



Lumbar disc degeneration is linked to a carbohydrate sulfotransferase 3 variant

You-Qiang Song,¹ Tatsuki Karasugi,^{2,3} Kenneth M.C. Cheung,⁴ Kazuhiro Chiba,⁵ Daniel W.H. Ho,¹ Atsushi Miyake,^{2,5} Patrick Y.P. Kao,¹ Kit Ling Sze,¹ Anita Yee,¹ Atsushi Takahashi,⁶ Yoshiharu Kawaguchi,⁷ Yasuo Mikami,⁸ Morio Matsumoto,⁵ Daisuke Togawa,⁹ Masahiro Kanayama,⁹ Dongquan Shi,^{10,11} Jin Dai,^{10,11} Qing Jiang,^{10,11} Chengai Wu,¹² Wei Tian,¹³ Na Wang,¹³ John C.Y. Leong,¹⁴ Keith D.K. Luk,⁴ Shea-ping Yip,¹⁵ Stacey S. Cherny,¹⁶ Junwen Wang,¹ Stefan Mundlos,¹⁷ Anthi Kelempisioti,¹⁸ Pasi J. Eskola,¹⁸ Minna Männikkö,¹⁸ Pirkka Mäkelä,¹⁹ Jaro Karppinen,^{20,21} Marjo-Riitta Järvelin,^{22,23} Paul F. O'Reilly,²³ Michiaki Kubo,²⁴ Tomoatsu Kimura,⁷ Toshikazu Kubo,⁸ Yoshiaki Toyama,⁵ Hiroshi Mizuta,³ Kathryn S.E. Cheah,¹ Tatsuhiko Tsunoda,²⁵ Pak-Chung Sham,¹⁶ Shiro Ikegawa,² and Danny Chan¹

¹Department of Biochemistry, The University of Hong Kong, Pokfulam, Hong Kong, People's Republic of China. ²Laboratory for Bone and Joint Diseases, Center for Genomic Medicine, RIKEN, Tokyo, Japan. ³Department of Orthopaedic and Neuro-Musculoskeletal Surgery, Faculty of Medical and Pharmaceutical Sciences, Kumamoto University, Kumamoto, Japan. ⁴Department of Orthopaedics and Traumatology, The University of Hong Kong, Pokfulam, Hong Kong, People's Republic of China. ⁵Department of Orthopaedic Surgery, School of Medicine, Keio University, Tokyo, Japan. ⁶Laboratory for Statistical Analysis, Center for Genomic Medicine, RIKEN, Tokyo, Japan. ⁷Department of Orthopaedic Surgery, Faculty of Medicine, University of Toyama, Toyama, Japan. ⁸Department of Orthopaedics, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan. ⁹Spine Center, Hakodate Central General Hospital, Hakodate, Japan. ¹⁰The Center of Diagnosis and Treatment for Joint Disease, Drum Tower Hospital Affiliated to Medical School of Nanjing University, Nanjing, People's Republic of China. ¹¹Laboratory for Bone and Joint Diseases, Model Animal Research Center, Nanjing University, Nanjing, People's Republic of China. ¹²Department of Orthopedics, Beijing Jishuitan Hospital, Beijing, People's Republic of China. ¹³Department of Spinal Surgery, Beijing Jishuitan Hospital, The 4th Clinical Medical College of Peking University, Beijing, People's Republic of China. ¹⁴Open University of Hong Kong, Hong Kong, People's Republic of China. ¹⁵Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China. ¹⁶Department of Psychiatry, The University of Hong Kong, Pokfulam, Hong Kong, People's Republic of China. ¹⁷Institute for Medical Genetics, Charité-Universitätsmedizin Berlin, Berlin, Germany. ¹⁸Oulu Center for Cell-Matrix Research, Biocenter and Department of Medical Biochemistry and Molecular Biology of Oulu, Oulu, Finland. ¹⁹Department of Surgery, University Hospital of Oulu, Oulu, Finland. ²⁰Finnish Institute of Occupational Health, Health and Work Ability, and Disability Prevention Centre, Oulu, Finland. ²¹Institute of Clinical Medicine, Department of Physical and Rehabilitation Medicine, and ²²Institute of Health Sciences, Public Health and General Practice, University of Oulu, Oulu, Finland. ²³Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, Faculty of Medicine, London, United Kingdom. ²⁴Laboratory for Statistical Analysis, Center for Genomic Medicine, RIKEN, Tokyo, Japan. ²⁵Laboratory for Medical Informatics, Center for Genomic Medicine, RIKEN, Yokohama, Japan.

Lumbar disc degeneration (LDD) is associated with both genetic and environmental factors and affects many people worldwide. A hallmark of LDD is loss of proteoglycan and water content in the nucleus pulposus of intervertebral discs. While some genetic determinants have been reported, the etiology of LDD is largely unknown. Here we report the findings from linkage and association studies on a total of 32,642 subjects consisting of 4,043 LDD cases and 28,599 control subjects. We identified carbohydrate sulfotransferase 3 (*CHST3*), an enzyme that catalyzes proteoglycan sulfation, as a susceptibility gene for LDD. The strongest genome-wide linkage peak encompassed *CHST3* from a Southern Chinese family-based data set, while a genome-wide association was observed at rs4148941 in the gene in a meta-analysis using multiethnic population cohorts. rs4148941 lies within a potential microRNA-513a-5p (miR-513a-5p) binding site. Interaction between miR-513a-5p and mRNA transcribed from the susceptibility allele (A allele) of rs4148941 was enhanced in vitro compared with transcripts from other alleles. Additionally, expression of *CHST3* mRNA was significantly reduced in the intervertebral disc cells of human subjects carrying the A allele of rs4148941. Together, our data provide new insights into the etiology of LDD, implicating an interplay between genetic risk factors and miRNA.

Introduction

Low back pain (LBP) is one of the most common symptoms of spinal abnormalities, with an annual point prevalence averaging 30% (1). It is a major factor affecting quality of life and has a significant social and economic impact worldwide (2). One cause of LBP is lumbar disc degeneration (LDD) (3–6), a subset of inter-

vertebral disc (IVD) degeneration (IDD; OMIM 603932), which includes lumbar disc herniation (LDH) and sciatica (7).

The etiology of LDD is not fully understood. While factors such as body weight, mechanical loading, physical activities, and smoking may have a role (8–11), familial predisposition (12, 13) and twin studies (6, 14–16) have demonstrated a significant genetic contribution, with heritability estimates as high as 74% (16).

