# Generalised Osteoarthritis: from Mendelian Disorder to Complex Disease

Min, J.L.

CHAPTER

# 5.2 Mutation analysis of candidate genes within the 2q33.3 linkage area for familial early onset generalised osteoarthritis

Josine L. Min, Ingrid Meulenbelt, Margreet Kloppenburg, Cornelia M. van Duijn and P. Eline Slagboom

*Arthritis and Rheumatism 2006: Submitted for publication*

## Abstract

In a genome-wide linkage scan of seven families with familial early onset osteoarthritis (FOA), we mapped a FOA locus to a 5 cM region on chromosome 2q33.3-2q34 with a maximum LOD score of 6.05. To identify causal variants, 20 positional candidate genes were sequenced for coding, splicing, and 5′ and 3′ untranslated regions. The pathogenicity of possible disease causing variants was evaluated using predicted effects on protein structure and function, splicing enhancers, degree of conservation and frequency in 790 unrelated subjects from the population-based Rotterdam study scored for the presence of radiographic signs of OA (ROA). Nine novel variants, identified in *NRP2*, *KIAA1571*, *ADAM23*, *IDH1*, *PIP5K3* and *PTHR2*, cosegregated with FOA, of which two were promising. *IDH1* Y183C cosegregated in one family, involved a conserved amino acid change and showed a damaging effect predicted by PolyPhen and SIFT. In the Rotterdam sample, carriers of *IDH1* Y183C (0.02) had an increased but insignificant risk for generalised ROA. The second variant, *NRP2* c.1938-21T>C cosegregated in three families. In the Rotterdam sample, carriers conferred a significant increased risk of 2.1 (95% CI, 1.1-4.1, $P$ = 0.032), to have generalised ROA. Furthermore, two variants (*NRP2* c.1938-21T>C and *IDH1* c.933-28C>T) occurred together on the haplotypes segregating with FOA in two out families. This haplotype was rare in the Rotterdam sample (0.0013). Two promising variants in or near *NRP2* and *IDH1,* may be not sufficient for the onset of FOA alone but might have a modulating role with FOA. Confirmation in other OA populations is required.

## Introduction

Osteoarthritis (OA) is the most common joint disease and is the leading cause of disability among the elderly. It is a degenerative disease of the joints characterised by cartilage degeneration and subchondral bone remodeling. Clinically, OA manifests itself with pain, stiffness, disability and loss of joint function. Twin, sibling and segregation studies have highlighted substantial heritability for primary OA ranging between 39%-80% depending on the sex and joint location[1,2]. OA often arises as a complex trait, but rare familial forms with autosomal dominant transmission have been reported[3-5]. In familial forms of osteochondrodysplasia often displayed with secondary early onset OA, mutations in several structural genes of the extracellular cartilage matrix (ECM) including *COL2A1*, *COL9A1*, *COMP* have been identified[6,7]. There is no convincing evidence yet that these genes are implicated in the susceptibility for primary early onset OA occurring without dysplasia[7-9]. Linkage and association studies on the basis of joint- or sex-specific OA definitions have yielded several loci associated with OA susceptibility including the interleukin-1 cluster (*IL1A*, *IL1B*, *IL1RN*)[10-12], *ESR1*[13,14], *VDR*[15,16], *MATN3*[17,18], *FRZB*[19,20], *IL4R*[21], *ASPN*[22] and *CALM1*[23].

Previously, we reported on a genome-wide linkage scan which revealed evidence for linkage at 2q33.3-2q34 with a two point LOD score of 6.05 (theta = 0.00) and a multipoint NPL score of 4.70 ($P$ = 0.0013) in seven extended families with early onset generalised OA (FOA) without dysplasia[24]. This region of 5 cM highlighted two strong candidate genes, *PTHR2* and *FZD5*, which were unlikely to explain the observed linkage signal[24]. None of the OA susceptibility genes mentioned are located in this linkage region. However, this interval is nearby the locus 2q32 identified in the UK study in which female siblings with hip replacement were investigated. In this study, a functional polymorphism in *FRZB,* a gene that antagonises the Wingless signalling, conferred susceptibility for hip OA in females[25]. Our subsequent investigation of *FRZB* demonstrated a significant association with radiographic generalised OA in a random sample from the population-based Rotterdam study and with clinical generalised OA in affected sibling pairs from the GARP study[26]. The high LOD score in the early onset OA families and absence of variants in the two most promising genes prompted us to perform an extended systemic mutation analysis of candidate genes in the 2q33.3-2q34 linkage area. In the present study, we report results of mutation screening of the entire coding region, splice sites, and 5′ and 3′ untranslated regions (UTR) of 19 positional candidate genes and the *FRZB* gene in affected family members with early onset generalised OA.

## Materials and methods

### *Families*

Families containing patients that express primary generalised OA in each generation were collected from different parts of the Netherlands. Informed consent was obtained from all patients and the medical ethics committee of the Leiden University Medical Centre approved the study. Probands were recognised through Rheumatology outpatient clinics. Family members were recruited via probands. Initially, we used questionnaires to select eligible families. For eligible families, complete medical history and available radiographs were obtained from Rheumatologists of almost all affected family members (81%). Radiographs were re-evaluated for signs of chondrodysplasia, spinal dysplasia or abnormal development of the epiphyses of the peripheral joints. These features were absent in all families. The presence of radiographic OA (ROA) was assessed according to Kellgren/Lawrence criteria[27] by an experienced reader. Some individuals had marked Heberden's nodes and ankle OA. A selection of affected and unaffected individuals of family 1, 2, 4, and 7 were additionally visited for physical examination in order to prevent misclassification. The mean age of onset of OA in these patients was 33 years ranging between 20 and 50 years. The phenotype within these families is characterised by distinct progressive OA in the absence of mild or severe chondrodysplasia, however, with symptoms and ROA at multiple joint sites simultaneously including involvement of the hands with noduli, knees, hips, ankle and spine. Individuals with clinical and radiographic evidence of OA in two or more joint sites before the age of 50 years were considered affected. Extensive description of the phenotype in family 1, which is representative for the phenotypes also of the other extended families included, is described elsewhere [28]. All clinical diagnostic decisions were made independent to genetic linkage analysis and homogeneity of the phenotype between different families was checked.

