A whole exome study identifies novel candidate genes for vertebral bone marrow signal changes (Modic changes)

Kraatari Minna MD [a,b,c], Skarp Sini MSc [a,b,d], Niinimäki Jaakko MD PhD [e,c] Karppinen Jaro MD PhD [a,c,f], Männikkö Minna PhD [a,b,d]

[a] Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Finland

[b] Center for Cell – Matrix Research, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Finland

[c] Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Finland

[d] Biocenter Oulu, University of Oulu, Finland

[e] Research Unit of Medical Imaging, Physics and Technology, Faculty of Medicine, University of Oulu, Finland

[f] Finnish Institute of Occupational Health, Oulu, Finland

Corresponding author: Minna Männikkö, PhD, Research Director, Northern Finland Birth Cohorts, Center for Life Course Health Research, Faculty of Medicine, P.O. Box 5000, 90014 University of Oulu, Finland, Tel +358-(0) 294 -485751, e-mail: minna.mannikko@oulu.fi

Abstract

Study Design. Family-based study

Objective. To identify rare genetic factors predisposing to Modic changes (MC).

Summary of Background Data. Lumbar disc degeneration (LDD) is one of the contributing factors behind low back pain (LBP). Lumbar MC visualized as bone marrow signal intensity changes on magnetic resonance imaging (MRI) represent a specific phenotype of LDD, which has stronger association with LBP than LDD without MC.

Methods. The study set consisted of two Finnish families: Family I included seven affected and four unaffected individuals and Family II eight affected and seven unaffected individuals. MC were evaluated in 26 individuals using MRI. Whole exome sequencing was used to identify alleles co-segregating with MC. ANNOVAR was used to carry out functional annotation of alleles and their frequencies were evaluated using 1000Genomes, Sequencing Initiative Suomi (SISu) and the Exome Aggregation Consortium (ExAC) databases.

Results. We identified predisposing genetic alleles for MC in two Finnish families. In each family only single allele co-segregated with MC. In Family I the observed allele was an insertion and deletion in the *HSPG2* gene, resulting in a premature termination codon. In Family II a single nucleotide polymorphism (rs61753465) in the *MAML1* gene was identified in all affected family members.

Conclusions. We have identified two novel candidate genes, *MAML1* and *HSPG2*, associating with MC. These genes are important in cartilage structure and joint cartilage maintenance. Our findings are novel among lumbar spine degenerative phenotypes.

Introduction

Low back pain (LBP) is the most debilitating disease globally [1] lumbar disc degeneration (LDD) being one of its major risk factors.[2] Lumbar disc consists of outer annulus fibrosus containing type I and type II collagens and inner nucleus pulposus containing proteoglycans responsible for maintenance of tissue hydration.[3] In classical twin studies LDD was found to be genetically influenced with heritability estimations ranging from 34% to 74%.[4, 5] In comparison, the heritability estimate for sciatica was 20.8%.[6] LDD is a common phenotype influenced by several genetic and environmental factors.[7] The genetic background of common multifactorial phenotypes is composed of both common and rare genetic components.[8] In many common diseases such as osteoarthritis and osteoporosis, the common alleles typically explain only modest portion of heritability.[9-11] Thus, it is thought that rare alleles (minor allele frequency (MAF) <0.01) and alleles with low frequency (0.01 < MAF < 0.05) could explain a substantial fraction of the heritability in common diseases.[8]

Modic changes (MC), pathological bone marrow signal changes adjacent to vertebral endplates, represent a distinct specific phenotype of LDD. They are visible only on magnetic resonance imaging (MRI) [12] and have been associated with LBP.[13]

The prevalence of MC in the general population ranges from 22% to 56%.[13] The heritability of MC is estimated to be 16-43%.[14] Based on candidate gene approach, interleukin 1-α (IL1A) and matrix metalloproteinase 3 (MMP3) polymorphisms may play a role in MC,[15, 16] but so far no genome-wide association studies or next-generation sequencing have been performed. We have earlier described 14 Finnish families with a history of LDD characterized by sciatica.[17] In order to identify rare predisposing variation behind MC and to find genetic factors contributing to the underlying molecular mechanisms, we performed whole exome sequencing, i.e. sequencing of the whole protein coding region of the human genome using high-throughput DNA sequencing methods in two of these families.

Materials and methods

Out of the 14 families with history of LDD characterized by sciatica,[17] two large families with several affected family members in at least two generations were chosen for MRI evaluation and exome sequencing analysis (Figure 1). The family members did not have other musculoskeletal abnormalities.