LDD presents with a complex cascade of degenerative events. The current gold standard for phenotype analysis is an assessment from T2-weighted MRI of the spine, which provides key information on disc dehydration as an indicator of degeneration,

Authorship note: You-Qiang Song and Tatsuki Karasugi contributed equally to this work.

Conflict of interest: The authors have declared that no conflict of interest exists.

Citation for this article: *J Clin Invest.* 2013;123(11):4909–4917. doi:10.1172/JCI69277.



Table 1
Cohorts analyzed at different stages of gene discovery

Cohort	Population	Phenotype	Cases	Controls	Total	Stage of discovery	SNPs analyzed ^A
J1	Japanese	LDD with sciatica	366	3,331	3,697	GWAS stage 1	464,755
J2	Japanese	LDD with sciatica	544	15,800	16,344	GWAS stage 2	1,349
J3	Japanese	LDD with sciatica	242	622	864	GWAS stage 3	10
NC	Northern Chinese	LDD with sciatica	572	776	1,348	GWAS stage 3	10
SC-1S	Southern Chinese ^B	Severe LDD	270	271	541	GWAS stage 4	1
F1	Finnish	LDD with sciatica	281	393	674	GWAS stage 4	1
F2	Finnish	LDD with sciatica	118	4,642	4,760	GWAS stage 4	1
SC-1	Southern Chinese ^C	LDD	–	–	1,379	Fine-mapping 1st and 2nd stage	69
SC-2	Southern Chinese ^D	LDD	–	–	1,620	Fine-mapping 2nd stage	13
SC-1S + SC-2S	Southern Chinese ^E	LDD	600	600	1,200	Final stage	2
F3	Finnish	LDD with sciatica	420	215	635	Final stage	2
J1'	Japanese ^F	LDD with sciatica	366	888	1,254	Final stage	2
J2'	Japanese ^G	LDD with sciatica	544	433	977	Final stage	2

^ANumber passing quality-control threshold. ^BSubset of SC-1, representing the 20th percentile from the 2 ends of the normalized spectrum of the age-adjusted LDD scores. ^CFrom the first-stage recruitment between 1999–2006. ^DFrom the second-stage recruitment between 2007–2011. ^ECases and controls selected from the top and bottom 20th percentiles from a normalized spectrum of age-adjusted LDD score from 2,999 individuals. ^FSubset of the J1 cohort. ^GSubset of the J2 cohort.

based on the age-adjusted LDD score (20): individuals with age-adjusted LDD scores 0.5 SD above the mean value for their age group were assigned as cases; all others were assigned as controls. Nonparametric linkage (NPL) analysis of the data identified candidate regions on chromosomes 1, 5, 8, 10, and 20 (Figure 1A).

In the second-stage linkage analysis, we added 37 individuals from 8 new families (families 11–18; Supplemental Figure 1), with 19 new microsatellite markers added at 5-cM intervals around the candidate regions. Only markers in the chromosome 10 region showed a substantial increase in NPL Z-score, while markers in other chromosomal regions showed no change or a decrease in NPL Z-score (Figure 1A). The maximum NPL Z-score was 3.72 at 98.96 cM (D10S569) within the 10q21.2–q23.1 region (72.065–79.223 Mb on chromosome 10).

Case-control GWAS identified variants in CHST3 (in 10q21.2–q23.1) at $P < 5 \times 10^{-8}$. We performed case-control association studies using 7 independent cohorts (Table 1) in a GWAS composed of 4 stages, in which the number of SNPs was reduced and the sample size increased at each stage (Figure 1B). In stage 1, we examined a Japanese cohort (J1; 366 LDD cases, 3,331 controls) using the Human-Hap550v3 Genotyping BeadChip (Illumina). After data quality assessment, we performed the Cochran-Armitage trend test on these case-control data at 464,775 SNPs genome-wide. Principal component analysis found no strong genetic heterogeneity (Supplemental Figure 2A), and the genomic control inflation factor (λ_{GC}) was 1.03 (Supplemental Figure 2B), indicative of weak population substructure. We then selected the top 1,500 SNPs according to P value ($7.59 \times 10^{-7} \leq P \leq 3.53 \times 10^{-3}$) for stage 2, which consisted of genotyping in a second independent Japanese cohort (J2; 544 cases, 15,800 controls), and analyzed 1,349 SNPs after data quality assessment. However, none of the SNPs achieved genome-wide significance after meta-analysis using Mantel-Haenszel combined minimum P . Therefore, we took the top 10 SNPs according to P value to stage 3, which included an additional Japanese cohort (J3; 242 cases, 622 controls) and a Northern Chinese cohort (NC; 572 cases, 776 controls). The smallest P value from stage 3 was for rs1245582 ($P = 1.46 \times 10^{-6}$; Table 2). While the association at rs1245582 was not genome-wide

significant, the SNP lies within the 10q21.2–q23.1 region identified by the linkage study, so we took this SNP forward for genotyping in stage 4.

Additional samples for stage 4 of the GWAS included a subset of the first Southern Chinese cohort (SC-1) containing individuals with severe LDD (SC-1S; 270 cases, 271 controls), and 2 Finnish cohorts (F1; 281 cases, 393 controls, ref. 27; and F2; 118 cases, 4,642 controls, ref. 28). Meta-analysis of the stage 4 data gave a genome-wide significant association with LDD at rs1245582 (odds ratio, 1.20; 95% CI, 1.13–1.29; $P = 3.73 \times 10^{-8}$). There was no evidence for heterogeneity in the stage 4 results ($P = 0.63$).

rs1245582 is located within an LD block of approximately 85 kb, according to the Japanese information in the International HapMap Project database (release 24) (Supplemental Figure 3A). Only 3 SNPs (rs1245582, rs751450, and rs4148917) in the region showed low P values ($P < 2.10 \times 10^{-4}$) in stage 2 of our study, and each is located within the LD block, which overlaps with *CHST3*. Carbohydrate sulphotransferase-3 catalyzes the transfer of sulphate from 3'-phosphoadenosine 5'-phosphosulphate (PAPS) to position 6 of the N-acetylgalactosamine (GalNAc) residue of chondroitin (29). The genome-wide significant association observed here was supported by our family linkage study, which found *CHST3* within the strongest linkage peak genome-wide.