### *The Rotterdam sample*

The Rotterdam study, which comprises 7,983 Caucasian participants is a prospective, population-based cohort study of the determinants and prognosis of chronic diseases in the elderly[29]. The medical ethics committee of the Erasmus University Medical Centre approved the study, and informed consent was obtained from all subjects. In a random sample of 809 unrelated subjects (the Rotterdam sample), ages 55-65 years, radiographs were scored for the presence of radiographic OA (ROA) in two knees, two hips[30], 36 hand joints and three levels of the thoracocolumbar spine[30;31]. All radiographs were scored according to the Kellgren/Lawrence grading system (grades 0-4)[27] by two independent readers, blinded to all other data of the

participant. Definite ROA at a particular joint site was defined as a Kellgren/Lawrence score of two or more[27]. In the hands, 36 separate joints were scored comprising eight joint groups: distal interphalangeal joints, the interphalangeal joint of the thumb, the proximal interphalangeal joints, the metacarpalphalangeal joints, the first carpometacarpal joints, the trapezoscaphoideal joints, the radionavicular joints and the distal radioulnar joints. By definition, ROA of the spine is confined to the apophyseal joints, but these joints could not be assessed on the lateral radiographs of the spine that were available. Instead, we assessed disc degeneration (DD) of the spine, at three levels i.e. thoracic (Th4 to Th12), lumbar (L1 to L4 or L5) and lumbosacral (L5-S1 or L5-L6).

We analysed and evaluated the occurrence and the generalised OA status of carriers of novel variants in the population-based sample as a qualitative trait using previously described definitions[32]. In brief, subjects with two or more of the following four criteria were considered as affected with generalised ROA: hand ROA in three or more hand joint groups (the right and left hands were considered separately), spinal DD in two or more disc levels, knee ROA in one or two knees and hip ROA in one or two hips. Subjects affected with generalised ROA were compared to the complete Rotterdam sample from which subjects with generalised ROA were excluded. Phenotypic data for assessment of generalised ROA was available for 790 subjects.

## *Mutation analysis*

Genomic DNA was isolated from EDTA blood of affected and unaffected family members. Initially, three affected family members (individual 14 from family 1, individual 10 from family 2 and individual 9 from family 4) were screened for possible mutations by direct forward and reverse sequencing from both ends (Figure 1). If a novel variant was identified, unaffected family members of these families and family members of remaining families were sequenced. Reference sequences corresponding to all coding and 5′ and 3′ UTR regions of the genes were obtained from the UCSC genome browser assembly May 2004 (http://genome.ucsc.edu/) or the Ensembl Genome database v35 (www.ensembl.org), NCBI build 35. Table 1 shows the Genbank numbers. To amplify exons, forward and reverse primer sets (primer sequences upon request) were designed with at least 25 bp flanking intronic sequences using Primer3 (http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi) with the conditions described by[33]. 3′UTR of *NRP2* and exons 1-27 of *PIP5K3* have not been sequenced due to current genome browser updates. PCR amplifications were carried out in a volume of 15 µl that contained 15 ng genomic DNA, 4.1 pmol of the PCR primers, 1.5 mM $MgCl_2$, 0.2 mM and 0.6 units of rTaq polymerase (Amersham Biosciences) or 0.6 units of HotfirePol® DNA

polymerase and solution S (Solis Biodyne) for GC-rich regions or standard conditions of the GC-rich PCR system (Roche). Reactions were cycled at 94°C for respectively 1 min or 15 min for GC-rich regions and then cycled for 35 cycles of 94°C for 30 sec, 57°C for 1 min 15 sec, 72°C for 30 sec, and finally incubated for 6 min at 72°C on B&L primus HT cyclers. PCR products were purified using Multiscreen 96 well plates (Millipore) filled with Sephadex (Amersham Biosciences) and quantified on 1.5% agarose gels. PCR products were sequenced for possible mutations using an ABI3730 capillary sequencer with Big Dye chemistry (Applied Biosystems).

## *Validation of possible mutations*

To investigate whether variants were novel or common SNPs, all variants observed in affected family members were blasted for existence using National Centre of Biotechnology Information (NCBI) SNP blast, build 124 (http://www.ncbi.nlm.nih.gov/SNP/snpblastByChr.html). We applied the nomenclature of den Dunnen et al. (2001) to describe these variants[34]. When a novel variant was found, segregation analysis of the variant with OA within all family members of the seven families was performed. The impact of a novel variant involving an amino acid change was examined using PolyPhen (http://tux.embl-heidelberg.de/ramensky/index.shtml) or SIFT (http://blocks.fhcrc.org/sift/SIFT.html)[35]. PolyPhen is a computational tool that predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations[36]. SIFT (sorting intolerant from tolerant) estimates tolerance indices that predict tolerated and deleterious substitutions for every position of the query protein sequence using multiple alignment information[37]. To test the effect on the splicing process, exonic variants were screened for exonic splicing enhancers sequences using http://exon.cshl.edu/ESE[38]. Conservation was determined using the Multiz Alignments and Conservation track of the UCSC genome browser version 140 (http://genome.ucsc.edu/) which shows a measure of evolutionary conservation in 17 vertebrates, including mammalian, amphibian, bird, and fish species.

## *Genotyping*

Nine novel variants were genotyped in 790 random subjects from the population-based Rotterdam study[39]. All variants were in Hardy-Weinberg equilibrium (HWE). Variants were genotyped using Sequenom homogenous Mass Extend MassARRAY System (Sequenom Inc, San Diego, CA) using standard conditions. Genotypes were analysed using Genotyper version 3.0 software (Sequenom Inc.).

### *Statistical analysis*

HWE was calculated with an exact HWE test for rare alleles implemented in R version 2.3.1, (http://www.r-project.org/). A logistic regression model was fitted to measure the strength of association, which is expressed as odds ratios (OR) with 95% confidence intervals (95% CI) adjusted for age (years), body mass index (BMI in $kg/m^2$), and sex. In these analyses, homo- and heterozygous carriers of the risk allele were pooled. All analyses were performed with SPSS version 11 software (SPSS, Chicago, IL).