Family I consisted of 18 family members, 10 males and eight females, aged 35 to 76 (mean=56). Family II had total of 33 family members, 16 males and 17 females, aged 30 to 78 (mean=54). Family members were interviewed by telephone at the beginning of the study in 2001-2003 and again 2016. Clinical characteristics and family demographics are displayed in the supplementary Table S1. MC were evaluated in eleven family members from Family I and fifteen family members from Family II using lumbar spine MRI. Two observers (J.K and J.N.) coded all available scans. Rostral and caudal endplates of each of the five lumbar segments were evaluated. Modic changes were evaluated as Type 1, Type 1/2, Type 2, Type 2/3, and Type 3, as previously defined.[12, 18] The maximum depth of each MC was graded from all sagittal slices compared to height of the vertebral body as (0) no MC, (1) MC in the endplate, (2) MC less than 25%, (3) MC from 25 to 50% and (4) MC more than 50% of the height of the vertebral body. The horizontal length was assessed in three zones in the AP direction (anterior, midpoint and posterior lesions) as described in Määttä *et al* 2016.[19] The interobserver reliability was calculated for the presence of MC (kappa = 0.9). In Family I, seven family members had MC while four were unaffected. In Family II, seven were unaffected. MC were present in eight family members. MC types, sizes and locations of the affected family members are presented in Table 1.

Four affected individuals with MC from Family I and three from Family II and one healthy individual from both families were selected for whole exome sequencing (Figure 1). This study

was approved by the Ethics Committee of the Northern Ostrobothnia Hospital District and family members gave their informed written consent.

The whole exome sequence data was acquired through commercial service (BGI, Hong Kong, http://www.bgi.com/) including exome capture, alignment and allele calling. The exome capture was performed using the SureSelect 51M Capture Kit (Agilent Technologies), which targets 75 Mb of the human genome with > 350 000 exons. The exome targets were sequenced with the IlluminaHiSeq2000 100PE platform and the reads were aligned to the hg19 human reference genome. Variant calling was performed using Genome Analysis ToolKit (GATK).[20]

Alleles were processed using GATK to filter them according to quality parameters. We filtered out alleles according to threshold values which were for single nucleotide polymorphisms (SNP): the root mean square of the mapping quality < 40, Haplotype score > 13 and genotype quality < 20 and for small insertions and deletions (indels): quality by depth < 2.0, Phred-scaled p-value using Fisher's Exact Test > 200.0 and the u-based z-approximation from the Mann-Whitney Rank Sum Test for mapping qualities < -20.0.

Alleles shared by the affected family members were extracted and alleles which were found in the unaffected individuals or in the in-house exome set (N=71) or had minor allele frequency > 1% in 1000Genomes database were filtered out. The remaining rare and private alleles were annotated using ANNOVAR version 2013aug23 [21] to identify alleles having harmful pathogenicity estimations (SIFT, PolyPhen-2, MutationTaster) and alleles affecting active promoters or strong enhancer regions (ENCODE databases for chromHMM estimations for HSMM and GM12878 cells) (Table 2). All databases used were downloaded and all annotations were done according to the ANNOVAR manual.

The frequency for alleles co-segregating with MC in general Finnish population was obtained from the Sequencing Initiative Suomi (SISu) database (N=6118) (http://www.sisuproject.fi/)

and the Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org). MAF in European population was also obtained from the ExAC.

After the annotation alleles were validated using capillary sequencing with ABI3500xL Genetic Analyzer (Applied Biosystems) from all family members to find alleles co-segregating with MC.

Results

In each family only one rare and pathogenic allele co-segregated with MC. The c.4829_4831delAGCinsCACTGTG (NM_005529) allele leads to an insertion and deletion mutation p.Q1611delfsX20 (NP_005520.4) in the heparan sulfate proteoglycan 2 (*HSPG2)* gene co-segregated with MC in Family I with complete penetrance (Figure 1; Table 1). This mutation was considered deleterious by MutationTaster (Table 3). It locates in the exon 38 encoding the third domain of the protein and alters the open reading frame resulting in a premature termination codon.

The allele identified in Family II was a SNP, c.1657G>A (NM_014757.4) (rs61753465), causing glutamic acid to lysine change p.Glu553Lys (NP_055572.1) in the mastermind like transcriptional coactivator 1 (*MAML1*) gene (Table 2). The SNP was estimated to be probably damaging by PolyPhen-2 (Table 3) and was found from eight individuals with MC observed in MRI. The SNP was also carried by two unaffected family members aged 27 and 35 years (Figure 1).

