden,<sup>3</sup> Tim D Spector<sup>1</sup>

► Additional data are published online only. To view this file please visit the journal online (<http://dx.doi.org/10.1136/annrheumdis-2012-201551>).

For numbered affiliations see end of article.

## Correspondence to

Dr Frances M K Williams, Department Twin Research and Genetic Epidemiology, King's College London, St Thomas Hospital, Westminster Bridge Road, London SE1 7EH, UK; frances.williams@kcl.ac.uk

ATB and JBM contributed equally.

Accepted 19 July 2012

## ABSTRACT

**Objective** Lumbar disc degeneration (LDD) is an important cause of low back pain, which is a common and costly problem. LDD is characterised by disc space narrowing and osteophyte growth at the circumference of the disc. To date, the agnostic search of the genome by genome-wide association (GWA) to identify common variants associated with LDD has not been fruitful. This study is the first GWA meta-analysis of LDD.

**Methods** We have developed a continuous trait based on disc space narrowing and osteophytes growth which is measurable on all forms of imaging (plain radiograph, CT scan and MRI) and performed a meta-analysis of five cohorts of Northern European extraction each having GWA data imputed to HapMap V2.

**Results** This study of 4600 individuals identified four single nucleotide polymorphisms with  $p < 5 \times 10^{-8}$ , the threshold set for genome-wide significance. We identified a variant in the PARK2 gene ( $p = 2.8 \times 10^{-8}$ ) associated with LDD. Differential methylation at one CpG island of the PARK2 promoter was observed in a small subset of subjects ( $\beta = 8.74 \times 10^{-4}$ ,  $p = 0.006$ ).

**Conclusions** LDD accounts for a considerable proportion of low back pain and the pathogenesis of LDD is poorly understood. This work provides evidence of association of the PARK2 gene and suggests that methylation of the PARK2 promoter may influence degeneration of the intervertebral disc. This gene has not previously been considered a candidate in LDD and further functional work is needed on this hitherto unsuspected pathway.

## INTRODUCTION

Lumbar disc degeneration (LDD) is a common, age-related trait:<sup>1</sup> over a third of middle aged women have at least one degenerate disc. LDD contributes to low back pain<sup>2-3</sup> and as low back pain is common in the general population and costly to society,<sup>4</sup> LDD is of considerable public health importance. Discrete biochemical, histological, metabolic and functional changes occur in LDD, such that discs become dehydrated, lose disc height and there is accompanying outgrowth of osteophytes from the vertebral body margin.<sup>5</sup> There are similarities with peripheral joint

osteoarthritis (OA). LDD has been shown to be heritable, with estimates of 65%–80%<sup>6-7</sup> and so a considerable proportion of the variance in LDD is explained by genetic factors. Yet to date, candidate gene studies have detected only a small number of convincing associations of genetic variants with LDD (reviewed by Ryder *et al*<sup>8</sup>). A number of studies show conflicting results: these are likely due to small sample size or may reflect ethnic differences between Northern European and Asian populations, as seen in OA.<sup>9</sup> That some published genome-wide associations (GWAs) in common complex traits fail to replicate candidate gene findings suggests limitations to the candidate gene method.<sup>10</sup> As in other common complex traits,<sup>11</sup> a considerable proportion of the genetic variance in LDD remains unexplained. While GWA studies do not capture all variation in the genome, the approach does offer an agnostic search of the genome for variants associated with common complex traits. Their main limitation results from the inherent multiple testing in their design, meaning that power is lost and large samples are needed to address this. In order to optimise sample size in the present study we performed meta-analysis of GWAS using a number of cohorts having the LDD phenotype.

## PATIENTS AND METHODS

The cohorts available for inclusion in this study were all population samples, except Genetics of osteoArthritis and Progression study (GARP) which specifically recruited participants having a diagnosis of OA. A variable was derived from measures of disc height and osteophytes obtained from lateral images on MR, CT scan or plain radiograph. Summing this variable over the lumbar discs provided a continuous measure of disc degeneration. GWA of this summary variable was performed by each individual study group and summary statistics were sent to and collated by KCL. Meta-analysis was performed of imputed GWA data from five population cohorts (Framingham, GARP, Rotterdam study 1 and 3 and TwinsUK (TUK)) having imaging of the spine (see below). All cohorts had obtained fully informed consent

from their participants and appropriate ethics committee approval. In all studies, a cumulative degeneration score was constructed from the sum of scores of degenerative change at each level (disc space narrowing coded 0–3 and osteophytes, either anterior or posterior or both, coded 0–3). In those cohorts where only four disc levels were read (Framingham Heart study (FHS)), a fifth level was imputed by taking the mean reading for four discs as a surrogate for the fifth disc, and summing over five discs. The data underwent inverse normal transformation to generate a normally distributed variable.

### Phenotyping the cohorts

*The FHS* is a longitudinal cohort of a defined population in Massachusetts, initiated in 1948 (<http://www.framinghamheartstudy.org>). It began as a study sample of 5209 Framingham men and women between the ages of 30 and 60. Subsequently, offspring and third generation subjects were incorporated. Every other year, after an extensive baseline examination, subjects undergo testing that includes a medical history, blood profile, echocardiogram, and bone, eye and other tests. The subset of the Framingham subjects covered by the current analysis comprised 366 subjects from the Offspring and Generation three arms of the study who had undergone CT scanning of the spine, and the recruitment, conduct and specifications of CT scanning have been reported elsewhere.<sup>12</sup> Measurement of the lumbar spine CTs for disc height and scoring (0–3) for anterior and posterior osteophytes was performed by a spine specialist using the mid-sagittal plane at spinal levels L2-L3, L3-L4, L4-L5 and L5-S1 by author PS using the atlas of Jarosz *et al.*<sup>7</sup> Using sagittal CT reformatting, the mid-sagittal plane was determined at each spinal level and measurements of disc height in millimetres were made. The measured values for disc height (mm) were converted to 0–3 categorical scale for disc height loss. Using an imputed value for the fifth lumbar disc based on the mean value of the measured four discs, values for disc height loss and anterior and posterior osteophytes were summed over five lumbar disc levels.

*The GARP study* comprises white sibling pairs of Dutch origin affected by OA at multiple sites and is aimed at identifying determinants of OA susceptibility and progression. Proband (ages 40–70 years) and their siblings had OA at multiple joint sites of the hand or in two or more of the following joint sites: hand, spine (cervical or lumbar), knee or hip as described previously.<sup>13</sup> Subjects included in this study had undergone lateral radiographs of the spine (T4-S1). Each intervertebral disc level from L1/2 to L5/S1 was reviewed for the presence and severity of osteophytes (anterior) and disc narrowing, using the Lane atlas<sup>14</sup> where 0=none; grade 1=mild; grade 2=moderate; and grade 3=severe. The score at each level for anterior osteophytes and disc height loss were summed over the five lumbar levels.

*The Rotterdam study* is a prospective population-based follow-up study of the determinants and prognosis of chronic diseases in the elderly.<sup>15–16</sup> All persons living in Ommoord, a suburb of Rotterdam, who were aged 55 years and over were invited to participate. A total of 7983 participants were examined. For the current analysis, two subsets of the data were considered. Rotterdam cohort 1 (RS1) consists of 2440 subjects; Rotterdam cohort 3 (RS3) consists of 974 subjects. Subjects originating from the Rotterdam study underwent plain radiography and scoring of LDD as previously described.<sup>2</sup> In brief, lateral lumbar radiographs were scored by a single observer for the presence of the individual radiographic features of disc

degeneration. Each intervertebral disc from L1/2 to L5/S1 was reviewed for the presence and severity of osteophytes (anterior) and disc narrowing using the Lane atlas as described above.<sup>14</sup> The scores for the two traits over the five lumbar discs were summed.

*The TUK registry* was described previously.<sup>17</sup> The register was started in 1993 and now comprises of approximately 10 000 monozygotic and dizygotic adult Caucasian twins aged 16–85 years from all over the UK, plus some parents and siblings. It now incorporates previous twin registries from the Institute of Psychiatry and Aberdeen University. This is a volunteer sample recruited by successive media campaigns without selecting for particular diseases or traits. All twins receive a series of detailed disease and environment questionnaires. The majority of twins have been assessed in detail clinically at several time points for several hundred phenotypes related to common diseases or intermediate traits. The subset of TUK covered by the current analysis consisted of 744 subjects who had participated in the spine MR study (scanned 1996–2000) using a Siemens MR machine with (Munich, Germany) 1.0-tesla superconducting magnet. Serial sagittal images of the cervical, thoraco-lumbar junction and lumbar spine (T9-L5) were obtained.<sup>7</sup> Images were coded for disc height loss and anterior osteophytes using a 0–3 scale in each case, where 0 is normal and 3 maximal degeneration as per the atlas of Jarosz *et al.*<sup>7</sup> All five lumbar discs were scored and the scores summed to give a combined LDD variable.