## Results

### *Families*

Previously, we performed a genome-wide linkage scan of seven extended families with early onset OA without any dysplasia. The scan revealed significant evidence for linkage to chromosome 2q33.3-2q34, implicating a 5 cM interval (4.6 Mb) between markers D2S1384 and D2S2178[24]. In family 2 and 4, most affected individuals carry identical heterozygous haplotypes resulting in two possible haplotypes explaining the linkage (Figure 1). In family 2, the most likely haplotype 3678 (A1) allows one phenocopy (individual 9) and the second likely haplotype 7658 (A2) allows two phenocopies (subject 6 and 17). In family 4, the most likely haplotype 5515232 (A1) allows no phenocopies and the second likely haplotype 5144984 (A2) allows one phenocopy (individual 7). As illustrated in Figure 1, recombinations in individuals conferring either haplotype A1 or A2 confined this linkage interval. To detect sequence changes, we initially screened three affected family members from three different families contributing most to the linkage: individual 14 in family 1, individual 10 in family 2 and individual 9 in family 4. When a novel variant was found, segregation analysis of the variant with OA within all family members of the seven families was performed.

### *Mutation screening of positional candidate genes*

On the basis of these result, we performed extended mutation analysis of the chromosome 2 locus. Using the human genome resources, 18 known RefSeq genes, nine predicted RefSeq genes (four genes with model status and five genes with predicted status in NCBI build 35) have been identified within the linkage area of chromosome 2 as shown in Table 1. Nearby the linkage region, at marker D2S72, the

*ICOS-CTLA4-CD28* cluster is located which is implicated in cytokine secretion and T-cell immunity[40-42]. *FRZB*, previously associated with hip OA and generalised OA is located more upstream at 2q32[43;44]. The entire coding region, splice sites, and 5' and 3' UTRs of *FRZB*, *CTLA4*, *CD28*, *ICOS*, *NRP2*, *NM_017759*, *NDUFS1*, *EEF1B2*, *GPR1*, *XM_371590*, *ADAM23*, *MDH1B*, *CPO*, *KLF7*, *CREB1*, *NM_030804*, *FZD5*, *IDH1*, *PIP5K3* and *PTHR2* were sequenced comprising 17 RefSeq genes and three predicted genes. Results of the screening of *PTHR2* and *FZD5* have previously been reported[24]. If the considered variant appeared novel in the current dbSNP database (build 124) and cosegregated with OA in the affected family, it was evaluated according to the criteria previously mentioned.

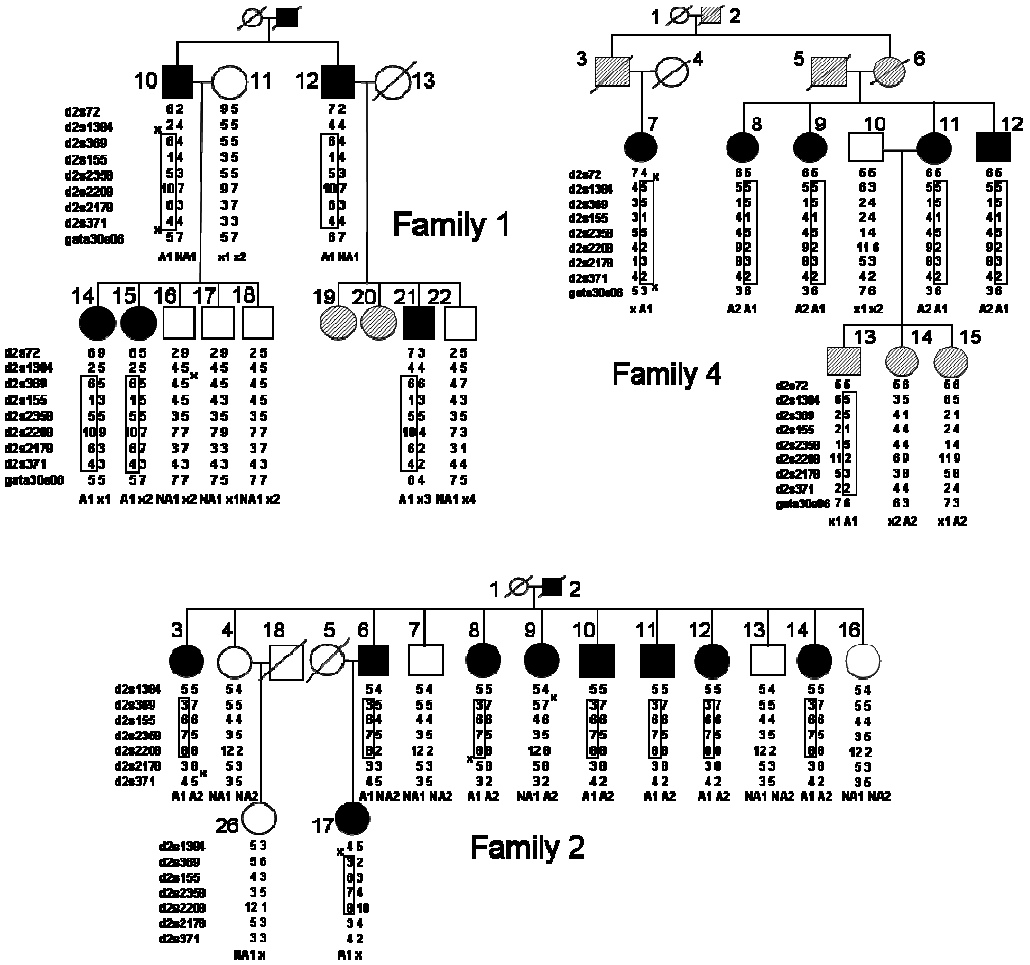


**Figure 1** Family pedigrees segregating for generalised early onset OA and chromosome 2q33.3-2q34 loci. Blackened circles and squares indicate affected females and males, respectively. White individuals represent unaffected family members. Crosses indicate recombinations. Shaded individuals indicate individuals with unknown disease status. The black lined box shows the most likely haplotype (A1), allowing one phenocopy (individual 9) in family 2.