Candidate causal variants in the 3' untranslated region (3'UTR) of CHST3 identified from fine-mapping study. To look for candidate causal variants within *CHST3*, we performed a 3-stage fine-mapping study (Figure 1B). In stage 1, we analyzed all 69 SNPs identified by the Japanese HapMap Phase II resource across *CHST3* and the relevant LD block (Supplemental Figure 3A) using our SC-1 cohort ($n = 1,379$; Supplemental Table 2 and ref. 20). Similar to the Japanese HapMap data, the LD map of *CHST3* from this dense set of SNPs in the SC-1 cohort showed 1 large LD block (Supplemental Figure 3B). Association testing of the 1,379 SC-1 individuals using linear regression at the 69 SNPs identified 14 SNPs with $P < 0.05$ (Supplemental Table 2). These were taken forward to stage 2 for genotyping in the combined Southern Chinese cohort (SC-1 + SC-2; $n = 2,999$). After data quality control analysis, 13 SNPs were tested for association (Supplemental Table 3), which revealed a cluster

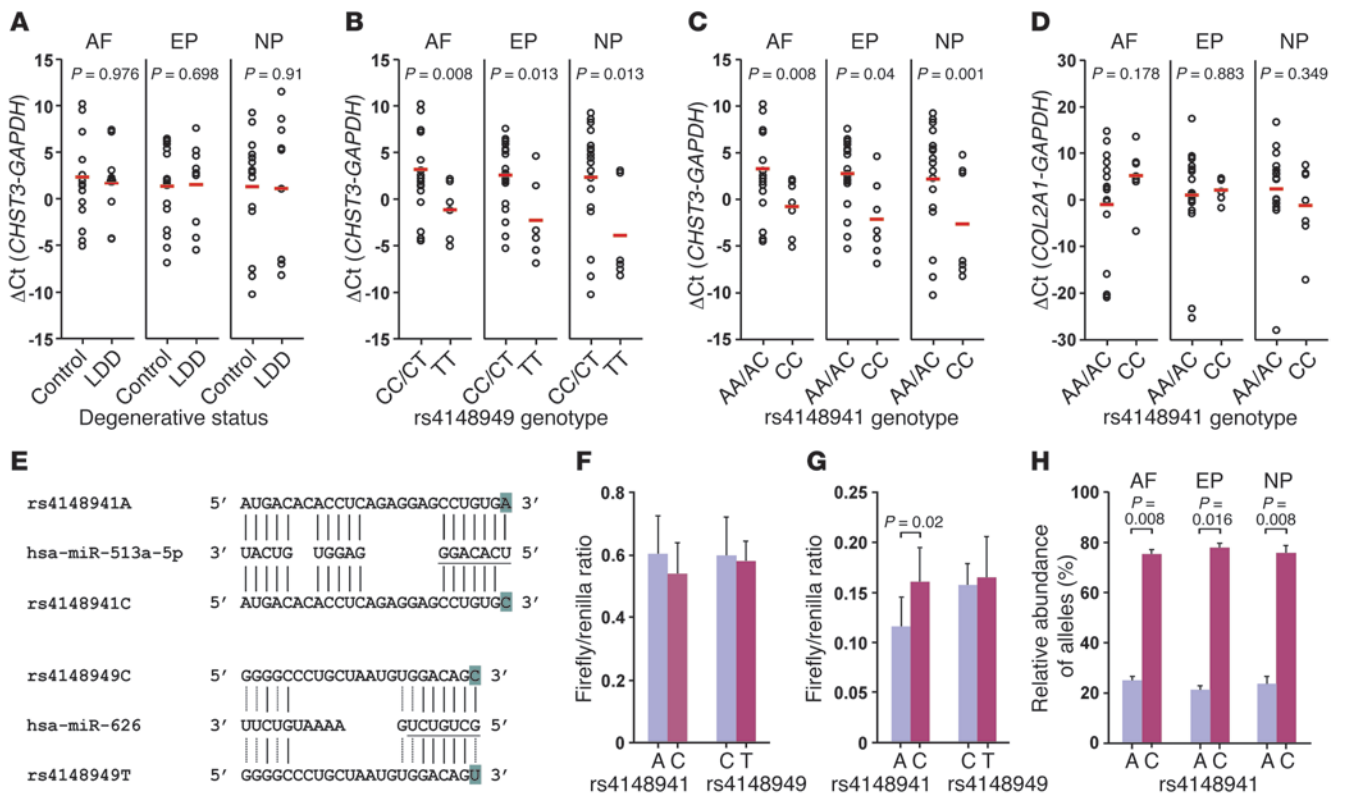


Figure 2

CHST3 expression in human IVDs and the effect of miRNA on allelic variants. (A–C) Comparison of *CHST3* expression by quantitative RT-PCR, (A) between control and degenerated samples and between samples differentiated according to genotypes of (B) CC/CT and TT for rs4148949 and (C) AA/AC and CC for rs4148941. (D) Comparison of *COL2A1* for samples differentiated according to genotypes of AA/AC and CC for rs4148941. (A–D) Levels are expressed as ΔC_t values above *GAPDH* (endogenous housekeeping gene for mRNA). Data are presented as dot plots; red bars denote mean values. (E) Predicted alignment of allelic variation at rs4148941 and rs4148949 as recognition sites for the binding of miR-513a-5p and miR-626, respectively. Both SNPs (shaded) occurred in the 7-bp seed sequence of complementarity (underlined) at the 5' end of the miRNAs. (F and G) Luciferase reporter assays were performed using C2812 cells to determine the effect of the A/C allele at rs4148941 and C/T allele at rs4148949 in the absence (F) and presence (G) of miR-513a-5p and miR-626, respectively. (H) Relative allelic expression difference of *CHST3* mRNA, determined by pyrosequencing for human IVD tissues with heterozygous genotype at rs4148941. Data are mean \pm SD. AF, annulus fibrosus; EP, cartilage end-plate.

of 8 SNPs at the 3' region of the gene with elevated association signals; these were contained within an LD block (rs6480592 to rs1245582) (Figure 1C) with pairwise correlation ($r^2 > 0.49$). Of these SNPs, 5 were within the 3'UTR of *CHST3*, while the others were intronic or downstream of the gene, which suggests that the causal variant may be in the 3'UTR.

We hypothesized the involvement of miRNA binding sites and performed an analysis using the PolymiRTS database (30). We found that rs4148941 and rs4148949 were within predicted miRNA binding sites for miR-513a-5p (also known as miR-513) and miR-626, respectively. With these SNPs as the best candidate causal variants, we performed replication studies using cohorts from Japan and Finland (J1', J2', and F3; Table 1). This was followed by a final meta-analysis using only severely affected individuals (top 20th percentile) from the SC-1S + SC-2S cohorts as a more appropriate subgroup, with both rs4148941 and rs4148949 achieving genome-wide significance using an allelic model (Table 3 and Figure 1C). Analysis using a dominant model gave similar *P* values, although not genome-wide significance, with a Bonfer-

roni corrected threshold set at 1.07×10^{-7} for a total of 467,709 SNPs analyzed throughout the study.