Each identified allele was private to the respective families, meaning that the affected family members with HSPG2 mutation in Family I did not share any rare pathogenic alleles in the MAML1 gene. Similarly, affected members in Family II shared no rare pathogenic alleles in the HSPG2 gene. The exome alignment covered well both gene loci in both families, the

average coverage of the exome sequencing being 67x, supporting the findings that these families each have their own allele associating with the disease.

Discussion

We have identified plausible causative or predisposing alleles for MC in two families. Firstly, an insertion and deletion mutation in the *HSPG2* gene (p.Q1611delfsX20) that encodes a protein called perlecan with an important role in the structure and function of musculoskeletal tissues. Secondly, we identified a glutamate to lysine amino acid change (p.Glu553Lys) in the *MAML1* gene. MAML1 acts as a transcriptional co-activator in the Notch signaling pathway participating in the joint cartilage maintenance.[22]

Perlecan is a heparan sulfate proteoglycan expressed in mammalian cartilage and basement membranes.[23] *HSPG2* has 94 exons and the encoded protein consists of five domains.[24] While some mutations in *HSPG2* cause fatal recessive disorders, more subtle alleles can predispose to less severe phenotypes. Rare SNPs in *HSPG2* have been associated with another spinal disorder, idiopathic scoliosis (IS).[25] Mutations in *HSPG2* also cause rare autosomal recessive disorders, dyssegmental dysplasia Silverman-Handmaker type (DDSH) and Schwarzt-Jampel syndrome (SJS) [26, 27] (Supp. Figure S1).

DDSH is a fatal condition characterized by flat facial appearance, narrow thorax, short neck, short and bowed limbs, decreased joint mobility and atypical ossification centers of the vertebral bodies.[28] SJS is characterized by reduced stature, chondrodysplasia, and skeletal anomalies including kyphoscoliosis and flattened vertebral bodies.[29] Differences in SJS and DDSH phenotypes are proposed to result from the differences in the amount of functional perlecan in the extra cellular matrix.[26, 27] Here the affected individuals in Family I carried a novel heterozygous mutation p.Q1611delfsX20 in HSPG2 resulting in premature termination codon. Previously, a premature termination codon in perlecan has been shown to result to mRNA instability through nonsense mRNA-mediated decay.[26] Thus, the mutation in HSPG2 (p.Q1611delfsX20) is likely to affect the level of perlecan expression. The functioning allele may provide enough perlecan to protect against more severe phenotypes.

MAML1 forms complex with intracellular Notch receptors and is considered to be transcriptional co-activator in the Notch signaling pathway.[30] Notch signaling pathway regulates variety of biological functions including cartilage development and homeostasis.[31-34] Notch signaling is required during joint cartilage maintenance and loss of notch signaling in postnatal murine joints leads to osteoarthritis.[35-37] Osteoarthritis is, similarly to disc degeneration, characterized by degeneration of cartilage, joint space narrowing and formation of osteophytes.

MAML1 has been shown to have several functions independent of notch signaling in variety of biological processes [38-40] Study of maml1 knock out mice showed that the mouse homolog maml1 enhances the transcriptional activity of runx2 in Notch independent manner. The knock out mice were characterized with impaired chondrocyte maturation.[41] In humans, RUNX2 is a transcription factor essential in the differentiation and proliferation of osteoblasts and chondrocytes.[42] RUNX2 has been reported to be highly expressed in the degenerated discs [43] and is upregulated in discs with MC compared to discs with no MC.[44] Based on these studies, we hypothesize that *MAML1* could affect the disc structure via notch signaling pathway and/or by affecting RUNX2 activity. The p.Glu553Lys (MAML1) alters the protein structure as glutamic acid is replaced by lysine. Glutamic acid has an acidic and negatively charged side chain, whereas the one in lysine is basic and positively charged. The p.Glu553Lys (MAML1) is present also in two family members unaffected at the time of MRI (year 2014). These individuals were only 27 and 35 years of age at the time, however, and since age has been shown to be a risk factor for MC especially in the lower lumbar spine,[45, 46] a follow-up MRI of these two individuals is needed to fully confirm the role of this allele.

We acknowledge limitations that may have impacted our study. The sample size was relatively small and the role of these genes needs to be investigated. Each identified allele was private to one single family. This is not unexpected, however, since founder mutations i.e. mutations

descending from a single ancestor have been described for example in genes causing familial breast cancer.[47-51] Chondrocytic cell lines were not available for the predictions of the promoter and enhancer regions. Instead data from HSMM and GM12878 cell lines were used. The activity of the promoter and enhancer regions can vary between tissues or cell lines,[52] and this may have affected our results. It should also be noted that the presence of the MC (or disc degeneration) does not always correlate with symptoms.

