EXTENDED REPORT

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

Frances M K Williams,1 Aruna T Bansal,2 Joyce B van Meurs,3,4 Jordana T Bell,1 Ingrid Meulenbelt,4,5 Pradeep Suri,6 Fernando Rivadeneira,3,4 Philip N Sambrook,7 Albert Hofman,8 Sita Bierma-Zeinstra,9 Cristina Menni,1 Margreet Kloppenburg,10 P Eline Slagboom,4,5 David J Hunter,7 Alex J MacGregor,1 Andre G Uitterlinden,3 Tim D Spector1

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 aetogenic 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 genome-wide association (GW A) data imputed to HapMap V2.

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

Conclusions LDD accounts for a considerable proportion of low back pain and the pathogenesis of LDD is poorly understood. This study 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:1 over a third of middle aged women have at least one degenerate disc. LDD contributes to low back pain2,3 and as low back pain is common in the general population and costly to society,4 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.5 There are similarities with peripheral joint osteoarthritis (OA). LDD has been shown to be heritable, with estimates of 65%–80%6,7 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 al8). 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.9 That some published genome-wide associations (GWAs) in common complex traits fail to replicate candidate gene findings suggests limitations to the candidate gene method.10 As in other common complex traits,11 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 aetogenic 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

For numbered af 201551).

1136/annrheumdis-2012-201551.


1141
Clinical and epidemiological research

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 50 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 specification of CT scanning having been reported elsewhere.12 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.14 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. Probands (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.13 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 as described above.14 The scores for the two traits over the five lumbar discs were summed.

The TUK registry was described previously.17 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.7 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.14 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.10

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 V21.2021 Strict selection criteria were applied to the measured genotypes using a high information content (r2 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.21

RSl 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 >0.38, 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 Hap500 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.22 Genotyping results had been sent to...
KCL for collation and analysis using the statistical package, STATA (StataCorp). Strict quality control was applied: 314,075 SNPs were retained for analysis (98.7%); 735 were excluded because their call rates were ≤90% and 725 SNPs had minor allele frequency <0.01. In TUK, significant population substructure was excluded using the STRUCTURE program.

### 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 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 or the genotype probabilities from IMPUTE.

### Meta-analysis of the five study groups

Genotypes for 2.5–3 million autosomal SNPs were imputed separately to increase coverage using HapMap V2 (http://www.hapmap.org) as the reference panel. In GARP and TUK, imputation was performed with IMPUTE V2 and in the other studies with MACH. The common reference panel led to the commonality 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 β 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 V1.06 (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 I² (ratio of true heterogeneity to total observed variation). SNPs were excluded from the meta-analysis if the cohort-specific imputation quality, as assessed by r² (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. 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 β, 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 (cg15832456, cg21926612 and cg24816866) mapped uniquely to the human genome (hg18) within two mismatches (see Bell et al28). 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 (β). The genotyping and imputation methods are shown in online supplementary table S1.

Quantile–quantile plots for the unadjusted LDD GWA 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, was low (λGC unadjusted=1.02; λGC 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

<table>
<thead>
<tr>
<th></th>
<th>FHS</th>
<th>GARPA</th>
<th>RS1</th>
<th>RS3</th>
<th>TUK</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>330</td>
<td>192</td>
<td>2440</td>
<td>974</td>
<td>744</td>
</tr>
<tr>
<td>Age (years)</td>
<td>54.3 (11.0)</td>
<td>60.3 (7.1)</td>
<td>65.7 (6.7)</td>
<td>54.7 (3.4)</td>
<td>53.6 (8.3)</td>
</tr>
<tr>
<td>Women (%)</td>
<td>42.2</td>
<td>79.7</td>
<td>57</td>
<td>59</td>
<td>96.8</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>28.1 (5.1)</td>
<td>26.4 (4.8)</td>
<td>26.3 (3.4)</td>
<td>27.12 (4.6)</td>
<td>24.9 (4.4)</td>
</tr>
<tr>
<td>Lumbar spine imaging</td>
<td>CT</td>
<td>Radiograph</td>
<td>Radiograph</td>
<td>Radiograph</td>
<td>MRI</td>
</tr>
<tr>
<td>LDD variable</td>
<td>2.49 (0.97)</td>
<td>0.02 (0.958)</td>
<td>0.006 (0.978)</td>
<td>0.011 (0.965)</td>
<td>3.46 (1.14)</td>
</tr>
</tbody>
</table>

Values are mean (SD) unless specified otherwise.