### Genotyping and imputation

*FHS* subjects were genotyped using Affymetrix GeneChip Human Mapping 500 K array set (Affymetrix, Santa Clara, CA, USA) and/or the 100 K array set and/or the 50 K array. Methods and quality controls have been described previously.<sup>18</sup>

*GARP subjects* were genotyped using Illumina Human660W Quad BeadChips (HumanHap550v3, HumanHap610; Illumina, San Diego, CA, USA). Genotyping was performed at the genotyping Rotterdam Genotyping Centre. Positive strand genotypes were called by clustering in Genome studio and imputation was performed using IMPUTE software and hapmap phase II V21.<sup>19–20</sup> Strict selection criteria were applied to the measured genotypes using a high information content ( $r^2$  of >95%) and a minor allele frequency >0.0025. Association analyses were performed using an inhouse developed software package that allows the analyses of family data using all information available in the cases and controls by extending the Cochran-Armitage trend test.<sup>21</sup>

*RS1 and RS3* subjects in the Rotterdam Study sets were genotyped on the HumanHap550v3 (RS1) or HumanHap610 (RS3) Genotyping BeadChip (Illumina, San Diego, California, USA). The following sample quality control criteria were applied: sample call rate >97.5%, gender mismatch with typed X-linked markers, evidence for DNA contamination in the samples using the mean of the autosomal heterozygosity >0.33, exclusion of duplicates or first-degree relatives identified using Identity by State probabilities and exclusion of outliers (four SD away from the population mean using multidimensional scaling analysis with four principal components). Filtering criteria for imputation are summarised in supplementary table S1.

*TUK* subjects were genotyped using a combination of Illumina arrays (Human Hap300 and the Human Hap610Q). Genotyping was performed by the Wellcome Trust Sanger Institute using the Infinium assay (Illumina) across three genome-wide single nucleotide polymorphism (SNP) sets, as described previously.<sup>22</sup> Genotyping results had been sent to

KCL for collation and analysis using the statistical package, STATA (StataCorp).<sup>23</sup> Strict quality control was applied: 314 075 SNPs were retained for analysis (98.7%); 733 were excluded because their call rates were  $\leq 90\%$  and 725 SNPs had minor allele frequency  $< 0.01$ . In TUK, significant population substructure was excluded using the STRUCTURE program.<sup>24</sup>

### GWA analysis

All analyses were performed on inverse normal transformed summary LDD score as described above. Each study performed GWA analysis for LDD scores using either MACH2QTL (<http://www.sph.umich.edu/csg/abecasis/MACH/index.html>) (RS1 and RS3) or SNPTTEST (<http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html>) (GARP), which use genotype dosage value as continuous additive predictors of LDD score in a linear regression framework, or ProbABEL<sup>25</sup> using an additive genetic model while accounting for relatedness between the members of a family. Analysis of imputed genotype data accounted for uncertainty in each genotype prediction by using either the dosage information from MACH<sup>26</sup> or the genotype probabilities from IMPUTE.<sup>19</sup>

### Meta-analysis of the five study groups

Genotypes for 2.5–3 million autosomal SNPs were imputed separately to increase coverage using HapMap V.2 (<http://www.hapmap.org>) as the reference panel. In GARP and TUK, imputation was performed with IMPUTE V.2<sup>19</sup> and in the other studies with MACH.<sup>26</sup> The common reference panel led to the reporting of results for the positive strand for all cohorts. In addition, allele pairs were compared between cohorts and no detectable strand-flips were found; the minor allele frequency was also compared between datasets. The distributions of  $\beta$  values of the cohorts were found to be similar and therefore suitable for meta-analysis. All directly genotyped or imputed autosomal SNPs having information from more than one study group ( $n=2\,552\,511$ ) were included in the meta-analysis. Association results were combined using inverse variance weighted fixed effects meta-analysis using PLINK V.1.06 (<http://pngu.mgh.harvard.edu/purcell/plink/>). Two meta-analyses were run: the first was unadjusted; the second was adjusted for age and sex as both are known risk factors for LDD and each risk factor was correlated with LDD in each study group. Heterogeneity of estimated effect was expressed using  $Q$  (weighted sum of squares) and  $I^2$  (ratio of true heterogeneity to total observed variation). SNPs were excluded from the meta-analysis if the cohort-specific imputation quality, as assessed by  $r^2$  (MACH) or Information Score (IMPUTE) metric, was  $< 0.40$ . On this basis, one marker was excluded from the unadjusted association and one from the adjusted association.

### DNA methylation data and analysis

Whole blood DNA methylation levels were obtained for 38 individuals in the TUK cohort using the Illumina HumanMethylation27 DNA Analysis BeadChip assay, as previously described.<sup>27</sup> The sample included four monozygotic twin pairs, eight dizygotic twin pairs and 14 unrelated individuals. At each CpG site within an individual the methylation level was presented as  $\beta$ , which represents the ratio of intensity signal obtained from the methylated beads in the array over the sum of methylated and unmethylated bead signals. Following quality control checks, we obtained DNA methylation at three CpG sites in the promoter region of the *PARK2* gene within 2kb of the transcription start site. The three probes (cg15832436, cg21926612 and cg24816866) mapped uniquely to the human genome (hg18) within two mismatches (see Bell *et al*).<sup>28</sup> We fitted linear mixed effects models to assess association between DNA methylation levels at the three CpG sites in the *PARK2* promoter and LDD. We regressed the raw methylation levels on fixed-effect terms including methylation chip and LDD, and random-effect terms denoting family structure and zygosity, and compared the association of differentially methylated regions with a null model, which excluded LDD from the fixed-effects terms. We also repeated the association analyses by normalising the methylation values at each CpG site to  $N(0, 1)$ .

### RESULTS

The study samples for the meta-analysis included 4683 individuals of European ancestries. Table 1 shows sample size, demographic characteristics, LDD and lumbar spine imaging method for each independent cohort. The majority of participants were female subjects (67.0%) and the samples had a mean age of 57.7 years. Across the cohorts, the mean level of LDD varied from 0.011 to 3.46, reflecting differences in imaging methods. However, the variance of the LDD variables was broadly similar (range 0.958–1.14), as were the distributions of the estimated genetic effect sizes ( $\beta$ ). The genotyping and imputation methods are shown in online supplementary table S1.

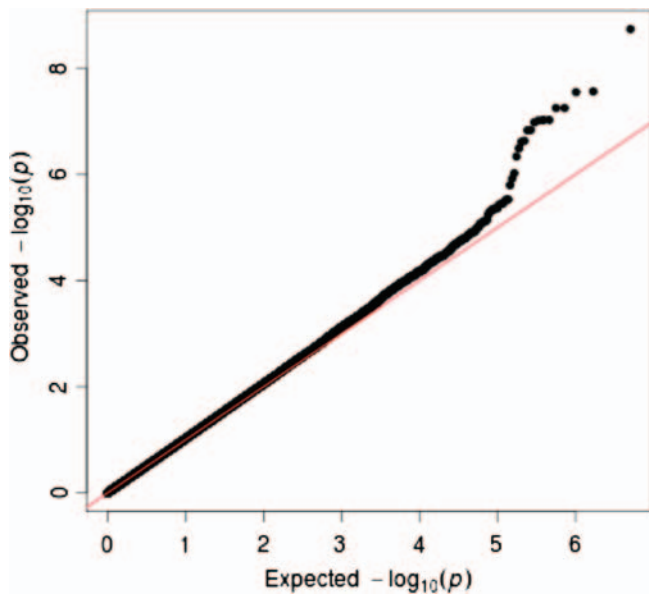
Quantile–quantile plots for the unadjusted LDD GWA meta-analysis are presented in figure 1 (see online supplementary figure S1, adjusted). Test statistic inflation post meta-analysis, as measured by the genomic control statistic,<sup>29</sup> was low ( $\lambda_{GC}$  unadjusted=1.02;  $\lambda_{GC}$  adjusted=1.03). Results of the unadjusted and adjusted association analyses were broadly similar, with the  $p$  values of the adjusted analysis somewhat attenuated. A Manhattan plot for the unadjusted analysis is shown in figure 2 with numeric results in table 2 (unadjusted) and online supplementary table S2 (adjusted) for SNPs having  $p < 10^{-5}$ .