In summary, we have identified two new promising candidate genes for MC: *MAML1* and *HSPG2*. The SNP p.Glu553Lys in the *MAML1* gene and the indel p.Q1611delfsX20 in the *HSPG2* gene co-segregated with MC in families. These genes have an important role in the maturation and proliferation of chondrocytes. Perlecan is a structural protein in the disc and MAML1 may affect the disc structure through Notch signaling and/or RUNX2 activation. The role of the genes needs to be studied further in other sample sets with data on MC. Our findings are novel in spinal degenerative phenotypes.

References

1. Murray CJ, Vos T, Lozano R *et al.* Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the global burden of disease study 2010. *Lancet.* 2013;380:2197-2223.

2. Livshits G, Popham M, Malkin I *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;70:1740-1745.

3. Urban J, Maroudas A. The chemistry of the intervertebral disc in relation to its physiological function. *Clinics in Rheumatic Diseases.* 1980;6:51-76.

4. Williams FM, Popham M, Sambrook PN *et al.* Progression of lumbar disc degeneration over a decade: A heritability study. *Ann Rheum Dis.* 2011;70:1203-1207.

5. Battie MC, Videman T, Gibbons LE *et al.* 1995 volvo award in clinical sciences. determinants of lumbar disc degeneration. A study relating lifetime exposures and magnetic resonance imaging findings in identical twins. *Spine.* 1995;20:2601-2612.

6. Heikkilä JK, Koskenvuo M, Heliovaara M *et al.* Genetic and environmental factors in sciatica. evidence from a nationwide panel of 9365 adult twin pairs. *Ann Med.* 1989;21:393-398.

7. Powell MC, Wilson M, Szypryt P *et al.* Prevalence of lumbar disc degeneration observed by magnetic resonance in symptomless women. *Lancet.* 1986;2:1366-1367.

8. Manolio TA, Collins FS, Cox NJ *et al*. Finding the missing heritability of complex diseases. *Nature.* 2009;461:747-753.

9. Estrada K, Styrkarsdottir U, Evangelou E *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet.* 2012;44:491-501.

10. arcOGEN Consortium, arcOGEN Collaborators, Zeggini E *et al*. Identification of new susceptibility loci for osteoarthritis (arcOGEN): A genome-wide association study. *Lancet.* 2012;380:815-823.

11. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007;447:661-678.

12. Modic MT, Steinberg PM, Ross JS *et al*. Degenerative disk disease: Assessment of changes in vertebral body marrow with MR imaging. *Radiology.* 1988;166:193-199.

13. Jensen TS, Karppinen J, Sorensen JS *et al.* Vertebral endplate signal changes (modic change): A systematic literature review of prevalence and association with non-specific low back pain. *European Spine Journal.* 2008;17:1407-1422..

14. Määttä JH, Kraatari M, Wolber L *et al.* Vertebral endplate change as a feature of intervertebral disc degeneration: A heritability study. *Eur Spine J.* 2014;23:1856-1862.

15. Karppinen J, Daavittila I, Solovieva S *et al.* Genetic factors are associated with modic changes in endplates of lumbar vertebral bodies. *Spine (Phila Pa 1976).* 2008;33:1236-1241.

16. Karppinen J, Solovieva S, Luoma K *et al.* Modic changes and interleukin 1 gene locus polymorphisms in occupational cohort of middle-aged men. *Eur Spine J.* 2009;18:1963-1970.

17. Virtanen IM, Noponen N, Barral S *et al.* Putative susceptibility locus on chromosome 21q for lumbar disc disease (LDD) in the finnish population. *J Bone Miner Res.* 2007;22:701-707.

18. Kuisma M, Karppinen J, Niinimäki J *et al.* A three-year follow-up of lumbar spine endplate (modic) changes. *Spine* 2006;31:1714-1718.

19. Määttä JH, Karppinen J, Paananen M *et al.* Refined phenotyping of modic changes: Imaging biomarkers of prolonged severe low back pain and disability. *Medicine (Baltimore)* 2016;95:e3495.

20. McKenna A, Hanna M, Banks E *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297-1303.

21. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.