**Table 1** Known genes between the markers D2S72 and D2S2178 based on RefSeq, mRNA, TrEMBL and Swiss-Prot prediction using Ensembl genome database v35

| Marker | Genbank ID | HUGO ID | Description | No. Exons |
|---|---|---|---|---|
| | NM_001037631, NM_005214 | CTLA4 | Cytotoxic T-lymphocyte protein 4 precursor | 3,4 |
| | NM_006139 | CD28 | T-cell specific surface glycoprotein CD28 precursor | 4,5 |
| D2S72 | NM_012092 | ICOS | Inducible T-cell co-stimulator precursor | 5 |
| D2S1384 | NM_057177, NM_205863, NM_152526 | ALS2CR19 | Amyotrophic lateral sclerosis 2 chromosomal region candidate gene protein 19 | 22,23,23 |
| | BC009222, NM_201264, NM_018534, NM_201279, NM_003872, NM_201266 | NRP2 | Neurophilin-2 precursor | 5,16,16, 16,17,17 |
| | NM_017759 | NM_017759 | Unknown (NP_060229) | 12 |
| D2S155 | NM_005006 | NDUFS1 | NADH-ubiquinone oxidoreductase 75 kDA subunit | 19 |
| | NM_001959, NM_021121 | EEF1B2 | Elongation factor 1-beta | 6,7 |
| | NM_005279 | GPR1 | Probable G protein-coupled receptor | 3 |
| | XM_371590 | XM_371590 | Unknown (Q9HCK1) | 1 |
| | XM_496742 | XM_496742 | Unknown (Q9BZ60) | 1 |
| D2S369 | NM_003812 | ADAM23 | A disintegrin and metalloproteinase 23 preproprotein | 26 |
| D2S2358 | NM_206892 | MDH1B | malate dehydrogenase 1B (NP_996775) | 10 |
| | NM_014929 | NM_014929 | Unknown (KIAA0971) | 12 |
| | NM_173077 | CPO | Carboxypeptidase O | 9 |
| | NM_003709 | KLF7 | Krueppel-like factor 7 | 4 |
| | NM_004379, NM134442 | CREB1 | cAMP-response element binding protein | 8,9 |
| | NM_145280 | NM_145280 | hepatocellular carcinoma-associated antigen (NP_660323) | 4 |
| | NM_152523 | NM_152523 | hypothetical protein FLJ40432 (NP_689736) | 11 |
| | NM_030804 | NM_030804 | Unknown (CB031) | 1 |
| | NM_003468 | FZD5 | Frizzled 5 precursor | 2 |
| D2S2208 | XM_371591 | XM_371591 | Unknown (Q8WW68) | 1 |
| | NM_006891 | CRYGD | Gamma crystallin D | 3 |
| | NM_020989 | CRYGC | Gamma crystallin C | 3 |
| | NM_005210 | CRYGB | Gamma crystallin B | 3 |
| | NM_014617 | CRYGA | Gamma crystallin A | 3 |
| | XM_371592 | XM_371592 | similar to RIKEN cDNA D630023F18 | 9 |
| | NM_005896 | IDH1 | Isocitrate dehydrogenase 1 | 10 |
| | NM_152671, NM_015040 | PIP5K3 | FYVE finger containing phosphonositide | 10,42 |
| D2S2178 | NM_005048 | PTHR2 | Parathyroid hormone receptor precursor | 13 |

**Table 2** Possible mutations segregating in FOA families

| Gene | Gene Location | DNA nt change[1] | aa change | Sequence conservation[2] | Splicing[3] | PolyPhen[4] | SIFT[5] | Segregation family[6] |
|---|---|---|---|---|---|---|---|---|
| *NRP2* | 3'UTR[7] | c.941A>C | | no | | | | F4:A1 |
| *NRP2* | intron 7[8] | c.1938-21T>C | | no | possible | | | F2:A1, F4:A2, F1:A1 |
| *KIAA1571* | exon 1 | c.6368G>T | R2133S | no | yes | benign (1.35)[9] | tolerated (0.00)[9] | F4:A1, F7:A1 |
| *ADAM23* | intron 20 | c.2065+24C>T | | no | possible | | | F2:A2 |
| *IDH1* | exon 6 | c.782A>G | Y183C | yes | no | damaging (2.89) | not tolerated (0.00) | F2:A2 |
| *IDH1* | intron 6 | c.933-28C>T | | no | possible | | | F2:A1, F4:A2 |
| *PIP5K3* | 3'UTR | c.8429T>A | | yes | | | | F4:A2 |
| *PIP5K3* | 3'UTR | c.8434insC | | no | | | | F4:A2 |
| *PTHR2* | exon 6 | c.786G>T | A225S | no | no | benign (0.12) | tolerated (0.77) | F4:A2 |

[1] Numbers are based on positions in BC009222 (gi:33874364), NM_201267 (gi:41872566), XM_371590 (gi:42656560), NM_003812 (gi:73765550), NM_005896 (gi:28178824), NM_015040 (gi:50881947),NM_005048 (gi:39995097).
[2] Sequence conservation based on vertebrate Multiz Alignment and Conservation of 17 species (UCSC Genome browser).
[3] Splicing effects in exons have been predicted by ESE finder.
[4] PolyPhen prediction PSIC score difference.
[5] SIFT prediction with normalised probability.
[6] A1 represents the most likely haplotype and A2 the second likely affected haplotype.
[7] This variant is located in one of the six transcripts (BC009222, gi:33874364).
[8] This variant is located in six of the seven transcripts (NM_201267 (gi:41872566), NM_003872 (gi:41872532), NM_201279 (gi:41872571), NM_018534 (gi:41872543), NM_201266 (gi:41872561), NM_201264 (gi:41872556)).
[9] Poor predictions due to little information.

## *Evaluation of novel variants*

The initial screening indicated 26 novel variants (17 SNPs and 9 insertion/deletion polymorphisms). From these 26 variants, only nine promising variants cosegregated with OA within one or more families as illustrated in Table 2. Three of these variants were found in coding regions and involved an amino acid change: (*KIAA1571* R2133S, *IDH1* Y183C and *PTHR2* A225S). Six variants were located in UTR regions (*PIP5K3* c.8429T>A, *PIP5K3* c.8434insC and *NRP2* c.941A>C) or in vicinity of exon/intron boundaries (*NRP2* c.1938-21T>C, *ADAM23* c.2065+24C>T and *IDH1* c.933-28C>T). Using PolyPhen, SIFT and ESE finder analysis to predict possible functional effects of these variants, two variants emerged as potential mutations: *KIAA1571* R2133S and *IDH1* Y183C (Table 2). In addition, three variants (*NRP2* c.941A>C, *PIP5K3* c.8429T>A and *PIP5K3* c.8434insC) were conserved across other species and might be of functional importance.