**Table 1** Characteristics of the study samples

	FHS	GARP	RS1	RS3	TUK
N	330	192	2440	974	744
Age (years)	54.3 (11.0)	60.3 (7.1)	65.7 (6.7)	54.7 (3.4)	53.6 (8.3)
Women (%)	42.2	79.7	57	59	96.8
BMI (kg/m <sup>2</sup> )	28.1 (5.1)	26.44 (4.8)	26.3 (3.4)	27.12 (4.6)	24.9 (4.4)
Lumbar spine imaging	CT	Radiograph	Radiograph	Radiograph	MRI
LDD variable	2.49 (0.97)	0.02 (0.958)	0.006 (0.978)	0.011 (0.965)	3.46 (1.14)

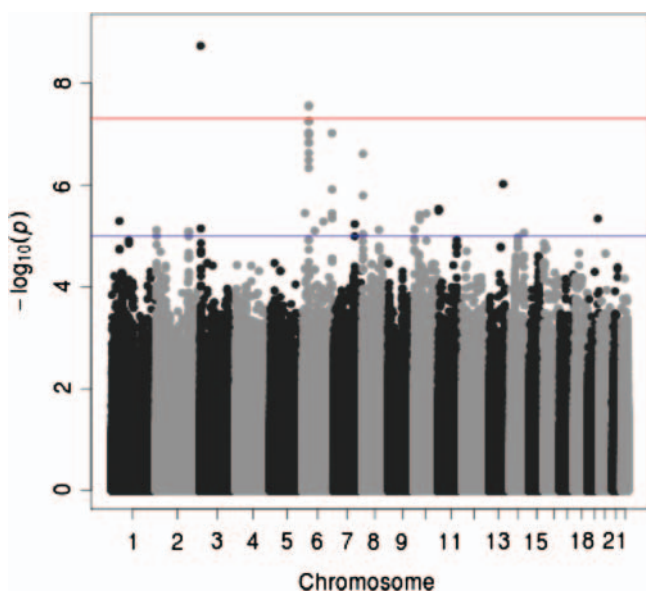
Values are mean (SD) unless specified otherwise.

BMI, body mass index; FHS, Framingham Heart Study; GARP, Genetics of OsteoArthritis and Progression study; LDD, Lumbar disc degeneration; RS1, Rotterdam study cohort 1; RS3, Rotterdam study cohort 3; TUK, TwinsUK.



**Figure 1** Quantile-quantile plot of observed versus expected distribution of p values for the genome-wide association (GWA) meta-analysis. The plots show GWA meta-analysis quantile-quantile plot of observed against expected results, unadjusted for covariates.

Four markers achieved genome-wide significance in the unadjusted GWAS, of which three were on chromosome 6 (rs926849; rs2187689; rs7767277) and an intergenic marker on chromosome 3 (rs17034687). The results of the meta-analysis adjusted for age and gender were broadly similar: the strongest signal was for SNP rs926849. This SNP lies on an intronic region of the Parkinson protein 2, E3 ubiquitin protein ligase (PARK2) gene on chromosome 6. A Forest plot of the groups' results and their meta-analysis is shown in figure 3. Data were



**Figure 2** Manhattan plot for meta-analysis of unadjusted genome-wide association results. Plot shows combined results for the five studies included in the meta-analysis, unadjusted results. The blue and red horizontal lines mark the levels of suggestive and likely significance, respectively.

available from four study groups and the range of estimated minor allele frequencies was 0.23–0.32. Imputation quality was high for all four studies contributing this SNP ( $>0.90$ , table 2) and the estimated between-study heterogeneity was acceptable ( $I^2=0\%$ ;  $p(Q)=0.67$ ). The results of a four-study meta-analysis which excluded GARP show the marker to have  $p=9.5e^{-8}$  (unadjusted). The minor or C allele of rs926849 was associated with a lower level of LDD implying that the minor allele is protective. Figure 4 shows a regional plot of both genotyped and imputed SNPs within 200 Kb of the PARK2 gene, along with recombination rates.

Two of the other strongly associated SNPs are in perfect linkage disequilibrium (LD): rs2187689 and rs7767277 on chromosome 6. Data were available for four studies and the range of estimated allele frequency was 0.06–0.09. Imputation quality was high for all four studies ( $>0.90$ ). Both SNPs are in strong LD ( $r^2=0.76$ ) with an intronic marker on the proteasome subunit,  $\beta$  type 9, large multifunctional peptidase 2 gene (PSMB9) that is located in the class II region of the major histocompatibility complex (MHC). Both genotyped and imputed SNPs within 400 Kb of rs2187689, along with recombination rates, are shown in a regional plot in online supplementary figure S3. None of these top SNPs is in LD with known functional SNPs in either PARK2 or PSMB9.

We tested for an association between LDD and DNA methylation variants at three CpG sites in the PARK2 promoter. A significant association between DNA methylation at CpG site cg15832436 and LDD ( $\beta=8.74 \times 10^{-4}$ ,  $SE=2.49 \times 10^{-4}$ ,  $p=0.006$ ) was observed. The pattern of hypermethylation with increasing LDD levels was reflected at the remaining promoter CpG sites; however, these did not reach nominal significance (cg21926612  $\beta=0.003$ ,  $p=0.14$ ; cg24816866  $\beta=6.76 \times 10^{-4}$ ,  $p=0.39$ ). We repeated the analyses using normalised methylation levels and observed that the association between cg15832436 and LDD remained nominally significant.

## DISCUSSION

GWA offers an unbiased scan of common genetic variants (minor allele frequency  $>5\%$ ) and thus may deliver novel variants in genes not hitherto suspected of playing a role in disc degeneration. This work is among the first to report on a genome-wide meta-analysis being conducted for LDD. LDD is an age-related process which occurs in all people to some extent and may be detected as early as the teenage years.<sup>1</sup> LDD is known to have genetic determinants<sup>7,16</sup> and its expression is also influenced by gender (women develop LDD later), body mass index<sup>30–34</sup> and smoking.<sup>35</sup> Occupational factors also play a small role in LDD.<sup>36,37</sup> LDD as determined by MRI has been implicated in the development of episodes of severe and disabling low back pain.<sup>3</sup> We undertook this large meta-analysis in order to identify novel genetic variants associated with LDD and to shed light on the underlying pathology of disc degeneration.

GWA data obtained using differing chip technology may be readily compared using imputation with HapMap. In total, 2 552 511 overlapping markers were available in each cohort. We identified four markers having significant association with the LDD phenotype ( $p < 5 \times 10^{-8}$ ). There was similarity in the results obtained with and without adjustment for the covariates age and sex. A total of 26 markers had  $p < 10^{-5}$  in both meta-analyses. As expected, results of the adjusted analyses had slightly attenuated p values (see online supplementary table S2) which likely reflect the confounding effect of age. In both analyses, there were multiple associations to the Human

**Table 2** Results of the genome-wide association meta-analysis (unadjusted, showing those SNPs having  $p < 10^{-5}$ ).