BMI, body mass index; FHS, Framingham Heart Study; GARPA, Genetics of OsteoArthrosis and Progression study; LDD, Lumbar disc degeneration; RS1, Rotterdam study cohort 1; RS3, Rotterdam study cohort 3; TUK, TwinsUK.
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 5 (rs17034687). The results of the meta-analysis adjusted for age and gender were broadly similar: the strongest association was on chromosome 3 (rs17034687). The results of the meta-analysis (rs926849; rs2187689; rs7767277) and an intergenic marker on chromosome 5 were 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 $\beta=6.76 \times 10^{-4}$, $p=0.59$). We repeated the analyses using normalised methylation levels and observed that the association between cg15832436 and LDD remained nominally significant.

**DISCUSSION**

GWAS 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. LDD is known to have genetic determinants and its expression is also influenced by gender (women develop LDD later), body mass index and smoking. Occupational factors also play a small role in LDD. LDD as determined by MRI has been implicated in the development of episodes of severe and disabling low back pain. 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.

GWAS 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 $p<10^{-7}$ 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 Genome Project.
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 0.25–0.54 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 pro-mal degradation.

Clinical and epidemiological research
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 39). Parkin may account for the inverse relation between Parkinson’s disease and cancer incidence. 40 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 eukaroytic 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 41 and anti-TNF has been used successfully to treat disc herniation-induced sciatica. 42 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. 43 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 association. Many published GWAS studies have identified SNPs in intergenic regions and the
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 ($p=3.76\times10^{-6}$, 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.proteintatlas.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$^{44}$ and a peak reported by the Framingham group for hand OA.$^{45}$ 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.$^{46}$

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.$^{47}$ 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.$^{7}$ 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 0–5) 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$^{48}$ 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 disproportinate 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$^{49}$ 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

1Department Twin Research and Genetic Epidemiology, King’s College London, London, UK
2Acclarogen Ltd, St John’s Innovation Centre, Cambridge, UK
3Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands
4The Netherlands Genomics Initiative sponsored Netherlands Consortium for Healthy Aging (NG-NCHA), Leiden/Rotterdam, The Netherlands
5Department of Molecular Epidemiology, Leiden University Medical Center, Leiden/Rotterdam, The Netherlands
6VA Boston Healthcare System, Harvard Medical School, Spaulding Rehabilitation Hospital, New England Baptist Hospital, Boston, USA
7Department of Rheumatology, University of Sydney and Royal North Shore Hospital, Sydney, Australia
8Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands
9Department of General Practice, Erasmus MC, Rotterdam, The Netherlands
10Department 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 Papom 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 Abusereis 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 Bundesminister 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 68344 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 K12HD01097 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.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is
REFERENCES