The predicted *KIAA1571* gene probably belongs, as predicted in Unigene, to the fibronectin type III and M protein repeat family in C. elegans. Fibronectin is a component of the ECM and *KIAA1571* may therefore be an excellent candidate gene. The G/T nucleotide change in the third exon of *KIAA1571* (Q9HCK1) results in disruption of exonic splicer enhancer motifs which serves as binding site for Serine/Arginine protein 40 and 55 and might be therefore a functional variant. However, because this gene is a predicted gene, little is known about other possible predicted functional effects on the protein. The novel variant *KIAA1571* R2133S cosegregated in family 4 and 7 with OA and showed a rare population frequency of 0.01 corresponding to nine carriers of 763 genotyped (Table 3). In the Rotterdam sample, we did not observe a significant association of this variant with generalised ROA. It is unlikely that this variant is a causal mutation.

Isocitrate dehydrogenase 1 (*IDH1*) encodes a cytoplasmic enzyme which catalyzes the oxidative decarboxylation of isocitrate to 2-oxoglutarate and has a significant role in cytoplasmic NADPH production[45]. In *IDH1*, two variants (Y183C and c.933-28C>T) cosegregated with the OA phenotype. *IDH1* Y183C cosegregated in affected family members in family 2 (A2, Figure 1). This variant was located in exon 6 encoding the isocitrate/isopropylmalate dehydrogenase domain (PF00180) of IDH1, predicted to be probably damaging for the protein structure/function by SIFT and PolyPhen, and highly conserved across all species investigated. Based on these results, this variant could be functional for the onset of generalised OA. In the Rotterdam sample, we observed 14 carriers out of 767 genotyped corresponding to a frequency of 0.02. In addition, carriers of this variant conferred an odds ratio, adjusted for age, BMI and sex of 2.8 (95% CI, 0.82-9.7, $P = 0.10$) to have generalised ROA, as shown in Table 3.

Another variant in this gene, IDH1 c.933-28C>T, was identified in family 2 and 4, near the intron/exon boundary of exon 7. This variant was not conserved across other species, and was not associated with generalised ROA in the Rotterdam sample (frequency 0.04).

Neuropilin 2 (*NRP2)*, is an interesting gene because it encodes for the co-receptor of vascular endothelial growth factor$_{165}$ (VEGF $_{165}$) which is an essential factor for endochondral ossification[46;47]. Furthermore, VEGF and its receptors are expressed in OA cartilage and VEGF stimulates production of ECM degrading matrix metalloproteinases (MMPs)[48;49]. In the *NRP2* gene, two novel variants were found: c.941A>C and c.1938-21T>C. In family 4, *NRP2* c.941A>C was identified in a residue with a low conservation score. This variant showed a frequency of 0.03 in the random population and no significant association with generalised ROA was observed. The second *NRP2* variant, c.1938-21T>C, was not conserved and cosegregated in three families (1, 2, and 4) and was more frequent in the population (0.07). Carriers of at least one risk allele of the *NRP2* c.1938-21T>C variant, conferred a significant increased risk of 2.1 (95% CI, 1.1-4.1, $P$ = 0.032), adjusted for age, sex, BMI, to have generalised ROA (Table 3).

Phosphatidylinositol-3-phosphate/phosphatidylinositol 5-kinase, type III *(PIP5K3)* catalyzes the phosphorylation of phosphatidylinositol-4-phosphate and has a role in endosome-related membrane trafficking[50]. We found two novel variants (*PIP5K3* c.8429T>A and *PIP5K3* c.8434insC) in the 3'UTR region of *PIP5K3* in family 4. *PIP5K3* c.8429T>A involved a highly conserved residue. *PIP5K3* c.8434insC was not conserved. In the Rotterdam sample, we observed that *PIP5K3* c.8429T>A and *PIP5K3* c.8434insC showed a population frequency of 0.04 and were in complete LD (D'= 1, $r^2$ = 1). In the Rotterdam sample, no significant associations of these variants with generalised ROA were observed excluding a possible pathogenic role in relation to the onset of FOA. Even though *PIP5K3* c.8429T>A occurred in a conserved residue, it is likely that these variants are neutral polymorphisms.

**Table 3** Frequencies of novel segregating variants in population-based sample scored for generalised ROA (GOA)

| Variant | Total ($n$ = 790) | No GOA[1] ($n$ = 659) | GOA[1] ($n$ = 131) | $P$[2] |
|---|---|---|---|---|
| *NRP2* c.941A>C | 0.03(743) | 0.03(601/17) | 0.02(123/2) | 0.44 |
| *NRP2* c.1938-21T>C[3] | 0.07(741) | 0.07(574/40) | 0.12(112/15) | 0.03 |
| *KIAA1571* c.6368G>T | 0.01(763) | 0.01(630/7) | 0.02(124/2) | 0.47 |
| *ADAM23* c.2065+24C>T | 0.04(741) | 0.04(591/22) | 0.03(124/4) | 0.79 |
| *IDH1* c.782A>G | 0.02(767) | 0.01(631/9) | 0.04(122/5) | 0.10 |
| *IDH1* c.933-28C>T | 0.04(773) | 0.04(622/23) | 0.04(123/5) | 0.94 |
| *PIP5K3* c.8429T>A[3] | 0.04(756) | 0.04(603/25) | 0.04(123/5) | 0.83 |
| *PIP5K3* c.8434insC[3] | 0.04(734) | 0.04(592/24) | 0.04(113/5) | 0.96 |

[1] Frequency (number of carriers of two common alleles/ number of carriers of one or two rare alleles).
[2] $P$ values adjusted for age, BMI and sex.
[3] One, one and two homozygous carriers for *NRP2* c.1938-21T>C, *PIP5K3* c.8429T>A and *PIP5K3* c.8434insC respectively.

Finally, we also examined whether some novel variants were inherited together in different families to identify a possible LD pattern or genetic interaction resulting in a high LOD score linked to OA. In the seven families, only two variants, *NRP2* c.1938-21T>C and *IDH1* c.933-28C>T occurred together on haplotype A1 in family 2 and on haplotype A2 in family 4. In the random population, this inheritance pattern was observed only once in 754 genotyped subjects (0.0013). This individual had spinal DD at three disc levels which has a prevalence of 0.04 in the random population.