SNP	Chr	Gene	N	MAF					Imputation quality*					Eff all	$\beta$	SE	p Value
				RS1	RS3	TUK	FHS	GARP	RS1	RS3	TUK	FHS	GARP				
rs17034687	3	NA	4429	0.09	0.1	0.05	0.08	NA	0.92	0.93	0.88	0.82	NA	C	0.23	0.038	1.82E-09
rs2187689	6	NA	4457	0.08	0.07	0.06	0.09	NA	0.98	0.99	0.95	0.93	NA	C	0.23	0.041	2.72E-08
rs7767277	6	NA	4457	0.08	0.07	0.06	0.09	NA	0.98	0.99	0.95	0.93	NA	A	0.23	0.041	2.81E-08
rs926849	6	PARK2	3939	0.31	0.32	NA	0.31	0.23	0.98	0.99	NA	0.9	95.04	C	-0.13	0.024	3.25E-08
rs7744666	6	NA	4466	0.1	0.09	0.06	0.1	NA	0.99	1	0.97	0.96	NA	C	0.2	0.037	5.58E-08
rs11969002	6	NA	4466	0.1	0.09	0.06	0.1	NA	0.99	1	0.97	0.96	NA	A	0.2	0.037	5.59E-08
rs6457690	6	NA	4464	0.1	0.09	0.07	0.11	NA	0.98	1	0.96	0.97	NA	A	0.19	0.036	9.36E-08
rs1029296	6	NA	4464	0.1	0.09	0.07	0.11	NA	0.98	1	0.96	0.97	NA	C	0.19	0.036	9.39E-08
rs6936004	6	NA	4462	0.1	0.09	0.07	0.11	NA	0.98	1	0.96	0.97	NA	C	0.19	0.036	1.04E-07
rs3749982	6	NA	4458	0.1	0.09	0.06	0.1	NA	0.99	1	0.96	0.96	NA	A	0.19	0.037	1.46E-07
rs9469300	6	NA	4482	0.1	0.09	0.07	0.1	NA	0.99	1	0.92	0.96	NA	A	0.19	0.037	1.47E-07
rs10214886	6	NA	4479	0.1	0.09	0.07	0.11	NA	0.98	1	0.92	0.97	NA	A	0.19	0.036	2.32E-07
rs10046257	6	NA	4461	0.1	0.09	0.08	0.11	NA	0.98	1	0.96	0.97	NA	A	0.19	0.037	3.22E-07
rs4875102	8	NA	4608	0.26	0.26	0.27	0.25	0.29	0.99	0.99	0.95	0.91	97.65	A	-0.12	0.024	3.61E-07
rs3019449	6	PARK2	4636	0.32	0.32	0.31	0.31	0.28	0.98	0.98	0.98	0.99	97.18	A	-0.12	0.023	3.68E-07
rs1029295	6	NA	3747	0.1	0.09	NA	0.11	NA	0.98	1	NA	0.97	NA	C	0.19	0.038	4.58E-07
rs9301951	13	GPC6	4397	0.04	0.04	0.03	0.05	NA	0.96	0.97	0.74	0.78	NA	C	-0.26	0.054	9.47E-07
rs7896691	10	PFKP	4652	0.1	0.1	0.09	0.13	0.07	0.99	1	0.96	0.67	97.61	C	0.17	0.035	2.02E-06
rs6602024	10	PFKP	4673	0.1	0.1	0.09	0.13	0.07	0.99	1	0.98	0.67	97.62	A	0.17	0.035	2.20E-06
rs1884158	6	PARK2	4491	0.31	0.32	0.31	0.31	NA	0.98	0.99	0.99	0.91	NA	C	-0.11	0.023	3.56E-06
rs10998466	10	NA	4446	0.01	0.01	0.01	0.02	NA	0.75	0.97	0.73	0.51	NA	A	-0.53	0.114	3.59E-06
rs17132175	10	PFKP	4659	0.1	0.1	0.09	0.13	0.07	0.98	0.99	0.97	0.63	97.13	C	0.16	0.036	3.73E-06
rs1981483	16	PIGQ	3939	0.42	0.43	NA	0.42	0.47	0.99	0.98	NA	0.44	99.64	A	0.11	0.023	3.75E-06
rs1154053	8	NA	4639	0.17	0.17	0.18	0.16	0.2	0.99	1	0.96	0.89	96.17	C	-0.13	0.028	3.99E-06
rs737631	6	PARK2	4303	0.28	0.29	0.23	0.28	NA	0.92	0.91	0.89	0.83	NA	A	-0.11	0.025	4.32E-06
rs2484990	10	NA	4429	0.01	0.01	0	0.02	NA	0.56	0.82	0.47	0.93	NA	C	0.68	0.148	4.39E-06
rs1250307	10	NA	4415	0.01	0.01	0	0.02	NA	0.56	0.82	0.45	0.94	NA	A	0.68	0.148	4.39E-06
rs7204439	16	RAB40C	3939	0.42	0.43	NA	0.43	0.47	0.97	0.98	NA	0.43	97.37	C	0.11	0.023	4.43E-06
rs4802666	19	MYH14	4094	0.27	0.27	0.18	0.27	NA	0.7	0.74	0.76	0.64	NA	A	-0.13	0.029	4.55E-06
rs2484992	10	NA	4440	0.01	0.01	0	0.02	NA	0.56	0.82	0.48	0.93	NA	C	0.68	0.149	4.81E-06
rs9488238	6	NA	4464	0.04	0.03	0.02	0.03	NA	0.87	0.88	0.87	0.97	NA	A	-0.28	0.061	5.17E-06
rs763014	16	RAB40C	3939	0.42	0.43	NA	0.43	0.47	0.98	0.99	NA	0.42	99.4	C	0.1	0.023	5.43E-06
rs1205863	6	NA	3939	0.06	0.06	NA	0.06	0.07	0.99	1	NA	0.99	98.23	G	0.21	0.047	5.64E-06
rs11918654	3	ARL8B	4310	0.27	0.27	0.29	0.27	NA	0.97	0.98	0.9	0.99	NA	C	-0.11	0.024	7.10E-06
rs4881085	10	PFKP	4490	0.1	0.1	0.1	0.13	NA	1	1	0.99	0.67	NA	A	0.16	0.036	7.44E-06
rs2657195	8	NA	4316	0.22	0.21	0.18	0.21	NA	0.85	0.85	0.87	0.74	NA	A	-0.13	0.028	7.54E-06
rs11754641	6	NA	4478	0.03	0.03	0.02	0.04	NA	0.93	0.95	0.91	1	NA	C	0.29	0.064	7.84E-06
rs12805875	11	NA	4674	0.42	0.41	0.42	0.43	0.41	1	1	1	1	98.12	A	0.09	0.021	8.51E-06
rs980238	8	NA	4491	0.3	0.29	0.31	0.29	NA	1	1	1	1	NA	A	-0.1	0.023	9.11E-06
rs7103004	11	NA	4667	0.42	0.41	0.42	0.43	0.41	1	1	0.99	1	97.09	C	0.09	0.021	9.20E-06
rs4554859	11	NA	4674	0.42	0.41	0.42	0.43	0.41	1	1	1	1	98.12	G	0.09	0.021	9.24E-06
rs7118412	11	NA	4669	0.42	0.41	0.42	0.43	0.41	1	1	0.99	1	97.09	A	0.09	0.021	9.37E-06
rs2017567	16	NA	4683	0.42	0.43	0.42	0.42	0.47	0.99	1	1	0.45	100	C	0.1	0.021	9.38E-06
rs710924	16	PIGQ	4683	0.42	0.43	0.42	0.42	0.47	1	1	1	0.45	99.64	C	0.09	0.021	9.93E-06

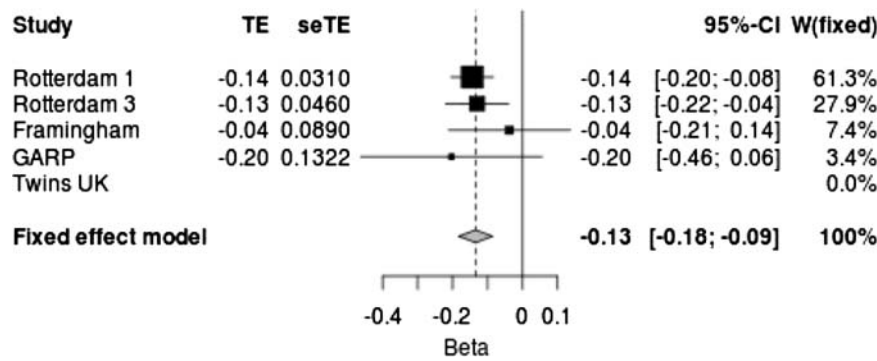
\* $r^2$  from MACH for RS1, RS3, FHS and GARP; information score from IMPUTE for TUK.

Studies contributing data are denoted.

SNP, single nucleotide polymorphism; Chr, chromosome; N, number of subjects studied; MAF, minor allele frequency; RS1, Rotterdam study cohort 1; RS3, Rotterdam study cohort 3; TUK, TwinsUK; FHS, Framingham Heart Study; GARP, Genetics of OsteoArthrosis and Progression study; Eff all, effect allele;  $\beta$ , effect size; SE, SE of  $\beta$ .