Allelic products of CHST3 are expressed differentially in IVD tissues. We examined *CHST3* expression in various human tissues using quantitative RT-PCR and detected high, specific *CHST3* expression in IVD tissues, bone, and cartilage (Supplemental Figure 4). Comparison of *CHST3* expression in normal and degenerative IVD tissues found similar expression in the NP, annulus fibrosus, and cartilage end-plates (Figure 2A).

We then genotyped the tissue samples, and *CHST3* mRNA levels were tested against rs4148941 and rs4148949 under both additive and dominant models. Differential expression was significant under the dominant model, but not the additive model. Individuals with the risk genotypes for rs4148941 (AA/AC) and rs4148949 (CC/CT) had significantly lower levels of *CHST3* mRNA than did individuals with the CC and TT genotypes, respectively (Figure 2, B and C), which suggests the risk alleles directly affect *CHST3* mRNA levels. We detected no significant difference in expression of *COL2A1*, another gene expressed in disc cells, with respect to the genotypes for rs4148941 (Figure 2D), which suggests the effect is specific to *CHST3*.



Table 2
Association of the top 10 SNPs selected from the combined analysis of GWAS data with LDD

SNP	Chr	Nearest gene	Models	Risk allele frequency (cases/controls)	J1 P	J2 P	J1 + J2 ^A P _{het}	J3 P	NC P	J1 + J2 + J3 + NC ^A P _{het}	P
rs2646143	2	LOC1-13311-9	A	0.81/0.78	2.73 × 10 ⁻⁴	7.57 × 10 ⁻³	1.84 × 10 ⁻¹	2.02 × 10 ⁻¹	9.21 × 10 ⁻¹	4.86 × 10 ⁻²	2.54 × 10 ⁻⁴
rs1245582	10	CHST3	A	0.50/0.44	5.80 × 10 ⁻⁵	2.38 × 10 ⁻²	8.21 × 10 ⁻²	1.77 × 10 ⁻¹	1.03 × 10 ⁻¹	2.94 × 10 ⁻¹	1.46 × 10 ⁻⁶
rs17344217	8	CSMD1	R	0.23/0.19	4.58 × 10 ⁻⁴	1.00 × 10 ⁻²	2.60 × 10 ⁻¹	7.51 × 10 ⁻¹	5.05 × 10 ⁻¹	1.44 × 10 ⁻⁴	1.44 × 10 ⁻⁴
rs973527	7	AMPH	D	0.11/0.09	4.33 × 10 ⁻⁵	2.90 × 10 ⁻²	1.38 × 10 ⁻¹	5.95 × 10 ⁻¹	4.16 × 10 ⁻¹	9.75 × 10 ⁻³	2.14 × 10 ⁻⁴
rs613516	3	RBMS3	A	0.61/0.58	8.28 × 10 ⁻⁴	9.62 × 10 ⁻³	2.64 × 10 ⁻¹	1.19 × 10 ⁻¹	1.03 × 10 ⁻¹	1.35 × 10 ⁻²	8.26 × 10 ⁻⁵
rs6459154	6	COL21A1	R	0.90/0.87	2.90 × 10 ⁻⁴	1.60 × 10 ⁻²	1.59 × 10 ⁻¹	6.71 × 10 ⁻¹	1.66 × 10 ⁻¹	2.73 × 10 ⁻¹	2.15 × 10 ⁻⁵
rs9689200	6	THBS2	A	0.46/0.44	1.20 × 10 ⁻³	6.99 × 10 ⁻³	3.87 × 10 ⁻¹	4.61 × 10 ⁻¹	5.48 × 10 ⁻¹	9.28 × 10 ⁻³	9.98 × 10 ⁻⁴
rs874840	2	DIRC3	A	0.42/0.38	2.30 × 10 ⁻³	5.31 × 10 ⁻³	5.05 × 10 ⁻¹	8.99 × 10 ⁻¹	3.04 × 10 ⁻¹	1.82 × 10 ⁻¹	3.92 × 10 ⁻⁴
rs7810842	7	TWISTNB	D	0.16/0.13	2.16 × 10 ⁻⁴	2.30 × 10 ⁻²	1.57 × 10 ⁻¹	2.07 × 10 ⁻¹	2.14 × 10 ⁻¹	3.89 × 10 ⁻¹	1.75 × 10 ⁻⁵
rs1109255	1	HIVEP3	A	0.42/0.37	1.03 × 10 ⁻³	7.28 × 10 ⁻³	4.21 × 10 ⁻¹	9.02 × 10 ⁻¹	6.21 × 10 ⁻¹	1.48 × 10 ⁻¹	5.45 × 10 ⁻⁴

^AMeta-analysis. A, allelic; R, recessive; D, dominant.

rs4148941 affects the interaction with miR-513a-5p, reducing CHST3 mRNA levels. rs4148941 and rs4148949 are within potential binding sites for the seed region of miR-513a-5p and miR-626, respectively (Figure 2E). We hypothesized that the risk allele would function through interaction with miR-513a-5p or miR-626. We therefore tested the effect of the CHST3 3'UTR sequence containing rs4148941-A/C or rs148949-C/T alleles using a luciferase reporter system (Figure 2F) in an immortalized human chondrocyte cell line, C28I2. We detected little or no expression of miR-513a-5p or miR-626 in C28I2 cells (data not shown), and C28I2 cells transfected with the rs4148941-A/C or rs1418949-C/T reporter constructs showed no differences in luciferase activity (Figure 2F). When we repeated the experiments with exogenous miR-513a-5p, the luciferase activity for the rs4148941-A reporter was 27% less than that in cells transfected with the counterpart rs4148941-C reporter ($P = 0.02$; Figure 2G). In contrast, cells transfected with either the rs4148949-C or -T reporter in the presence of miR-626 showed no significant difference in luciferase activity. This is consistent with mRNAs transcribed from the risk allele for rs4148941 being less stable.