22. Wu L, Aster JC, Blacklow SC, Lake R *et al* MAML1, a human homologue of drosophila mastermind, is a transcriptional co-activator for NOTCH receptors. *Nat Genet* 2000;26:484-489.

23. Iozzo RV, Cohen IR, Grassel S *et al*. The biology of perlecan: The multifaceted heparan sulphate proteoglycan of basement membranes and pericellular matrices. *Biochem J* 1994;302 (Pt 3):625-639.

24. Murdoch AD, Dodge GR, Cohen I *et al*. Primary structure of the human heparan sulfate proteoglycan from basement membrane (HSPG2/perlecan). A chimeric molecule with multiple domains homologous to the low density lipoprotein receptor, laminin, neural cell adhesion molecules, and epidermal growth factor. *J Biol Chem* 1992;267:8544-8557.

25. Baschal EE, Wethey CI, Swindle K *et al.* Exome sequencing identifies a rare HSPG2 variant associated with familial idiopathic scoliosis. *G3 (Bethesda)* 2014;5:167-174.

26. Stum M, Davoine CS, Vicart S *et al.* Spectrum of HSPG2 (perlecan) mutations in patients with schwartz-jampel syndrome. *Hum Mutat* 2006;27:1082-1091.

27. Arikawa-Hirasawa E, Wilcox WR, Le AH *et al.* Dyssegmental dysplasia, silverman-handmaker type, is caused by functional null mutations of the perlecan gene. *Nat Genet* 2001;27:431-434.

28. Handmaker SD, Campbell JA, Robinson LD *et al.* Dyssegmental dwarfism: A new syndrome of lethal dwarfism. *Birth Defects Orig Artic Ser* 1977;13:79-90.

29. Nicole S, Davoine CS, Topaloglu H *et al.* Perlecan, the major proteoglycan of basement membranes, is altered in patients with schwartz-jampel syndrome (chondrodystrophic myotonia). *Nat Genet* 2000;26:480-483.

30. Wu L, Aster JC, Blacklow SC *et al.* MAML1, a human homologue of drosophila mastermind, is a transcriptional co-activator for NOTCH receptors. *Nat Genet* 2000;26:484-489.

31. Shang X, Wang J, Luo Z *et al.* Notch signaling indirectly promotes chondrocyte hypertrophy via regulation of BMP signaling and cell cycle arrest. *Sci Rep* 2016;6:25594.

32. Dong Y, Jesse AM, Kohn A *et al.* RBPjkappa-dependent notch signaling regulates mesenchymal progenitor cell proliferation and differentiation during skeletal development. *Development* 2010;137:1461-1471.

33. Kohn A, Dong Y, Mirando AJ *et al.* Cartilage-specific RBPjkappa-dependent and -independent notch signals regulate cartilage and bone development. *Development* 2012;139:1198-1212.

34. Mead TJ, Yutzey KE. Notch pathway regulation of chondrocyte differentiation and proliferation during appendicular and axial skeleton development. *Proc Natl Acad Sci U S A* 2009;106:14420-14425.

35. Liu Z, Chen J, Mirando AJ *et al.* A dual role for NOTCH signaling in joint cartilage maintenance and osteoarthritis. *Sci Signal* 2015;8:ra71.

36. Liu Z, Ren Y, Mirando AJ *et al.* Notch signaling in postnatal joint chondrocytes, but not subchondral osteoblasts, is required for articular cartilage and joint maintenance. *Osteoarthritis Cartilage* 2016;24:740-751.

37. Mirando AJ, Liu Z, Moore T *et al* RBP-jkappa-dependent notch signaling is required for murine articular cartilage and joint maintenance. *Arthritis Rheum* 2013;65:2623-2633.

38. Shen H, McElhinny AS, Cao Y *et al.* The notch coactivator, MAML1, functions as a novel coactivator for MEF2C-mediated transcription and is required for normal myogenesis. *Genes Dev* 2006;20:675-688.

39. Zhao Y, Katzman RB, Delmolino LM *et al.* The notch regulator MAML1 interacts with p53 and functions as a coactivator. *J Biol Chem* 2007;282:11969-11981.

40. Alves-Guerra MC, Ronchini C, Capobianco AJ. Mastermind-like 1 is a specific coactivator of beta-catenin transcription activation and is essential for colon carcinoma cell survival. *Cancer Res* 2007;67:8690-8698.

41. Watanabe T, Oyama T, Asada M *et al.* MAML1 enhances the transcriptional activity of Runx2 and plays a role in bone development. *PLoS Genet* 2013;9:e1003132.