Leukocyte Antigen (HLA) region and to markers in PARK2 (Parkinson protein 2, E3 ubiquitin protein ligase). Among the most significant findings (table 2) is SNP rs926849 that lies at 6q25.2–27 within an intron in the PARK2 gene, a large gene of 1.3 Mb comprising 12 exons. The SNP encodes a change of base from T to C and is reported to have a minor allele frequency of 0.23–0.34 in dbSNP, which is keeping with the findings in our study groups (table 2, figures 3 and 4). Although this SNP has not been directly genotyped by any study group, estimates

suggest imputation to be accurate for rs926849 (range 95%–99%, table 2). PARK2 encodes a protein called parkin, which is a component of a multiprotein E3 ubiquitin ligase complex that mediates the targeting of unwanted proteins for proteasomal degradation. This complex also controls the level of proteins involved in cell activities such as cell division and growth: it may be a tumour suppressor protein. Parkin is also involved in mitophagy: it translocates from the cytosol and promotes the degradation of uncoupled mitochondria.<sup>38</sup> Parkin is widely



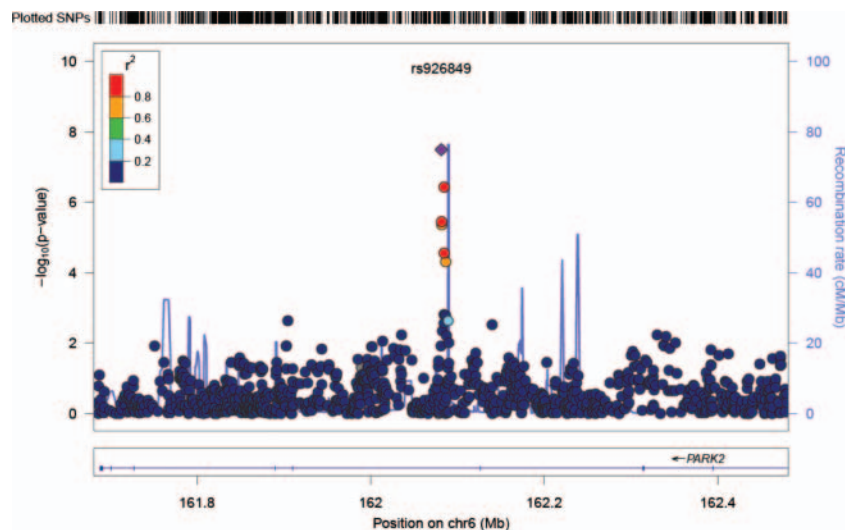
**Figure 3** Forest plot of rs926849 in PARK2 unadjusted for covariates. The contribution of the studies included in the meta-analysis is shown in this fixed effects model. The C allele is considered. Heterogeneity,  $I^2=0\%$ ;  $p(Q)=0.67$ . TE, treatment effect; seTE, standard error treatment effect.

expressed in solid organs as well as skeletal muscle (<http://www.proteinatlas.org/>). Mutations within PARK2 are associated with diverse conditions including autosomal recessive juvenile Parkinson's disease, Alzheimer's disease, diabetes mellitus and several solid tumours (reviewed in<sup>39</sup>). Parkin may account for the inverse relation between Parkinson's disease and cancer incidence.<sup>40</sup> Our findings of hypermethylation with increasing LDD score suggest that PARK2 expression is reduced with increasing disc degeneration but functional studies of intervertebral disc and other spine tissues are needed.

Three further markers in the unadjusted meta-analysis had  $p < 5 \times 10^{-8}$ . Marker rs17034687 is an intergenic marker on chromosome 3. Based on One Thousand Genomes (1KG)/CEU data, it is not in LD ( $r^2 > 0.3$ ) with any known gene-based markers. Markers rs2187689 (supplementary figure S2) and rs7767277 are HLA region markers, neither of which is included in the 1KG pilot data. Using data from HapMap V3 (release 2), rs2187689 and rs7767277 are in perfect LD with each other and in LD ( $r^2 = 0.76$ ) with an intronic marker in PSMB9 (proteasome (prosome, macropain) subunit,  $\beta$  type 9; large

multifunctional peptidase 2). Proteasomes are distributed throughout eukaryotic cells at high concentration and cleave peptides in an ATP/ubiquitin-dependent process in a non-lysosomal pathway. The gene is located in the class II region of the MHC. Expression of the gene is induced by interferon  $\gamma$  and this gene product replaces catalytic subunit 1 (proteasome  $\beta$  6 subunit) in the immunoproteasome.

While lumbar degeneration is not considered an inflammatory process and has not been reported to be auto-immune in aetiology, there is evidence of pro-inflammatory cytokine activation in herniated lumbar discs<sup>41</sup> and anti-TNF has been used successfully to treat disc herniation-induced sciatica.<sup>42</sup> Of note, the COL11A2 gene lies 169 KB upstream from rs2187689. An SNP (rs2076311) within this candidate gene has been shown to be associated with MR determined disc signal intensity in a candidate gene study of Finnish male twins.<sup>43</sup> SNP rs2076311 is not, however, in LD with our top hit, rs2187689 ( $r^2 = 0.017$ ) and so it seems unlikely that this collagen-encoding gene accounts for our observed SNP association. Many published GWA studies have identified SNPs in intergenic regions and the



**Figure 4** Regional plot of association results and recombination rates for the PARK2 gene, unadjusted results.  $-\log_{10} p$  values (y-axis) of the single nucleotide polymorphisms (SNPs) are shown according to their chromosomal positions (x-axis) with lead SNP shown as a purple diamond. The colour intensity of each symbol depicting an SNP reflects the extent of LD with the rs926849, coloured red ( $r^2 > 0.8$ ) through to blue ( $r^2 < 0.2$ ). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on build 36 (NCBI) of the human genome. Also shown are the relative positions of genes mapping to the region of association. Genes have been redrawn to show the relative positions and, therefore, the maps are not to physical scale.

precise role of these regions is yet to be defined. Long range enhancers, for example, could operate here and so an influence on COL11A2 expression cannot be ruled out.

Of suggestive significance is SNP rs4802666 ( $p=3.76 \times 10^{-6}$ , adjusted meta-analysis) which lies within the MYH14 gene which encodes myosin, heavy chain 14, non-muscle. It is expressed in cell lines derived from bone (<http://www.proteinatlas.org>) and is implicated in autosomal dominant hearing impairment. It is of interest in LDD because it lies on chromosome 19 under the linkage peak we have reported in twins for LDD<sup>44</sup> and a peak reported by the Framingham group for hand OA.<sup>45</sup> As there is a known relationship between these two phenotypes, this region on chromosome 19 forms a highly plausible candidate region for OA. It is not impossible that a muscle-expressed protein plays a role in LDD through mechanisms similar to those proposed for OA, considered by some to be a multi-tissue syndrome rather than simply a disease of cartilage.<sup>46</sup>

The main limitation of the study is one of obtaining an accurate phenotype on individuals which is known to be an important factor in the success of GWA.<sup>47</sup> There is at present no agreed gold standard imaging method in the determination of LDD, although it is recognised that MRI offers the most sensitive, widely available tool. Even so, MR is relatively expensive and many of the largest spine cohorts in the world have plain radiographs, which offer more limited phenotypic information. The coding method applied to the imaging is also yet to be formally standardised: our interest in the individual subtraits of LDD led us to devise a coding method in which they were separated, as reported previously.<sup>7</sup> In order to obtain sufficient sample size, a number of cohorts contributed having different imaging methods, but traits were selected to enable comparison across the cohorts. Thus, study groups recoded their imaging where necessary to meet uniform requirements for inclusion. We included measures of disc height (coded 0–3) and anterior osteophytes in RS1, RS3, GARP and TUK (also coded 0–3) and posterior osteophytes in FHS (coded 0–3). These subphenotypes were summed over the five discs and underwent inverse normal transformation to give a normal distribution. A further limitation is that four cohorts are population samples while GARP is derived from OA-affected sibling pairs. We included GARP because it has made a contribution to similar analyses performed for OA<sup>48</sup> and, with adjustment for relatedness, provides data comparable with other studies. While the differing methods of imaging provide different amounts of information so the LDD variable has lower mean in those cohorts with radiographs, the variance is comparable. Where GARP samples made a contribution to the meta-analysis (a number of the significant SNPs did not include a contribution from GARP, table 2), the minor allele frequency was similar to those of other groups. The TUK group has a disproportionate number of women, for historical reasons. The men were retained, however, as they did not differ significantly from women in the LDD variable or body mass index (data not shown). This study lacks a replication group. A second sample of similar size to the first is considered important to show that the findings of the first sample are true positives. Unfortunately there are, to our knowledge, no other collections of Northern Europeans having spine imaging which together would approach our sample size. There is considerable evidence in the literature that the genetic predisposition between Northern Europeans and Asians to OA is different<sup>49</sup> and, given the similarities between OA and LDD, we felt replication should be made in Northern Europeans. We elected to include all the subjects in

a single, powerful study rather than split the sample and reduce the chances of finding significant novel loci associated with LDD.