To assess whether this allelic difference occurs in human IVD samples, we quantified CHST3 allelic-specific transcripts in individuals with the AC genotype for rs4148941 by pyrosequencing. There was a significant difference between the relative levels of the A and C allelic products as a percentage of total CHST3 mRNA for all tissues of the IVD (Figure 2H). These data support rs4148941-A being the functional risk allele and causing a preferential reduction in CHST3 mRNA. Finally, we showed that miR-513a-5p was expressed in the IVD tissues that could influence the level of CHST3 mRNA. There was no significant difference in the expression level of miR-513a-5p between control and LDD samples (Supplemental Figure 5A), nor according to the risk genotype (Supplemental Figure 5B), which suggests that the allelic difference is primarily the consequence of the interaction between miR-513a-5p and the risk allele of CHST3.

Heterozygous individuals with mutations in CHST3 showed early indication of disc abnormalities consistent with degeneration. Rare mutations in CHST3 that disrupt its enzymatic activity have been reported in patients with recessive skeletal abnormalities, including spondyloepiphyseal dysplasia Omani type, Larsen syndrome, humero-spinal dysostosis, and chondrodysplasia with multiple dislocation (31–36). Despite the different diagnostic labels, patients with CHST3 mutations have similar clinical characteristics and can be generally classified as spondyloepiphyseal dysplasia with congenital joint dislocation and vertebral changes as the principal features (OMIM 603799). By late childhood, these features manifest and lead to arthritis of the hips and spine with IVD degeneration, rigid kyphoscoliosis, and trunk shortening.

Numerous mutations have been identified, including missense, nonsense, small insertions, and deletions. Biochemical analysis showed the majority of these mutations result in loss-of-function or severe reduction in the enzymatic activity (31, 33–36). Recessive inheritance would be consistent with most rare mutations that affect enzymes. While detailed clinical information for the heterozygous parents was not available, it is conceivable that a reduction of 50% or less in CHST3 activity causes a milder phenotype that may have preferential defects of the spine, leading to early-onset disc degeneration. We obtained and reviewed MRI scans of 3 parents with heterozygous missense mutations in CHST3, 2 of Chinese ethnicity (35) and 1 from Oman (33). Each showed evi-



Table 3
Meta-analysis of the 2 most significant SNPs of *CHST3*

Cohort ^A	Genotype count ^B		a ₁ frequency (%)		Allelic model (a ₁ vs. a ₂)		Dominant model (a ₁ a ₁ , a ₁ a ₂ vs. a ₂ a ₂)	
	Cases	Controls	Cases	Controls	Odds ratio (95% CI)	P	Odds ratio (95% CI)	P
rs4148941^C								
SC-1S + SC-2S	102/293/181	77/240/230	43.14	36.01	1.35 (1.14–1.60)	5.58 × 10 ⁻⁴	1.58 (1.24–2.02)	2.29 × 10 ⁻⁴
J1'	93/184/89	162/444/280	50.55	43.34	1.34 (1.12–1.59)	9.89 × 10 ⁻⁴	1.44 (1.09–1.90)	1.03 × 10 ⁻²
J2'	118/274/152	74/202/157	46.86	40.42	1.30 (1.09–1.56)	4.27 × 10 ⁻³	1.47 (1.12–1.92)	5.58 × 10 ⁻³
F3	52/79/29	26/45/17	57.19	55.11	1.09 (0.75–1.58)	6.56 × 10 ⁻¹	1.08 (0.56–2.10)	8.17 × 10 ⁻¹
Meta-analysis					1.31 (1.19–1.44)	4.50 × 10 ⁻⁸	1.48 (1.27–1.71)	2.40 × 10 ⁻⁷
rs4148949^D								
SC-1S + SC-2S	103/289/181	78/241/230	43.19	36.16	1.34 (1.13–1.59)	6.62 × 10 ⁻⁴	1.56 (1.22–1.99)	3.53 × 10 ⁻⁴
J1'	93/184/89	166/442/279	50.55	43.63	1.32 (1.11–1.57)	1.57 × 10 ⁻³	1.43 (1.08–1.89)	1.19 × 10 ⁻²
J2'	120/272/152	75/201/157	47.06	40.53	1.30 (1.09–1.56)	3.90 × 10 ⁻³	1.47 (1.12–1.92)	5.58 × 10 ⁻³
F3	124/209/55	63/102/37	58.89	56.44	1.11 (0.87–1.41)	4.17 × 10 ⁻¹	1.36 (0.86–2.14)	1.89 × 10 ⁻¹
Meta-analysis					1.29 (1.17–1.41)	8.40 × 10 ⁻⁸	1.48 (1.28–1.71)	1.14 × 10 ⁻⁷

^ASee Table 1. ^BData presented as a₁a₁/a₁a₂/a₂a₂. ^Ca₁a₂, A/C. ^Da₁a₂, C/T.

dence of IVD degeneration and associated disc abnormalities, such as herniation (Schmorl's nodes) of the cartilaginous end-plate, in the lumbar region (Figure 3). With respect to our Southern Chinese population cohort, their LDD scores with age adjustment were considered to be mild for their age group (30–35 years old; ref. 20). They would therefore be considered to have earlier onset of disc degeneration resulting from reduced *CHST3* activity. This is consistent with our hypothesis that reduced expression of *CHST3* mRNA is a risk factor for LDD.

Discussion

The phenotype definition for LDD is complex. In the present study, because LDH is caused by disc degeneration or other traumatic events in the lumbar region of the spine, we included it as a subset of LDD. Therefore, when performing the meta-analysis, we treated all cohorts as a single population with common heritable factors. Diagnostic heterogeneity in the case samples is unlikely, because heterogeneity tests were not significant. The fact that independent linkage and GWAS studies homed in on the same region in chromosome 10 provides strong evidence for a susceptibility locus for LDD in the region.

We adopted the strategy of initially searching for linkage and association loci by means of genome-wide approaches in more severe cases that may have a larger effect on the disease, followed by association analysis to fine-tune regions within significant linkage/association signals. Complex diseases often involve multiple genetic variants, each with small or moderate effect, reducing the power of linkage analysis compared with association studies (37). Our selection of early-onset cases with disc degeneration was aimed at reducing the heterogeneity, working on the hypothesis that some variants are common within these families. Furthermore, the relatively high heritability from familial aggregation and twin studies supports the notion that useful information can be obtained from linkage analysis for early-onset families. Our 2-stage approach enabled the filtering of potential false positives and provided a candidate region within chromosome 10.