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

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

**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 $p<10^{-5}$.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>gene</th>
<th>N</th>
<th>RS1</th>
<th>RS3</th>
<th>TUK</th>
<th>FHS</th>
<th>GARP</th>
<th>RS1</th>
<th>RS3</th>
<th>TUK</th>
<th>FHS</th>
<th>GARP</th>
<th>Eff All</th>
<th>BETA</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs926849</td>
<td>6</td>
<td>PARK2</td>
<td>3939</td>
<td>0.31</td>
<td>0.32</td>
<td>NA</td>
<td>0.31</td>
<td>0.23</td>
<td>0.98</td>
<td>0.99</td>
<td>NA</td>
<td>0.9</td>
<td>95.04</td>
<td>C</td>
<td>-0.12</td>
<td>0.023</td>
<td>9.50E-08</td>
</tr>
<tr>
<td>rs17034687</td>
<td>3</td>
<td>NA</td>
<td>4429</td>
<td>0.09</td>
<td>0.1</td>
<td>0.05</td>
<td>0.08</td>
<td>NA</td>
<td>0.92</td>
<td>0.93</td>
<td>0.88</td>
<td>0.82</td>
<td>NA</td>
<td>C</td>
<td>0.19</td>
<td>0.037</td>
<td>3.67E-07</td>
</tr>
<tr>
<td>rs2187689</td>
<td>6</td>
<td>NA</td>
<td>4457</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.09</td>
<td>NA</td>
<td>0.98</td>
<td>0.99</td>
<td>0.95</td>
<td>0.93</td>
<td>NA</td>
<td>C</td>
<td>0.19</td>
<td>0.039</td>
<td>6.63E-07</td>
</tr>
<tr>
<td>rs7767277</td>
<td>6</td>
<td>NA</td>
<td>4457</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.09</td>
<td>NA</td>
<td>0.98</td>
<td>0.99</td>
<td>0.95</td>
<td>0.93</td>
<td>NA</td>
<td>A</td>
<td>0.19</td>
<td>0.039</td>
<td>6.64E-07</td>
</tr>
<tr>
<td>rs9488238</td>
<td>6</td>
<td>NA</td>
<td>4464</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
<td>NA</td>
<td>0.87</td>
<td>0.88</td>
<td>0.87</td>
<td>0.97</td>
<td>NA</td>
<td>A</td>
<td>-0.28</td>
<td>0.058</td>
<td>7.30E-07</td>
</tr>
<tr>
<td>rs3019449</td>
<td>6</td>
<td>PARK2</td>
<td>4636</td>
<td>0.32</td>
<td>0.32</td>
<td>0.31</td>
<td>0.31</td>
<td>0.28</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.99</td>
<td>97.18</td>
<td>A</td>
<td>-0.11</td>
<td>0.022</td>
<td>7.42E-07</td>
</tr>
<tr>
<td>rs9301951</td>
<td>13</td>
<td>GPC6</td>
<td>4397</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.05</td>
<td>NA</td>
<td>0.96</td>
<td>0.97</td>
<td>0.74</td>
<td>0.78</td>
<td>NA</td>
<td>C</td>
<td>-0.25</td>
<td>0.051</td>
<td>1.35E-06</td>
</tr>
<tr>
<td>rs11136566</td>
<td>8</td>
<td>CSMD1</td>
<td>4610</td>
<td>0.39</td>
<td>0.39</td>
<td>0.36</td>
<td>0.39</td>
<td>0.46</td>
<td>0.99</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>97.39</td>
<td>A</td>
<td>-0.1</td>
<td>0.021</td>
<td>2.24E-06</td>
</tr>
<tr>
<td>rs7744666</td>
<td>6</td>
<td>NA</td>
<td>4466</td>
<td>0.1</td>
<td>0.09</td>
<td>0.06</td>
<td>0.1</td>
<td>NA</td>
<td>0.99</td>
<td>1</td>
<td>0.97</td>
<td>0.96</td>
<td>NA</td>
<td>C</td>
<td>0.16</td>
<td>0.035</td>
<td>3.07E-06</td>
</tr>
<tr>
<td>rs11969002</td>
<td>6</td>
<td>NA</td>
<td>4466</td>
<td>0.1</td>
<td>0.09</td>
<td>0.06</td>
<td>0.1</td>
<td>NA</td>
<td>0.99</td>
<td>1</td>
<td>0.97</td>
<td>0.96</td>
<td>NA</td>
<td>A</td>
<td>0.16</td>
<td>0.035</td>
<td>3.07E-06</td>
</tr>
<tr>
<td>rs365962</td>