In conclusion, this is the first large-scale GWA study of LDD and we have identified several novel variants in the PARK2 gene and in PSMB9 within MHC class 2. We have shown in a small subset that methylation at one of the PARK2 promoters is associated with MRI determined LDD. Both loci merit further investigation to shed light on the important clinical endpoint of low back pain.

#### Author affiliations

<sup>1</sup>Department Twin Research and Genetic Epidemiology, King's College London, London, UK

<sup>2</sup>Acclarogen Ltd, St John's Innovation Centre, Cambridge, UK

<sup>3</sup>Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands

<sup>4</sup>The Netherlands Genomics Initiative sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA), Leiden/Rotterdam, The Netherlands

<sup>5</sup>Department of Molecular Epidemiology, Leiden University Medical Center, Leiden/Rotterdam, The Netherlands

<sup>6</sup>VA Boston Healthcare System, Harvard Medical School, Spaulding Rehabilitation Hospital, New England Baptist Hospital, Boston, USA

<sup>7</sup>Department of Rheumatology, University of Sydney and Royal North Shore Hospital, Sydney, Australia

<sup>8</sup>Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands

<sup>9</sup>Department of General Practice, Erasmus MC, Rotterdam, The Netherlands

<sup>10</sup>Department of Rheumatology and Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

**Acknowledgements** FMKW is a Wellcome Trust Intermediate Fellow. The authors would like to acknowledge the contribution of Maria Popham to the collection and analysis of TwinsUK data. The GWA study was funded by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2) and the Netherlands Genomics Initiative (NGI)/the Netherlands Consortium for Healthy Aging (NCHA) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Dr Michael Moorhouse, Marijn Verkerk and Sander Bervoets for their help in creating the GWAS database. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organisation for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are very grateful to the participants and staff from the Rotterdam Study, the participating general practitioners and the pharmacists. We would like to thank Dr Tobias A Knoch, Luc V de Zeeuw, Anis Abuseiris and Rob de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources. The GARP study was supported by the Leiden University Medical Centre and the Dutch Arthritis Association. Pfizer Inc., Groton, CT, USA supported the inclusion of the GARP study. The genotypic work was supported by the Netherlands Organisation of Scientific Research (MW 904-61-095, 911-03-016, 917 66344 and 911-03-012), Leiden University Medical Centre and the Centre of Medical System Biology and the Netherlands Consortium for Healthy Aging both in the framework of the Netherlands Genomics Initiative (NGI). The research leading to these studies has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement n° 259679. Pradeep Suri was funded by K12HDO1097 from the National Institute of Health (US) and the Rehabilitation Medicine Scientist Training Programme (RMSTP). The Framingham Study is supported by the National Heart, Lung and Blood Institute's Framingham Heart Study contract (No. N01-HC-25195) for the recruitment, enrolment and examination of the Offspring and Third Generation cohorts and the imaging by CT scan.

**Contributors** The authors listed fulfilled the criteria for authorship.

**Funding** Fellowship to FMKW who led the project.

**Competing interests** None.

**Ethics approval** All bodies appropriate to the groups contributing data.

**Provenance and peer review** Not commissioned; externally peer reviewed.

#### REFERENCES

1. Powell MC, Wilson M, Szypryt P, *et al*. Prevalence of lumbar disc degeneration observed by magnetic resonance in symptomless women. *Lancet* 1986;**2**:1366–7.

2. **de Schepper EI**, Damen J, van Meurs JB, *et al*. The association between lumbar disc degeneration and low back pain: the influence of age, gender, and individual radiographic features. *Spine (Phila Pa 1976)* 2010;**35**:531–6.
3. **MacGregor AJ**, Andrew T, Sambrook PN, *et al*. Structural, psychological, and genetic influences on low back and neck pain: a study of adult female twins. *Arthritis Rheum* 2004;**51**:160–7.
4. **Phillips CJ**. Economic burden of chronic pain. *Expert Rev Pharmacoecon Outcomes Res* 2006;**6**:591–601.
5. **Roberts S**, Evans H, Trivedi J, *et al*. Histology and pathology of the human intervertebral disc. *J Bone Joint Surg Am* 2006;**88**(Suppl 2):10–14.
6. **Battie MC**, Levalahti E, Videman T, *et al*. Heritability of lumbar flexibility and the role of disc degeneration and body weight. *J Appl Physiol* 2008;**104**:379–85.
7. **Sambrook PN**, MacGregor AJ, Spector TD. Genetic influences on cervical and lumbar disc degeneration: a magnetic resonance imaging study in twins. *Arthritis Rheum* 1999;**42**:366–72.
8. **Ryder JJ**, Garrison K, Song F, *et al*. Genetic associations in peripheral joint osteoarthritis and spinal degenerative disease: a systematic review. *Ann Rheum Dis* 2007;**67**:584–91.
9. **Valdes AM**, Loughlin J, Oene MV, *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* 2006;**56**:137–46.
10. **Richards JB**, Kavvoura FK, Rivadeneira F, *et al*. Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture. *Ann Intern Med* 2009;**151**:528–37.
11. **Manolio TA**, Collins FS, Cox NJ, *et al*. Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747–53.
12. **Hoffmann U**, Siebert U, Bull-Stewart A, *et al*. Evidence for lower variability of coronary artery calcium mineral mass measurements by multi-detector computed tomography in a community-based cohort—consequences for progression studies. *Eur J Radiol* 2006;**57**:396–402.
13. **Riyazi N**, Meulenbelt I, Kroon HM, *et al*. Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* 2005;**64**:438–43.
14. **Lane NE**, Nevitt MC, Genant HK, *et al*. Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. *J Rheumatol* 1993;**20**:1911–18.
15. **Hofman A**, Breteler MM, van Duijn CM, The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol* 2009;**24**:553–72.
16. **Bijkerk C**, Houwing-Duistermaat JJ, Valkenburg HA, *et al*. Heritabilities of radiologic osteoarthritis in peripheral joints and of disc degeneration of the spine. *Arthritis Rheum* 1999;**42**:1729–35.
17. **Spector TD**, Williams FM. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* 2006;**9**:899–906.
18. **Suchindran S**, Rivedal D, Guyton JR, *et al*. Genome-wide association study of Lp-PLA(2) activity and mass in the Framingham Heart Study. *PLoS Genet* 2010;**6**:e1000928.
19. **Howie BN**, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;**5**:e1000529.
20. **Marchini J**, Howie B, Myers S, *et al*. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906–13.
21. **Uh HW**, Wijk HJ, Houwing-Duistermaat JJ. Testing for genetic association taking into account phenotypic information of relatives. *BMC Proc* 2009;**3**(Suppl 7):S123.
22. **Richards JB**, Rivadeneira F, Inouye M, *et al*. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* 2008;**371**:1505–12.
23. Stata for Windows. Version 10 edn. College Station, Texas, USA: Statacorp, 2007.
24. **Pritchard JK**, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;**155**:945–59.
25. **Aulchenko YS**, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* 2010;**11**:134.
26. **Li Y**, Willer CJ, Ding J, *et al*. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010;**34**:816–34.
27. **Bell JT**, Tsai PC, Yang TP, *et al*. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 2012;**8**:e1002629.
28. **Bell JT**, Pai AA, Pickrell JK, *et al*. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 2011;**12**:R10.
29. **Devlin B**, Roeder K. Genomic control for association studies. *Biometrics* 1999;**55**:997–1004.
30. **Liuke M**, Solovieva S, Lamminen A, *et al*. Disc degeneration of the lumbar spine in relation to overweight. *Int J Obes (Lond)* 2005;**29**:903–8.
31. **Pye SR**, Reid DM, Adams JE, *et al*. Influence of weight, body mass index and lifestyle factors on radiographic features of lumbar disc degeneration. *Ann Rheum Dis* 2007;**66**:426–7.
32. **Symons DP**, van Hemert AM, Vandenbroucke JP, *et al*. A longitudinal study of back pain and radiological changes in the lumbar spines of middle aged women. II. Radiographic findings. *Ann Rheum Dis* 1991;**50**:162–6.
33. **Williams FM**, Popham M, Livshits G, *et al*. A response to Videman, *et al*, 'challenging the cumulative injury model: positive effects of greater body mass on disc degeneration'. *Spine J* 2010;**10**:571–2.
34. **Yoshimura N**, Dennison E, Wilman C, *et al*. Epidemiology of chronic disc degeneration and osteoarthritis of the lumbar spine in Britain and Japan: a comparative study. *J Rheumatol* 2000;**27**:429–33.
35. **Battie MC**, Videman T, Gill K, *et al*. 1991 Volvo Award in clinical sciences. Smoking and lumbar intervertebral disc degeneration: an MRI study of identical twins. *Spine* 1991;**16**:1015–21.
36. **Videman T**, Battie MC. The influence of occupation on lumbar degeneration. *Spine* 1999;**24**:1164–8.
37. **Williams FM**, Sambrook PN. Neck and back pain and intervertebral disc degeneration: role of occupational factors. *Best Pract Res Clin Rheumatol* 2011;**25**:69–79.
38. **Tanaka A**. Parkin-mediated selective mitochondrial autophagy, mitophagy: Parkin purges damaged organelles from the vital mitochondrial network. *FEBS Lett* 2010;**584**:1386–92.
39. **Kay DM**, Stevens CF, Hamza TH, *et al*. A comprehensive analysis of deletions, duplications, and copy number variations in PARK2. *Neurology* 2010;**75**:1189–94.
40. **Gao X**, Ning Y. Cancer and Parkinson's disease: the odd couple. *Drugs Today (Barc)* 2011;**47**:215–22.
41. **Kang JD**, Georgescu HI, McIntyre-Larkin L, *et al*. Herniated lumbar intervertebral discs spontaneously produce matrix metalloproteinases, nitric oxide, interleukin-6, and prostaglandin E2. *Spine (Phila Pa 1976)* 1996;**21**:271–7.
42. **Korhonen T**, Karppinen J, Paimela L, *et al*. The treatment of disc-herniation-induced sciatica with infliximab: one-year follow-up results of FIRST II, a randomized controlled trial. *Spine (Phila Pa 1976)* 2006;**31**:2759–66.
43. **Videman T**, Saarela J, Kaprio J, *et al*. Associations of 25 structural, degradative, and inflammatory candidate genes with lumbar disc desiccation, bulging, and height narrowing. *Arthritis Rheum* 2009;**60**:470–81.
44. **Williams FM**, Kato BS, Livshits G, *et al*. Lumbar disc disease shows linkage to chromosome 19 overlapping with a QTL for hand OA. *Ann Rheum Dis* 2008;**67**:117–19.
45. **Demissie S**, Cupples LA, Myers R, *et al*. Genome scan for quantity of hand osteoarthritis: the Framingham Study. *Arthritis Rheum* 2002;**46**:946–52.
46. **Brandt KD**, Radin EL, Dieppe PA, *et al*. Yet more evidence that osteoarthritis is not a cartilage disease. *Ann Rheum Dis* 2006;**65**:1261–4.
47. **Amos CI**. Successful design and conduct of genome-wide association studies. *Hum Mol Genet* 2007;**16**:R220–R225.
48. **Zhai G**, van Meurs JB, Livshits G, *et al*. A genome-wide association study suggests that a locus within the ataxin 2 binding protein 1 gene is associated with hand osteoarthritis: the Treat-OA consortium. *J Med Genet* 2009;**46**:614–16.
49. **Valdes AM**, Spector TD. Genetic epidemiology of hip and knee osteoarthritis. *Nat Rev Rheumatol* 2011;**7**:23–32.