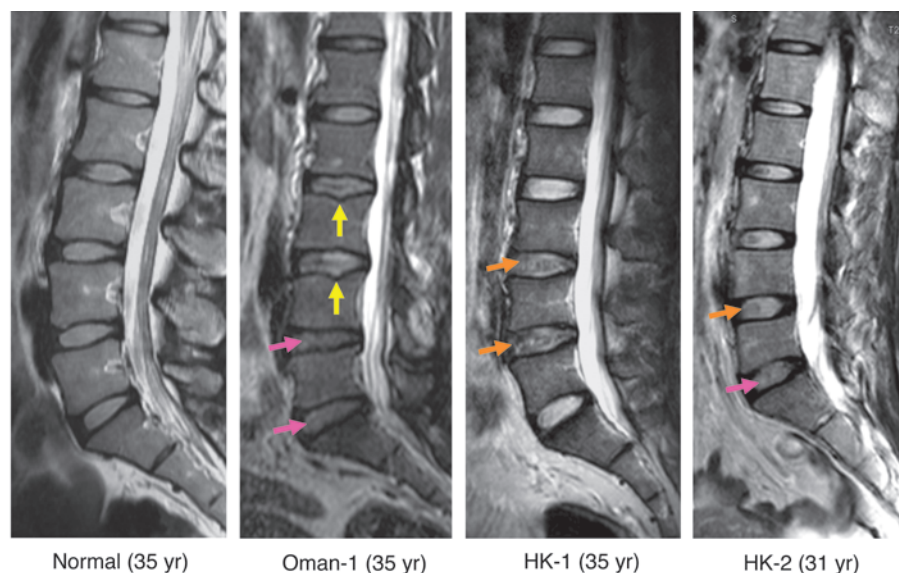
Using a GWAS comprising 4 stages allowed the identification of a SNP (rs1245582) closely associated with LDD, with *CHST3* as the nearest gene. This approach minimized the possibility of false-pos-

itive associations due to population stratification in stage 1 and reduction in number of SNPs required for genotyping in stage 2. Although there was an age difference between case and control groups in stages 1 and 2 (data not shown), it was not a confounding factor, as association remains significant after adjustment for age, gender, and BMI using the logistic regression model. In stages 3 and 4, while rs1245582 did not show significant association except in the Chinese study of replication 2, when we set the significance threshold to $P < 0.05$, the risk allele frequency in each of the cases and controls showed a common trend throughout all phases of the association studies. Importantly, this multistep association study and meta-analysis provided the first evidence associating LDD with rs1245582 by an unbiased approach. The fact that rs1245582 was within a large LD block that contained only *CHST3* provided us with a match in the gene list from our linkage analysis and the confidence to carry out a detailed association study that substantiated *CHST3* as a novel risk factor with genome-wide statistical significance.

CHST3 plays a key role in maintaining the hydration and function of cartilaginous tissues. Aggrecan is the most abundant proteoglycan in the disc matrix and contains abundant chondroitin sulphate glycosaminoglycan side chains. Proper sulfation of glycosaminoglycan side chains is critical for water retention within the disc to maintain proper osmotic pressure to resist compressive forces. Thus, the activity of *CHST3* would be important for IVD function.

CHST3 mutations result in recessive forms of spondyloepiphyseal dysplasia with congenital joint dislocation and vertebral changes as the principal features (OMIM 603799). Our finding that 3 heterozygous parents with previously known mutations in *CHST3* that disrupt its enzymatic activity showed disc abnormalities is consistent with earlier onset of degeneration in the corresponding age group, providing further support for *CHST3* as a genetic risk factor for disc degeneration.

Although we did not find a significant difference in the level of *CHST3* mRNA in IVD tissues from control and LDD patients, this could be due to the limited number of samples and consequent lack of statistical power or other unknown mechanisms. Clarification will require a much larger set of IVD samples. However, we did find evidence supporting rs4148941-A as a functional risk allele

**Figure 3**

MRI of heterozygous individuals with known mutations in *CHST3*. T2 MRI of a normal individual and parents of patients with recessive spondyloepiphyseal dysplasia (OMIM 603799) from Oman (Oman-1) and Hong Kong (HK-1 and HK-2). Disc degeneration with reduced NP intensity (pink arrows), abnormal end-plate (yellow arrows) consistent with Schmorl's nodes, and irregular darkened NP signals (orange arrows) were observed.

that reduces *CHST3* mRNA level. We speculate that mild *CHST3* reduction caused by the susceptibility SNP could result in disc degeneration in adults in conjunction with other risk factors.

A recent publication has demonstrated that miRNAs are likely to be involved in the degenerative process of IVD, with several either up- or downregulated in cells isolated from degenerated tissues (38). Although miR-513a-5p was not among the reported miRNAs with changed expression levels (38), we found that it interacted with certain alleles of *CHST3*, reducing its expression. miR-513a-5p also has a negative effect on the mRNA level of B7-H1 (also known as CD274 or programmed death receptor-1 [PD-1] ligand-1) in biliary epithelial cells (cholangiocytes) (39). B7-H1 is a key member of the B7 costimulator family with important regulatory functions in cell-mediated immune responses (40). miR-513a-5p binds to a site in the B7-H1 3'UTR, resulting in translational repression (39). The expression of miR-513a-5p itself can be downregulated by IFN- γ (39), a proinflammatory cytokine that can mediate many cellular events in cholangiocytes (41). Thus, it is possible that there is a dynamic interaction among miR-513a-5p and other miRNAs, regulated by proinflammatory cytokines in the degenerative process, that cannot be captured readily in disc tissue samples. The expression of miR-513a-5p in response to various cytokines needs to be studied carefully in disc cells or IVD explants in bioreactors.

In summary, the present study represents the largest systematic investigation thus far of the genetic risk factors for LDD. Using a combination of genome-wide linkage analysis and association studies, we identified *CHST3* as a novel risk factor, emphasizing the importance of *CHST3* as a musculoskeletal disease gene, and the involvement of miRNA, shedding light on the molecular pathogenesis of LDD.

Methods

Study populations. Case-control association studies were carried out using multiple cohorts from 3 different populations consisting of Chinese, Japanese, and Finnish subjects (see Supplemental Table 1 for composition and phenotype inclusion criteria). All recruited subjects in the Southern Chinese cohort underwent MRI scanning of the lumbar spine. LDD was diagnosed on the basis of signal intensity changes within the NP of the IVDs

of the lumbar spine and graded using Schneiderman classification (42) for signal intensity within the NP (43). Because age is a confounding factor in LDD, we adjusted for this using a sliding-window method (20, 27). In brief, the LDD score based on summation of the Schneiderman score in the lumbar region was normalized by logarithmic transformation. The age band for each individual was defined as age \pm 5 years. The age-adjusted LDD score for each individual was calculated by subtracting the mean and then dividing by the SD in the individual's age band.