<td>16</td>
<td>NA</td>
<td>4608</td>
<td>0.44</td>
<td>0.44</td>
<td>0.48</td>
<td>0.47</td>
<td>0.42</td>
<td>0.99</td>
<td>0.99</td>
<td>0.96</td>
<td>0.69</td>
<td>98.61</td>
<td>C</td>
<td>-0.1</td>
<td>0.020</td>
<td>3.10E-06</td>
</tr>
<tr>
<td>rs387953</td>
<td>16</td>
<td>NA</td>
<td>4683</td>
<td>0.46</td>
<td>0.45</td>
<td>0.48</td>
<td>0.46</td>
<td>0.43</td>
<td>1</td>
<td>0.99</td>
<td>1</td>
<td>0.67</td>
<td>99.99</td>
<td>G</td>
<td>-0.09</td>
<td>0.020</td>
<td>3.37E-06</td>
</tr>
<tr>
<td>rs737631</td>
<td>6</td>
<td>PARK2</td>
<td>4303</td>
<td>0.28</td>
<td>0.29</td>
<td>0.23</td>
<td>0.28</td>
<td>NA</td>
<td>0.92</td>
<td>0.91</td>
<td>0.89</td>
<td>0.83</td>
<td>NA</td>
<td>A</td>
<td>-0.11</td>
<td>0.023</td>
<td>3.57E-06</td>
</tr>
<tr>
<td>rs4802666</td>
<td>19</td>
<td>MYH14</td>
<td>4094</td>
<td>0.27</td>
<td>0.27</td>
<td>0.18</td>
<td>0.27</td>
<td>NA</td>
<td>0.7</td>
<td>0.74</td>
<td>0.76</td>
<td>0.64</td>
<td>NA</td>
<td>A</td>
<td>-0.13</td>
<td>0.027</td>
<td>3.76E-06</td>
</tr>
<tr>
<td>rs1154053</td>
<td>8</td>
<td>NA</td>
<td>4639</td>
<td>0.17</td>
<td>0.17</td>
<td>0.18</td>
<td>0.16</td>
<td>0.2</td>
<td>0.99</td>
<td>1</td>
<td>0.96</td>
<td>0.89</td>
<td>96.17</td>
<td>C</td>
<td>-0.12</td>
<td>0.026</td>
<td>4.68E-06</td>
</tr>
<tr>
<td>rs1205863</td>
<td>6</td>
<td>NA</td>
<td>3939</td>
<td>0.06</td>
<td>0.06</td>
<td>NA</td>
<td>0.06</td>
<td>0.07</td>
<td>0.99</td>
<td>1</td>
<td>NA</td>
<td>0.99</td>
<td>98.23</td>
<td>G</td>
<td>0.21</td>
<td>0.045</td>
<td>4.77E-06</td>
</tr>
<tr>
<td>rs3749982</td>
<td>6</td>
<td>NA</td>
<td>4458</td>
<td>0.1</td>
<td>0.09</td>
<td>0.06</td>
<td>0.1</td>
<td>NA</td>
<td>0.99</td>
<td>1</td>
<td>0.96</td>
<td>0.96</td>
<td>NA</td>
<td>A</td>
<td>0.16</td>
<td>0.035</td>
<td>4.97E-06</td>
</tr>
<tr>
<td>rs6457690</td>
<td>6</td>
<td>NA</td>
<td>4464</td>
<td>0.1</td>
<td>0.09</td>
<td>0.07</td>
<td>0.11</td>
<td>NA</td>
<td>0.98</td>
<td>1</td>
<td>0.96</td>
<td>0.97</td>
<td>NA</td>
<td>A</td>
<td>0.16</td>
<td>0.035</td>
<td>5.93E-06</td>
</tr>
<tr>
<td>rs1029296</td>
<td>6</td>
<td>NA</td>
<td>4464</td>
<td>0.1</td>
<td>0.09</td>
<td>0.07</td>
<td>0.11</td>
<td>NA</td>
<td>0.98</td>
<td>1</td>
<td>0.96</td>
<td>0.97</td>
<td>NA</td>
<td>C</td>
<td>0.16</td>
<td>0.035</td>
<td>5.93E-06</td>
</tr>
<tr>
<td>rs4875102</td>
<td>8</td>
<td>NA</td>
<td>4608</td>
<td>0.26</td>
<td>0.26</td>
<td>0.27</td>
<td>0.25</td>
<td>0.29</td>
<td>0.99</td>
<td>0.99</td>
<td>0.95</td>
<td>0.91</td>
<td>97.65</td>
<td>A</td>
<td>-0.1</td>
<td>0.023</td>
<td>6.24E-06</td>
</tr>
<tr>
<td>rs7234567</td>
<td>18</td>
<td>NA</td>
<td>4682</td>
<td>0.26</td>
<td>0.26</td>
<td>0.22</td>
<td>0.22</td>
<td>0.29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.89</td>
<td>97.67</td>
<td>A</td>
<td>-0.1</td>
<td>0.023</td>
<td>6.34E-06</td>
</tr>
</tbody>
</table>
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 7.29E-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;

*R2 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. Manhatten 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.