**Supplementary Table 1. Genotyping and imputation methods by study**

	<b>FHS</b>	<b>GARP</b>	<b>RS1</b>	<b>RS3</b>	<b>TUK</b>
<i><u>Genotyping</u></i>					
Platform	Affymetrix	Illumina	Illumina	Illumina	Illumina
Chip	500K or 100K or 10 K	Human660W quad	Human Hap550	Human Hap550	Human Hap300 or Human Hap610Q
Call rate	>0.97	> 0.95	> 0.975	> 0.975	>0.90
MAF	> 0.05	> 0.0025	> 0.01	> 0.01	>0.01
HWE p-value	> 10 <sup>-6</sup>	> 10 <sup>-6</sup>	> 10 <sup>-6</sup>	> 10 <sup>-6</sup>	> 10 <sup>-4</sup>
<i><u>Imputation</u></i>					
Variants included for imputation	534,982	462,361	512,349	514,073	314,075
Imputation software	MACH vs1	IMPUTE	MACH	MACH	IMPUTE vs 2
Genome build	HapMap Phase II release 22	Hapmap Phase II release 21	Hapmap Phase II release 22	Hapmap Phase II release 22	Hapmap Phase II release 36
Total number of SNPs available	2,543,887	1,830,498	2,543,887	2,543,887	3,044,064

**Legend to Supplementary Table 1**

FHS represents Framingham Heart Study; GARP, Genetics of OsteoArthrosis and Progression study; RS1, Rotterdam study cohort 1; RS3, Rotterdam study cohort 3; TUK, TwinsUK: BMI, body mass index; MAF, Minor Allele Frequency; HWE, Hardy-Weinberg equilibrium

**Supplementary Table 2. Results of the GWA meta-analysis adjusted for age and sex, showing those SNPs having  $p < 10^{-5}$ .**

SNP	Chr	gene	N	MAF					Imputation Quality*					Eff	BETA	SE	P
				RS1	RS3	TUK	FHS	GARP	RS1	RS3	TUK	FHS	GARP				
rs926849	6	PARK2	3939	0.31	0.32	NA	0.31	0.23	0.98	0.99	NA	0.9	95.04	C	-0.12	0.023	9.50E-08
rs17034687	3	NA	4429	0.09	0.1	0.05	0.08	NA	0.92	0.93	0.88	0.82	NA	C	0.19	0.037	3.67E-07
rs2187689	6	NA	4457	0.08	0.07	0.06	0.09	NA	0.98	0.99	0.95	0.93	NA	C	0.19	0.039	6.63E-07
rs7767277	6	NA	4457	0.08	0.07	0.06	0.09	NA	0.98	0.99	0.95	0.93	NA	A	0.19	0.039	6.64E-07
rs9488238	6	NA	4464	0.04	0.03	0.02	0.03	NA	0.87	0.88	0.87	0.97	NA	A	-0.28	0.058	7.30E-07
rs3019449	6	PARK2	4636	0.32	0.32	0.31	0.31	0.28	0.98	0.98	0.98	0.99	97.18	A	-0.11	0.022	7.42E-07
rs9301951	13	GPC6	4397	0.04	0.04	0.03	0.05	NA	0.96	0.97	0.74	0.78	NA	C	-0.25	0.051	1.35E-06
rs11136566	8	CSMD1	4610	0.39	0.39	0.36	0.39	0.46	0.99	0.98	0.97	0.96	97.39	A	-0.1	0.021	2.24E-06
rs7744666	6	NA	4466	0.1	0.09	0.06	0.1	NA	0.99	1	0.97	0.96	NA	C	0.16	0.035	3.07E-06
rs11969002	6	NA	4466	0.1	0.09	0.06	0.1	NA	0.99	1	0.97	0.96	NA	A	0.16	0.035	3.07E-06
rs365962	16	NA	4608	0.44	0.44	0.48	0.47	0.42	0.99	0.99	0.96	0.69	98.61	C	-0.1	0.020	3.10E-06
rs387953	16	NA	4683	0.46	0.45	0.48	0.46	0.43	1	0.99	1	0.67	99.99	G	-0.09	0.020	3.37E-06
rs737631	6	PARK2	4303	0.28	0.29	0.23	0.28	NA	0.92	0.91	0.89	0.83	NA	A	-0.11	0.023	3.57E-06
rs4802666	19	MYH14	4094	0.27	0.27	0.18	0.27	NA	0.7	0.74	0.76	0.64	NA	A	-0.13	0.027	3.76E-06
rs1154053	8	NA	4639	0.17	0.17	0.18	0.16	0.2	0.99	1	0.96	0.89	96.17	C	-0.12	0.026	4.68E-06
rs1205863	6	NA	3939	0.06	0.06	NA	0.06	0.07	0.99	1	NA	0.99	98.23	G	0.21	0.045	4.77E-06
rs3749982	6	NA	4458	0.1	0.09	0.06	0.1	NA	0.99	1	0.96	0.96	NA	A	0.16	0.035	4.97E-06
rs6457690	6	NA	4464	0.1	0.09	0.07	0.11	NA	0.98	1	0.96	0.97	NA	A	0.16	0.035	5.93E-06
rs1029296	6	NA	4464	0.1	0.09	0.07	0.11	NA	0.98	1	0.96	0.97	NA	C	0.16	0.035	5.93E-06
rs4875102	8	NA	4608	0.26	0.26	0.27	0.25	0.29	0.99	0.99	0.95	0.91	97.65	A	-0.1	0.023	6.24E-06
rs7234567	18	NA	4682	0.26	0.26	0.22	0.22	0.29	1	1	1	0.89	97.67	A	-0.1	0.023	6.34E-06