The NC, J1, J2, J3, F1 (27), and F3 (23) cohorts were recruited on the basis of LDH characterized by sciatica or LBP requiring surgical treatment, but not necessarily having had surgical treatment. The F2 cohort (28) consisted of individuals from the Northern Finland Birth Cohort 1966 (NFBC 66) who were hospitalized owing to sciatica. 18 families with early-onset LDD identified from recruitment of the Southern Chinese cohort were used for whole-genome linkage analysis (Supplemental Figure 1). See Supplemental Methods for details of the sample sets.

Laboratory methods. PRISM human linkage mapping set v2.5-MD10 (Applied Biosystems) was used for the linkage analysis. Illumina Human-Hap550v3 Genotyping BeadChip was used for stage 1 of the GWAS. For genotyping of specific SNPs, multiple methods were used, depending on cohort origin (see Supplemental Methods).

Quantitative assessment of gene expression was performed using quantitative RT-PCR on total RNA extracted from disc tissues collected postoperatively with informed consent, while pyrosequencing was used to determine the relative allelic mRNA products. Expression of specific miRNA was determined using TaqMan miRNA assays (Applied Biosystems). miRNA binding prediction of SNP variants was performed using the PolymiRTS database (30). Luciferase assays were carried out in transfected C2812 chondrocytes (provided by M.B. Goldring, Hospital for Special Surgery, New York, New York, USA). See Supplemental Methods and Supplemental Table 4 for details of gene and miRNA expression analyses, reporter construct generation, cell transfection, luciferase assay conditions, and primer sequences.

Statistics. Multipoint NPL analysis (44) was performed using the S_{all} scoring function of MERLIN (45); NPL Z-scores and corresponding *P* values are reported. For quantitative-trait linkage analysis, the Deviates method of MERLIN was used to test for excess sharing among individuals in the same tail of trait distribution, as it makes no assumptions regarding trait



distribution and is implemented in the general framework of Whittemore and Halpern (44) and Kong and Cox (46). Thus, the Deviates method was employed when using age-adjusted and raw LDD scores. Association and Hardy-Weinberg equilibrium analyses were assessed using the χ^2 test. Odds ratios and other statistical measures were calculated using SPSS 15.0. HapMap data were downloaded from <http://www.hapmap.org/index.html>. Linkage disequilibrium blocks were estimated, and Hardy-Weinberg equilibrium was tested using Haploview v4.2 software (47). GraphPad Prism 4 (GraphPad Software) was used for data presentation. Meta-analyses were carried out using R (R Foundation for Statistical Computing) and PLINK (48) using fixed-effect model analyses at both the GWAS and the fine-mapping stages. Multiple testing using the Bonferroni correction was used in all statistical analyses. A *P* value less than 0.05 was considered significant.

Study approval. The local ethics committees from the relevant institutions approved the study designs. Written informed consent was obtained from all participants in the study.

Acknowledgments

Supported by the University Grants Council of Hong Kong (AoE04/04), AOSPINE (AOSBRC-07-2), by Research Grants Council of Hong Kong (HKU7509/03M), by a grant-in-aid from the Ministry of Education, Culture, Sports, and Science of Japan (grant no.

21249080), and by the Academy of Finland (grant nos. 121620 and 129504). The authors thank the individuals and families who participated in this study as well as Walter Boldmer, Peter Goodfellow, Dong Yan Jin, and Kin Hang Kok for study design; Florence Mok for phenotyping; Pei Yu for data entry and patient recruitment; Sandy C.S. Poon for technical assistance; and Jockey Club MRI Centre (LKS Faculty of Medicine, The University of Hong Kong) and The Hong Kong Sanatorium and Hospital for MRI.

Received for publication February 13, 2013, and accepted in revised form August 8, 2013.

Address correspondence to: Danny Chan or You-Qiang Song, Department of Biochemistry, Faculty of Medicine Building, The University of Hong Kong, 21 Sassoon Road, Pokfulam, Hong Kong. Phone: 852.28199482; Fax: 852.28551254; E-mail: chand@hku.hk (D. Chan). Phone: 852.28199245; Fax: 852.28551254; E-mail: Songy@hku.hk (Y.-Q. Song). Or to: Shiro Ikegawa, Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Science, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan. Phone: 81.3.5449.5393; Fax: 81.3.5449.5393; E-mail: sikegawa@ims.u-tokyo.ac.jp.