## Discussion

Upon screening 20 genes localised within or near the area of linkage in three families that contributed most to the linkage, we identified nine novel variants cosegregating with OA. We evaluated the significance of the novel variants by prediction of pathogenicity using *in silico* functional analysis and by establishing the frequency in the random population. The *IDH1* Y183C variant was the most promising variant since it cosegregated on haplotype A2 with the OA phenotype in family 2, was predicted to be probably damaging for the protein structure/function and concerned a highly conserved residue. Among carriers of this variant in the general population (frequency 0.02), the risk of generalised ROA was 2.8 (95% CI, 0.82-9.7, $P = 0.10$). Given these results and the finding that family 2 already showed a significant linkage on its own[24], this variant may contribute to the FOA susceptibility in family 2.

The risk allele showed reduced penetrance among carriers in the Rotterdam study which might not be expected for a mutation causing a severe disorder such as FOA. It is expected that causal variants for severe OA are very rare and strongly correlated with generalised ROA. The reduced penetrance of this variant might reflect genetic interaction with other causal variants at this or other loci. Therefore, the pathogenic potential of this variant should be confirmed further in other populations with advanced OA or supported by functional assays. *IDH1* supplies NADPH for antioxidant systems suggesting a regulatory role in cellular defense against oxidative stress and in senescence[51]. Little is known about a possible role of *IDH1* in cartilage but we speculate that increased oxidative stress could make chondrocytes more susceptible to cell death which might contribute to the onset of OA.

A second promising variant, *NRP2* c.1938-21T>C, emerged from our mutation analysis which cosegregated in three families (1, 2, and 4) contributing most to the linkage. Carriers of this variant conferred a significant increased risk of 2.1 (95% CI, 1.1-4.1, $P = 0.032$), to have generalised ROA. Given the low effect size and the high frequency (0.07) in the random population, this variant may be not sufficient alone for the onset of FOA but it might be that this variant modulates OA susceptibility

and interacts genetically with another variant. *NRP2* acts as a co-receptor of $VEGF_{165}$, which is produced from hypertrophic chondrocytes and is also expressed in OA cartilage[52;53]. VEGF is an essential coordinator of growth plate morphogenesis and triggers cartilage remodeling[54]. VEGF may contribute to OA cartilage destruction through stimulation of MMPs[55-57].

The *NRP2* c.1938-21T>C occurred together with *IDH1* c.933-28C>T on haplotype A1 in family 2 and on haplotype A2 in family 4. This haplotype is extremely rare in the random population (0.0013). These results may suggest that family 2 and 4 might have a common ancestor with a FOA susceptibility haplotype. At this point, we conclude that possibly a causal variant in the LD pattern driven by both variants may have contribution to OA in family 2 and 4. Alternatively, both variants could be in LD with a causal variant each indicating an oligogenic disorder or a causal variant with a genetic modifier. More research is necessary to fully investigate the role of this haplotype in the OA phenotype.

How can we reconcile these results; what do they tell us about the proposed role of these genes in OA and why is there no compelling evidence for a genetic variant responsible for the observed LOD score of 6.0 in the families with early onset OA? Although the frequencies of carriers of novel variants identified in the OA families were rare (0.01-0.07%) in the random population-based sample as expected for such a severe phenotype as FOA, detected associations with generalised ROA did not achieve a high level of significance. This may be explained by the low frequency of the carriers, the absence of a clinical OA assessment and the relatively high frequency of generalised ROA above 55 years (0.17) which can result in reduced power of the statistical test, misclassification and spurious associations. We also analysed the presence of ROA among carriers in the random population as quantitative trait using a sumscore of the number of affected joints[24] which revealed similar findings (data not shown).

There is a possibility that the family members could be affected with the generalised ROA by chance alone although the phenotype is more severe (clinical OA before age 50). In family 2 and family 4, we observed two possible haplotypes segregating with the disease allowing one or two phenocopies suggesting that these individuals could be sporadic patients. Both haplotypes (A1 and A2) confined the linkage area. Consistent with an age-related disorder, we were not able to perform further segregation analysis because first degree relatives died or are too young to reveal symptoms of the disease.

We prioritised genes for sequencing based on the function known in the literature and human genome resources. Consequently, we have not selected genes which have clearly a function completely unrelated to OA like the gamma crystallins which is associated with cataract and lens morphology[58] or ALS2CR19 which plays a role in

asymmetrical cell division and cell polarization processes[59] or a few predicted genes like NM_014929 with unknown function. Obviously, these genes could carry an OA causing allele and will be screened for mutations in the future. In addition, we selected families that contributing most to the linkage for the initial mutation survey. Apparently, remaining four families could harbor causal variants in one of the screened genes. Alternatively, it is possible that affected family members may have a noncoding regulatory mutation in promoter or intron or a heterozygous deletion of one or more of the exons in one of the screened genes that has not yet been detected in our study. In addition, predictions of the on the basis of computational algorithms as PolyPhen, SIFT and ESE finder are difficult to interpret. Furthermore, even variants that are associated with disease may be in linkage disequilibrium with the true causal variant or are rare polymorphisms.

Although we were able to identify nine novel variants cosegregating with the FOA phenotype by our extended mutation analysis, we may found no robust evidence for a major disease gene responsible for the observed linkage to the FOA phenotype. Our results, however, might indicate a possible modulating role for variants in or near *NRP2* and *IDH1*. Further mutation analysis of the linkage area on chromosome 2q33.3-2q34 and confirmation of the most promising variants in other populations with advanced OA is needed.