rs6936004	6	NA	4462	0.1	0.09	0.07	0.11	NA	0.98	1	0.96	0.97	NA	C	0.16	0.035	6.52E-06
rs1884158	6	PARK2	4491	0.31	0.32	0.31	0.31	NA	0.98	0.99	0.99	0.91	NA	C	-0.1	0.022	6.54E-06
rs1981483	16	PIGQ	3939	0.42	0.43	NA	0.42	0.47	0.99	0.98	NA	0.44	99.64	A	0.1	0.022	6.79E-06
rs10890236	1	NA	4683	0.09	0.09	0.09	0.09	0.09	1	1	1	0.93	100	C	0.16	0.035	7.07E-06
rs1029295	6	NA	3747	0.1	0.09	NA	0.11	NA	0.98	1	NA	0.97	NA	C	0.16	0.037	7.08E-06
rs763014	16	RAB40C	3939	0.42	0.43	NA	0.43	0.47	0.98	0.99	NA	0.42	99.4	C	0.1	0.022	7.28E-06
rs10046257	6	NA	4461	0.1	0.09	0.08	0.11	NA	0.98	1	0.96	0.97	NA	A	0.15	0.034	8.21E-06
rs7204439	16	RAB40C	3939	0.42	0.43	NA	0.43	0.47	0.97	0.98	NA	0.43	97.37	C	0.1	0.022	8.23E-06
rs9469300	6	NA	4482	0.1	0.09	0.07	0.1	NA	0.99	1	0.92	0.96	NA	A	0.15	0.035	8.73E-06
rs7205409	16	RAB40C	4590	0.42	0.43	0.45	0.43	0.47	0.99	0.99	0.96	0.44	99.33	C	0.09	0.021	8.78E-06
rs10998466	10	NA	4446	0.01	0.01	0.01	0.02	NA	0.75	0.97	0.73	0.51	NA	A	-0.47	0.107	9.70E-06

### Legend to Supplementary Table 2

Studies contributing data are denoted RS1: Rotterdam study cohort 1; RS3: Rotterdam study cohort 3; TUK: TwinsUK; BMI: body mass index; FHS: Framingham Heart Study; GARP: Genetics of OsteoArthrosis and Progression study;

SNP single nucleotide polymorphism; Chr chromosome; position, SNP location in base pairs; MAF minor allele frequency;

\*R<sup>2</sup> from MACH for RS1, RS3, FHS and GARP; Information score from IMPUTE for TUK

Eff All, effect allele; beta, effect size; SE, standard error of beta; p, p value

Supplementary Figure 1. Quantile-quantile plot of meta-analysis adjusted GWA results

**Legend to Supplementary Figure 1**

The plots show GWA meta-analysis quantile-quantile plot of observed against expected results, adjusted for age and sex

**Supplementary Figure 2. Manhattan plot for meta-analysis of adjusted GWA results**

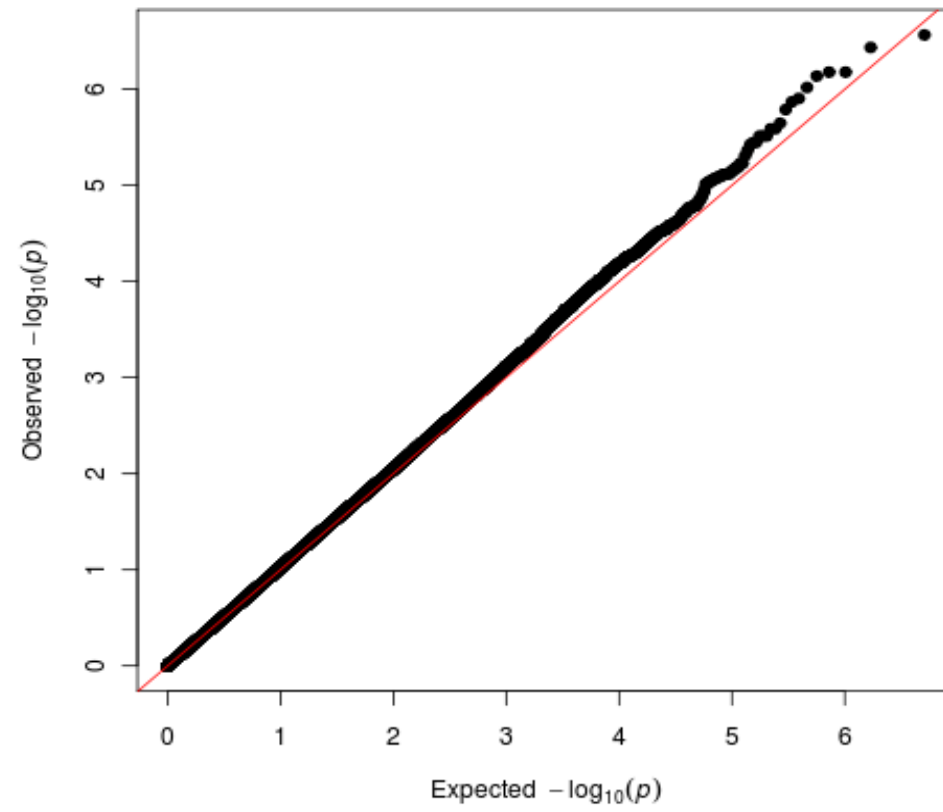
**Legend to Supplementary Figure 2**

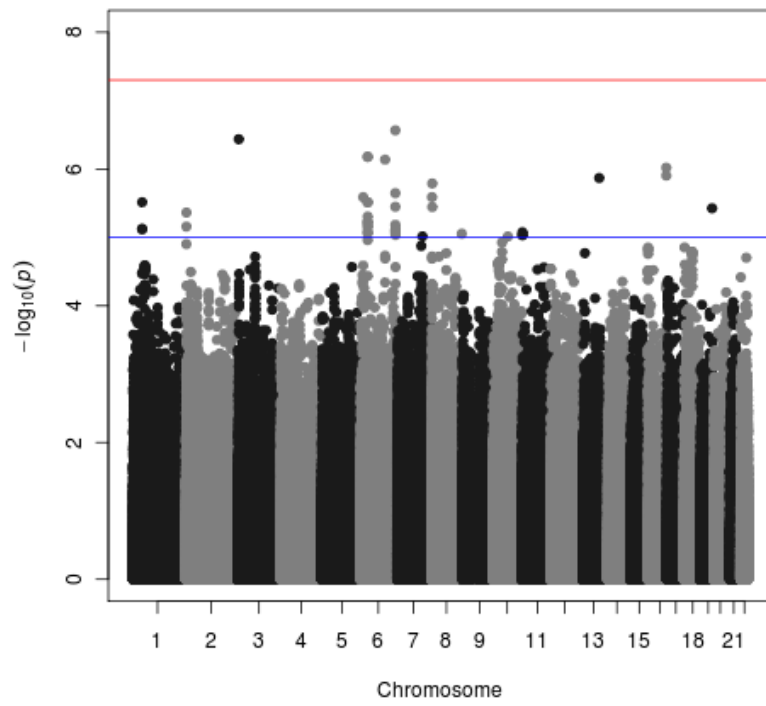
Plot shows combined results for the 5 studies included in the meta-analysis, adjusted for age and sex. The blue and red horizontal lines mark the levels of suggestive and likely significance, respectively

**Supplementary Figure 3. Regional plot of association results and recombination rates for the class II MHC region unadjusted for age and sex**

**Legend to Supplementary Figure 3**

$-\log_{10} P$  values ( $y$  axis) of the SNPs are shown according to their chromosomal positions ( $x$  axis) with lead SNP shown as a purple diamond. The colour intensity of each symbol depicting a SNP reflects the extent of LD with the rs926849, coloured red ( $r^2 > 0.8$ ) through to blue ( $r^2 < 0.2$ ). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on build 36 (NCBI) of the human genome. Also shown are the relative positions of genes mapping to the region of association. Genes have been redrawn to show the relative positions, and therefore, the maps are not to physical scale.







Plotted SNPs