42. Komori T, Yagi H, Nomura S *et al*. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* 1997;89:755-764.

43. Rutges JP, Duit RA, Kummer JA *et al.* Hypertrophic differentiation and calcification during intervertebral disc degeneration. *Osteoarthritis Cartilage* 2010;18:1487-1495.

44. Torkki M, Majuri ML, Wolff H *et al.* Osteoclast activators are elevated in intervertebral disks with modic changes among patients operated for herniated nucleus pulposus. *Eur Spine J* 2016;25:207-216.

45. Wang Y, Videman T, Battié MC. Modic changes: Prevalence, distribution patterns, and association with age in white men. *The Spine Journal* 2012;12:411-416.

46. Mok FP, Samartzis D, Karppinen J *et al.* Modic changes of the lumbar spine: Prevalence, risk factors, and association with disc degeneration and low back pain in a large-scale population-based cohort. *Spine J* 2016;16:32-41.

47. Bodmer W, Tomlinson I. Rare genetic variants and the risk of cancer. *Curr Opin Genet Dev* 2010;20:262-267.

48. Struewing JP, Hartge P, Wacholder S *et al.* The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among ashkenazi jews. *N Engl J Med* 1997;336:1401-1408.

49. Krajc M, De Greve J, Goelen G, Teugels E. BRCA2 founder mutation in slovenian breast cancer families. *Eur J Hum Genet* 2002;10:879-882.

50. Erkko H, Xia B, Nikkila J *et al.* A recurrent mutation in PALB2 in finnish cancer families. *Nature* 2007;446:316-319.

51. Southey MC, Goldgar DE, Winqvist R *et al.* PALB2, CHEK2 and ATM rare variants and cancer risk: Data from COGS. *J Med Genet* 2016.