1. Andersson GB. Epidemiological features of chronic low-back pain. *Lancet*. 1999;354(9178):581–585.
2. Brodke DS, Ritter SM. Nonsurgical management of low back pain and lumbar disk degeneration. *Instr Course Lect*. 2005;54:279–286.
3. Deyo RA, Weinstein JN. Low back pain. *N Engl J Med*. 2001;344(5):363–370.
4. Rannou F, Revel M, Poiraudou S. Is degenerative disk disease genetically determined? *Joint Bone Spine*. 2003;70(1):3–5.
5. Salminen JJ, Erkintalo MO, Pentti J, Oksanen A, Kormano MJ. Recurrent low back pain and early disc degeneration in the young. *Spine*. 1999; 24(13):1316–1321.
6. Videman T, et al. Intragenic polymorphisms of the vitamin D receptor gene associated with intervertebral disc degeneration. *Spine*. 1998; 23(23):2477–2485.
7. Luoma K, Riihimaki H, Luukkonen R, Raininko R, Viikari-Juntura E, Lamminen A. Low back pain in relation to lumbar disc degeneration. *Spine (Phila Pa 1976)*. 2000;25(4):487–492.
8. Kelsey JL, et al. Acute prolapsed lumbar intervertebral disc. An epidemiologic study with special reference to driving automobiles and cigarette smoking. *Spine (Phila Pa 1976)*. 1984;9(6):608–613.
9. Ishihara H, Tsuji H, Hirano N, Ohshima H, Terahata N. Effects of continuous quantitative vibration on rheologic and biological behaviors of the intervertebral disc. *Spine (Phila Pa 1976)*. 1992; 17(3 suppl):S7–S12.
10. Pope MH, Hansson TH. Vibration of the spine and low back pain. *Clin Orthop Relat Res*. 1992; (279):49–59.
11. Battie MC, et al. 1991 Volvo Award in clinical sciences. Smoking and lumbar intervertebral disc degeneration: an MRI study of identical twins. *Spine (Phila Pa 1976)*. 1991;16(9):1015–1021.
12. Matsui H, Kanamori M, Ishihara H, Yudoh K, Naruse Y, Tsuji H. Familial predisposition for lumbar degenerative disc disease. A case-control study. *Spine*. 1998;23(9):1029–1034.
13. Postacchini F, Lami R, Pugliese O. Familial predisposition to discogenic low-back pain. An epidemiologic and immunogenetic study. *Spine*. 1988; 13(12):1403–1406.
14. Battie MC, Haynor DR, Fisher LD, Gill K, Gibbons LE, Videman T. Similarities in degenerative findings on magnetic resonance images of the lumbar spines of identical twins. *J Bone Joint Surg Am*. 1995; 77(11):1662–1670.
15. Livshits G, et al. Lumbar disc degeneration and genetic factors are the main risk factors for low back pain in women: the UK Twin Spine Study. *Ann Rheum Dis*. 2011;70(10):1740–1745.
16. 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(2):366–372.
17. Battie MC, Videman T. Lumbar disc degeneration: epidemiology and genetics. *J Bone Joint Surg Am*. 2006; 88(suppl 2):3–9.
18. Kalb S, Martirosyan NL, Kalani MY, Broc GG, Theodore N. Genetics of the degenerated intervertebral disc. *World Neurosurg*. 2011;77(3–4):491–501.
19. Song YQ, Sham P, Cheung KMC, Chan D. Genetics of disc degeneration. *Curr Orthop*. 2008; 22:259–266.
20. Song YQ, et al. Association of the Asporin *D14* allele with lumbar disc degeneration in Asians. *Am J Hum Genet*. 2008;82(3):744–747.
21. Mio F, et al. A functional polymorphism in COL11A1, which encodes the alpha 1 chain of type XI collagen, is associated with susceptibility to lumbar disc herniation. *Am J Hum Genet*. 2007; 81(6):1271–1277.
22. Williams FM, et al. GDF5 single-nucleotide polymorphism rs143383 is associated with lumbar disc degeneration in Northern European women. *Arthritis Rheum*. 2011;63(3):708–712.
23. Karasugi T, et al. Association of the tag SNPs in the human SKT gene (K1AA1217) with lumbar disc herniation. *J Bone Miner Res*. 2009;24(9):1537–1543.
24. Hirose Y, et al. A functional polymorphism in THBS2 that affects alternative splicing and MMP binding is associated with lumbar-disc herniation. *Am J Hum Genet*. 2008;82(5):1122–1129.
25. Ioannidis JP, et al. Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol*. 2008;37(1):120–132.
26. Williams FM, et al. Novel genetic variants associated with lumbar disc degeneration in northern Europeans: a meta-analysis of 4600 subjects. *Ann Rheum Dis*. 2013;72(7):1141–1148.
27. Virtanen IM, et al. Phenotypic and population differences in the association between *CILP* and lumbar disc disease. *J Med Genet*. 2007;44(4):285–288.
28. Rivinoja AE, et al. Sports, smoking, and overweight during adolescence as predictors of sciatica in adulthood: a 28-year follow-up study of a birth cohort. *Am J Epidemiol*. 2011;173(8):890–897.
29. Tsutsumi K, Shimakawa H, Kitagawa H, Sugahara K. Functional expression and genomic structure of human chondroitin 6-sulfotransferase. *FEBS Lett*. 1998;441(2):235–241.
30. Bao L, et al. PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids Res*. 2007;35(Database issue):D51–D54.
31. Hermanns P, et al. Congenital joint dislocations caused by carbohydrate sulfotransferase 3 deficiency in recessive Larsen syndrome and humero-spinal dysostosis. *Am J Hum Genet*. 2008;82(6):1368–1374.
32. Rajab A, Kunze J, Mundlos S. Spondyloepiphyseal dysplasia Omani type: a new recessive type of SED with progressive spinal involvement. *Am J Med Genet A*. 2004;126A(4):413–419.
33. Thiele H, et al. Loss of chondroitin 6-O-sulfotransferase-1 function results in severe human chondro-dysplasia with progressive spinal involvement. *Proc Natl Acad Sci U S A*. 2004;101(27):10155–10160.
34. Tuysuz B, Mizumoto S, Sugahara K, Celebi A, Mundlos S, Turkmen S. Omani-type spondyloepiphyseal dysplasia with cardiac involvement caused by a missense mutation in *CHST3*. *Clin Genet*. 2009; 75(4):375–383.
35. Unger S, et al. Phenotypic features of carbohydrate sulfotransferase 3 (*CHST3*) deficiency in 24 patients: congenital dislocations and vertebral changes as principal diagnostic features. *Am J Med Genet A*. 2010;152A(10):2543–2549.
36. van Roij MH, et al. Spondyloepiphyseal dysplasia, Omani type: further definition of the phenotype. *Am J Med Genet A*. 2008;146A(18):2376–2384.
37. Hirschhorn JN. Genetic approaches to studying common diseases and complex traits. *Pediatr Res*. 2005; 57(5 pt 2):74R–77R.
38. Wang HQ, et al. Deregulated miR-155 promotes Fas-mediated apoptosis in human intervertebral disc degeneration by targeting *FADD* and *caspase-3*. *J Pathol*. 2011;225(2):232–242.
39. Gong AY, et al. MicroRNA-513 regulates B7-H1 translation and is involved in IFN-gamma-induced B7-H1 expression in cholangiocytes. *J Immunol*. 2009; 182(3):1325–1333.



40. Dong H, et al. Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. *Nat Med.* 2002;8(8):793–800.
41. Lazaridis KN, Strazzabosco M, LaRusso NF. The cholangiopathies: disorders of biliary epithelia. *Gastroenterology.* 2004;127(5):1565–1577.
42. Schneiderman G, Flannigan B, Kingston S, Thomas J, Dillin WH, Watkins RG. Magnetic resonance imaging in the diagnosis of disc degeneration: correlation with discography. *Spine.* 1987;12(3):276–281.
43. Jim JJ, et al. The TRP2 allele of COL9A2 is an age-dependent risk factor for the development and severity of intervertebral disc degeneration. *Spine.* 2005;30(24):2735–2742.
44. Whittemore AS, Halpern J. A class of tests for linkage using affected pedigree members. *Biometrics.* 1994;50(1):118–127.
45. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002;30(1):97–101.
46. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet.* 1997;61(5):1179–1188.
47. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2):263–265.
48. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–575.