## Acknowledgements

We would like to thank Dr. J.K. van der Korst (J. van Breemen Institute, Amsterdam), Dr. P.J.J.M. Rompa (Gooi Noord, Blaricum), Prof. Dr. J.J. Rasker (Med. Spectrum Twente, Enschede), W. Hissink-Muller, A.J.M. Schoffelen and P.J.M. Konings de Waal Malefijt (Maria Hospital, Tilburg), Prof. dr. R.G.T. Geesink and Prof. A.J. vd Linden (Academic Hospital, Maastricht), Dr. M.C. Kraan and Dr. N. Riyazi (Leiden Medical Hospital, Leiden), Dr. E.J. van Langelaan (Rijnland Hospital, Leiderdorp), Dr. Ronday, (Leijenburg Hospital The Hague) and Dr. H. Knoblauch (Universitätsklinikum Charite, Berlin) for providing us with the radiographs and medical histories of their patients. This work was supported by the Leiden University Medical Centre, the Dutch Arthritis Association (RF 001-301), the Centre of Medical System Biology and the Netherlands organization of Scientific Research (MW 904-61-095) of the Netherlands.

## References


1. Spector TD, MacGregor AJ. Risk factors for osteoarthritis: genetics. Osteoarthritis Cartilage 2004; 12 Suppl A:S39-S44.
2. Loughlin J. Genetic epidemiology of primary osteoarthritis. Curr Opin Rheumatol 2001; 13(2):111-116.
3. Ingvarsson T, Stefansson SE, Gulcher JR, Jonsson HH, Jonsson H, Frigge ML et al. A large Icelandic family with early osteoarthritis of the hip associated with a susceptibility locus on chromosome 16p. Arthritis Rheum 2001; 44(11):2548-2555.
4. Jakkula E, Melkoniemi M, Kiviranta I, Lohiniva J, Raina SS, Perala M et al. The role of sequence variations within the genes encoding collagen II, IX and XI in non-syndromic, early-onset osteoarthritis. Osteoarthritis Cartilage 2005; 13(6):497-507.
5. Meulenbelt I, Bijkerk C, Breedveld FC, Slagboom PE. Genetic linkage analysis of 14 candidate gene loci in a family with autosomal dominant osteoarthritis without dysplasia. J Med Genet 1997; 34(12):1024-1027.
6. Briggs MD, Chapman KL. Pseudoachondroplasia and multiple epiphyseal dysplasia: mutation review, molecular interactions, and genotype to phenotype correlations. Hum Mutat 2002; 19(5):465-478.
7. Reginato AM, Olsen BR. The role of structural genes in the pathogenesis of osteoarthritic disorders. Arthritis Res 2002; 4(6):337-345.
8. Jakkula E, Melkoniemi M, Kiviranta I, Lohiniva J, Raina SS, Perala M et al. The role of sequence variations within the genes encoding collagen II, IX and XI in non-syndromic, early-onset osteoarthritis. Osteoarthritis Cartilage 2005; 13(6):497-507.
9. Ikeda T, Mabuchi A, Fukuda A, Kawakami A, Ryo Y, Yamamoto S et al. Association analysis of single nucleotide polymorphisms in cartilage-specific collagen genes with knee and hip osteoarthritis in the Japanese population. J Bone Miner Res 2002; 17(7):1290-1296.
10. Loughlin J, Dowling B, Mustafa Z, Chapman K. Association of the interleukin-1 gene cluster on chromosome 2q13 with knee osteoarthritis. Arthritis Rheum 2002; 46(6):1519-1527.
11. Meulenbelt I, Seymour AB, Nieuwland M, Huizinga TW, Van Duijn CM, Slagboom PE. Association of the interleukin-1 gene cluster with radiographic signs of osteoarthritis of the hip. Arthritis Rheum 2004; 50(4):1179-1186.
12. Smith AJ, Keen LJ, Billingham MJ, Perry MJ, Elson CJ, Kirwan JR et al. Extended haplotypes and linkage disequilibrium in the IL1R1-IL1A-IL1B-IL1RN gene cluster: association with knee osteoarthritis. Genes Immun 2004; 5(6):451-460.
13. Bergink AP, van Meurs JB, Loughlin J, Arp PP, Fang Y, Hofman A et al. Estrogen receptor alpha gene haplotype is associated with radiographic osteoarthritis of the knee in elderly men and women. Arthritis Rheum 2003; 48(7):1913-1922.
14. Jin SY, Hong SJ, Yang HI, Park SD, Yoo MC, Lee HJ et al. Estrogen receptor-alpha gene haplotype is associated with primary knee osteoarthritis in Korean population. Arthritis Res Ther 2004; 6(5):R415-R421.
15. Keen RW, Hart DJ, Lanchbury JS, Spector TD. Association of early osteoarthritis of the knee with a Taq I polymorphism of the vitamin D receptor gene. Arthritis Rheum 1997; 40(8):1444-1449.
16. Uitterlinden AG, Burger H, Huang Q, Odding E, Duijn CM, Hofman A et al. Vitamin D receptor genotype is associated with radiographic osteoarthritis at the knee. J Clin Invest 1997; 100(2):259-263.
17. Min JL, Meulenbelt I, Riyazi N, Kloppenburg M, Houwing-Duistermaat JJ, Seymour AB et al. Association of matrilin-3 polymorphisms with spinal disc degeneration and with osteoarthritis of the CMC1 joint of the hand. Ann Rheum Dis 2006.
18. Stefansson SE, Jonsson H, Ingvarsson T, Manolescu I, Jonsson HH, Olafsdottir G et al. Genomewide scan for hand osteoarthritis: a novel mutation in matrilin-3. Am J Hum Genet 2003; 72(6):1448-1459.