52. Ong CT, Corces VG. Enhancer function: New insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* 2011;12:283-293.

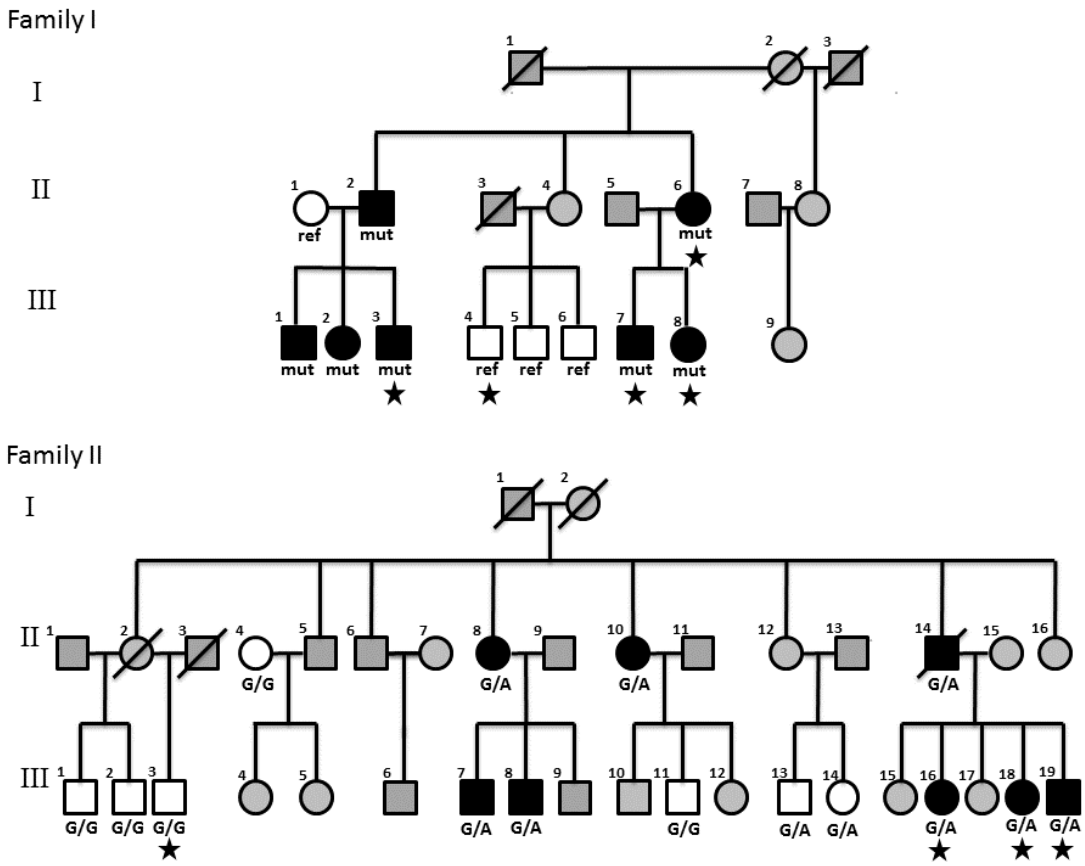


Figure 1. Family pedigrees. Individuals with MC are marked with black symbols, those without MC with white and gray symbols represent undetermined phenotype. Deceased individuals are marked with a diagonal line through the symbol. Genotypes of the family members validated by Sanger sequencing are indicated, and individuals included in exome sequencing are marked with stars. Abbreviations: ref = *HSPG2* reference allele (p.Q1611), mut = *HSPG2* mutation (c.4829_4831delAGCinsCACTGTG, p.Q1611delfsX20), G/A = *MAML1* (c.1657G>A, p.Glu553Lys), G/G = *MAML1* reference allele (p. Glu553).

Table 1. Types, sizes and locations of Modic changes (MC) in the affected family members. Depth is defined as 1 = MC in the endplate, 2 = MC less than 25%, 3 = MC from 25 to 50%, 4 = MC more than 50% of the height of the vertebral body. Abbreviations: R = rostral, C = caudal, a = anterior segment, m = middle segment, p = posterior segment, f = all segments.

		L1-2		L2-3		L3-4		L4-5		L5-S1	
Family I		R	C	R	C	R	C	R	C	R	C
II-2	Type	2	2	1/2	1/2	0	1	2	2	1/2	2
	Size	3f	3f	3f	3f	0	3p	3f	3f	3f	3f
II-6	Type	0	0	1/2	1/2	2	2	2	2	2	2
	Size	0	0	2p	2p	2f	2f	3f	3f	2a	1f
III-1	Type	0	0	0	0	0	0	0	0	1	1
	Size	0	0	0	0	0	0	0	0	1m	1mp
III-2	Type	0	0	0	0	2	2	0	2	1/2	1/2
	Size	0	0	0	0	1am	2am	0	2am	1mp	1mp
III-3	Type	0	0	0	0	0	0	1/2	1/2	2	2
	Size	0	0	0	0	0	0	1a	1a	2am	2f
III-7	Type	0	0	0	0	0	0	2	2	1/2	1/2
	Size	0	0	0	0	0	0	2f	3f	3am	2f
III-8	Type	0	0	0	0	0	0	1	1	0	0
	Size	0	0	0	0	0	0	2f	2am	0	0

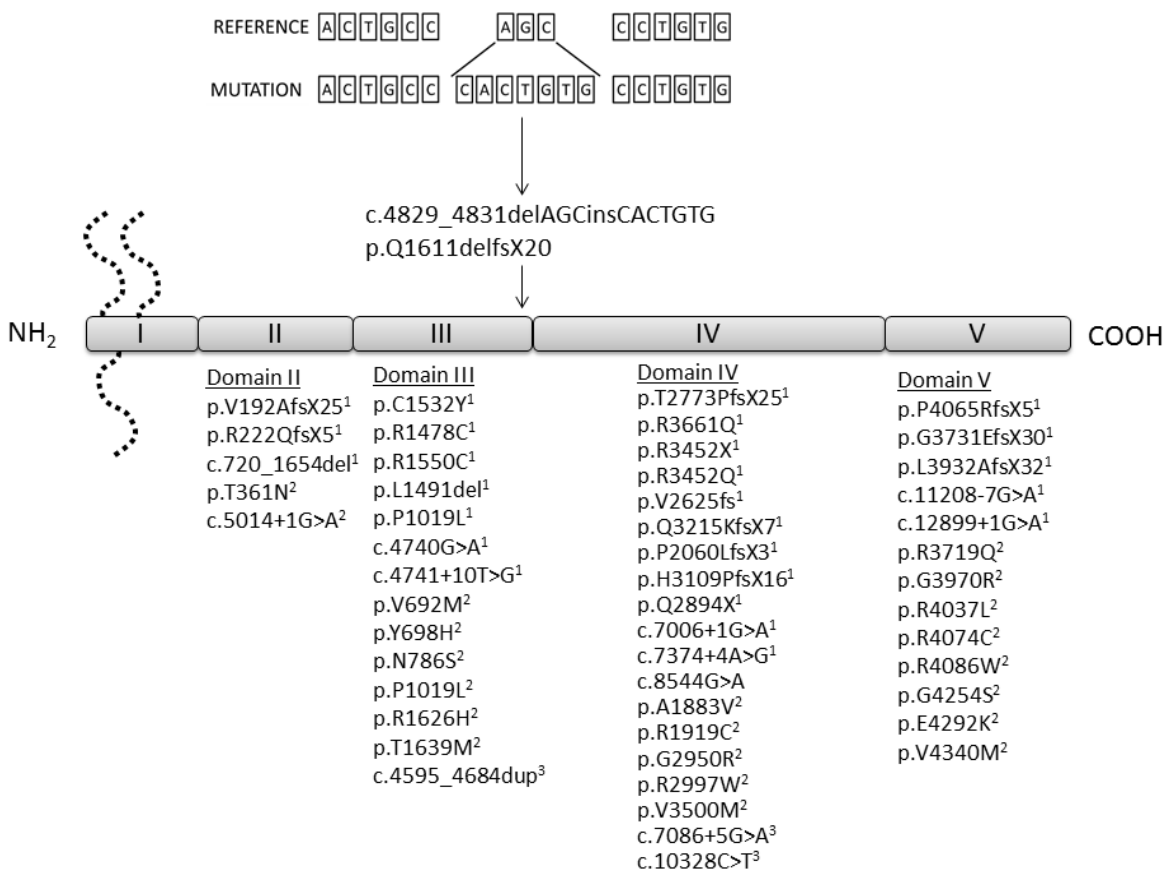
		L1-2		L2-3		L3-4		L4-5		L5-S1	
Family II		R	C	R	C	R	C	R	C	R	C
II-8	Type	0	0	0	0	0	0	2	0	2	2
	Size	0	0	0	0	0	0	2am	0	2f	2f
II-10	Type	0	0	0	0	0	0	1/2	1/2	0	0
	Size	0	0	0	0	0	0	2am	2f	0	0
II-14	Type	0	0	2	2	0	0	2	2	0	0
	Size	0	0	3f	3f	0	0	1am	2am	0	0
III-7	Type	0	0	0	0	0	0	0	0	2	0
	Size	0	0	0	0	0	0	0	0	1p	0
III-8	Type	0	0	0	0	0	0	1/2	1/2	0	0
	Size	0	0	0	0	0	0	1am	2am	0	0
III-16	Type	0	0	0	0	0	0	1/2	1/2	1	1
	Size	0	0	0	0	0	0	3f	3f	2m	1m
III-18	Type	0	0	0	0	0	0	2	2	2	2
	Size	0	0	0	0	0	0	1m	1m	3f	2f
III-19	Type	0	0	0	0	2	2	0	0	0	0
	Size	0	0	0	0	2a	2am	0	0	0	0

Table 2. Mutations co-segregating with Modic changes in the two Finnish families. 1000G = 1000 Genomes database, SISu = Sequencing Initiative Suomi database, ExAC Fin = Finnish population in Exome Aggregation Consortium database.

Family	Gene	Location	Mutation	Rs-number	1000G	SISu	ExAC Fin
I	*HSPG2*	chr1: 22188515-22188517	p.Q1611delfsX20	NA	NA	NA	NA
II	*MAML1*	chr5:179193668	p.Glu553Lys	rs61753465	0.001	0.001	0.0003024

Table 3. *In silico* pathogenicity estimations for the mutations. NA = not available.

In silico analysis	Predictions	c.4829_4831delAGCinsCACTGTG, p.Q1611delfsX20	c.1657G>A, p.Glu553Lys
SIFT	Deleterious: score ≤ 0.05 Tolerated: score >0.05	NA	Tolerated
PolyPhen-2	Probably damaging: score ≥ 0.957 Possibly damaging: 0.453≤score≤0.956 Benign: score ≤0.452	NA	Probably damaging
MutationTaster	Disease_causing_automatic Disease_causing Polymorphism Polymorphism_automatic	Disease_causing	Polymorphism



Supplementary Figure S1. Schematic presentation of the perlecan protein. The *HSPG2* gene has 94 exons and the encoded protein perlecan is divided into five domains (numbered from I to V). Three glycosaminoglycan (GAG) chains are attached to the domain I. The c.4829_4831delAGCinsCACTGTG mutation is located in the domain III. Mutations described in SJS, IS and DDSH are marked under the domain they locate to (Nicole et al., 2000; Arikawa-Hirasawa et al., 2001; Stum et al., 2006; Baschal et al., 2014; Arikawa-Hirasawa et al., 2002). [1] = SJS, [2] =IS, [3] =DDSH.