19. Loughlin J, Dowling B, Chapman K, Marcelline L, Mustafa Z, Southam L et al. Functional variants within the secreted frizzled-related protein 3 gene are associated with hip osteoarthritis in females. Proc Natl Acad Sci U S A 2004; 101(26):9757-9762.
20. Min JL, Meulenbelt I, Riyazi N, Kloppenburg M, Houwing-Duistermaat JJ, Seymour AB et al. Association of the Frizzled-related protein gene with symptomatic osteoarthritis at multiple sites. Arthritis Rheum 2005; 52(4):1077-1080.
21. Forster T, Chapman K, Loughlin J. Common variants within the interleukin 4 receptor alpha gene (IL4R) are associated with susceptibility to osteoarthritis. Hum Genet 2004; 114(4):391-395.
22. Kizawa H, Kou I, Iida A, Sudo A, Miyamoto Y, Fukuda A et al. An aspartic acid repeat polymorphism in asporin inhibits chondrogenesis and increases susceptibility to osteoarthritis. Nat Genet 2005; 37(2):138-144.
23. Mototani H, Mabuchi A, Saito S, Fujioka M, Iida A, Takatori Y et al. A functional single nucleotide polymorphism in the core promoter region of CALM1 is associated with hip osteoarthritis in Japanese. Hum Mol Genet 2005; 14(8):1009-1017.
24. Meulenbelt I, Min JL, Van Duijn CM, Kloppenburg M, Breedveld FC, Slagboom PE. Strong linkage on 2q33.3 to familial early onset generalized osteoarthritis and a consideration of two candidate genes. Eur J Hum Genet. In press.
25. Loughlin J, Dowling B, Chapman K, Marcelline L, Mustafa Z, Southam L et al. Functional variants within the secreted frizzled-related protein 3 gene are associated with hip osteoarthritis in females. Proc Natl Acad Sci U S A 2004; 101(26):9757-9762.
26. Min JL, Meulenbelt I, Riyazi N, Kloppenburg M, Houwing-Duistermaat JJ, Seymour AB et al. Association of the Frizzled-related protein gene with symptomatic osteoarthritis at multiple sites. Arthritis Rheum 2005; 52(4):1077-1080.
27. Kellgren JH, Jeffrey MR, Ball J. The epidemiology of chronic rheumatism. Volume II: Atlas of standard radiographs of arthritis. Oxford: Blackwell Scientific Publications, 1963.
28. Meulenbelt I, Bijkerk C, Breedveld FC, Slagboom PE. Genetic linkage analysis of 14 candidate gene loci in a family with autosomal dominant osteoarthritis without dysplasia. J Med Genet 1997; 34(12):1024-1027.
29. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol 1991; 7(4):403-422.
30. Odding E, Valkenburg HA, Stam HJ, Hofman A. Determinants of locomotor disability in people aged 55 years and over: the Rotterdam Study. Eur J Epidemiol 2001; 17(11):1033-1041.
31. Bijkerk C, Houwing-Duistermaat JJ, Valkenburg HA, Meulenbelt I, Hofman A, Breedveld FC et al. Heritabilities of radiologic osteoarthritis in peripheral joints and of disc degeneration of the spine. Arthritis Rheum 1999; 42(8):1729-1735.
32. Min JL, Meulenbelt I, Riyazi N, Kloppenburg M, Houwing-Duistermaat JJ, Seymour AB et al. Association of the Frizzled-related protein gene with symptomatic osteoarthritis at multiple sites. Arthritis Rheum 2005; 52(4):1077-1080.
33. Vieux EF, Kwok PY, Miller RD. Primer design for PCR and sequencing in high-throughput analysis of SNPs. Biotechniques 2002; Suppl:28-30, 32.
34. den Dunnen JT, Antonarakis SE. Nomenclature for the description of human sequence variations. Hum Genet 2001; 109(1):121-124.
35. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 2002; 12(3):436-446.
36. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002; 30(17):3894-3900.
37. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 2002; 12(3):436-446.
38. Wang J, Smith PJ, Krainer AR, Zhang MQ. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. Nucleic Acids Res 2005; 33(16):5053-5062.
39. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol 1991; 7(4):403-422.

40. Hutloff A, Dittrich AM, Beier KC, Eljaschewitsch B, Kraft R, Anagnostopoulos I et al. ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. Nature 1999; 397(6716):263-266.
41. Rudd CE, Schneider H. Unifying concepts in CD28, ICOS and CTLA4 co-receptor signalling. Nat Rev Immunol 2003; 3(7):544-556.
42. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 2003; 423(6939):506-511.
43. Loughlin J, Dowling B, Chapman K, Marcelline L, Mustafa Z, Southam L et al. Functional variants within the secreted frizzled-related protein 3 gene are associated with hip osteoarthritis in females. Proc Natl Acad Sci U S A 2004; 101(26):9757-9762.
44. Min JL, Meulenbelt I, Riyazi N, Kloppenburg M, Houwing-Duistermaat JJ, Seymour AB et al. Association of the Frizzled-related protein gene with symptomatic osteoarthritis at multiple sites. Arthritis Rheum 2005; 52(4):1077-1080.
45. Geisbrecht BV, Gould SJ. The human PICD gene encodes a cytoplasmic and peroxisomal NADP(+)-dependent isocitrate dehydrogenase. J Biol Chem 1999; 274(43):30527-30533.
46. Gerber HP, Vu TH, Ryan AM, Kowalski J, Werb Z, Ferrara N. VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation. Nat Med 1999; 5(6):623-628.
47. Gluzman-Poltorak Z, Cohen T, Herzog Y, Neufeld G. Neuropilin-2 is a receptor for the vascular endothelial growth factor (VEGF) forms VEGF-145 and VEGF-165 [corrected]. J Biol Chem 2000; 275(24):18040-18045.
48. Enomoto H, Inoki I, Komiya K, Shiomi T, Ikeda E, Obata K et al. Vascular endothelial growth factor isoforms and their receptors are expressed in human osteoarthritic cartilage. Am J Pathol 2003; 162(1):171-181.
49. Pufe T, Harde V, Petersen W, Goldring MB, Tillmann B, Mentlein R. Vascular endothelial growth factor (VEGF) induces matrix metalloproteinase expression in immortalized chondrocytes. J Pathol 2004; 202(3):367-374.
50. Sbrissa D, Ikonomov OC, Shisheva A. Phosphatidylinositol 3-phosphate-interacting domains in PIKfyve. Binding specificity and role in PIKfyve. Endomenbrane localization. J Biol Chem 2002; 277(8):6073-6079.
51. Kil IS, Huh TL, Lee YS, Lee YM, Park JW. Regulation of replicative senescence by NADP(+)-dependent isocitrate dehydrogenase. Free Radic Biol Med 2006; 40(1):110-119.
52. Enomoto H, Inoki I, Komiya K, Shiomi T, Ikeda E, Obata K et al. Vascular endothelial growth factor isoforms and their receptors are expressed in human osteoarthritic cartilage. Am J Pathol 2003; 162(1):171-181.
53. Gluzman-Poltorak Z, Cohen T, Herzog Y, Neufeld G. Neuropilin-2 is a receptor for the vascular endothelial growth factor (VEGF) forms VEGF-145 and VEGF-165 [corrected]. J Biol Chem 2000; 275(24):18040-18045.
54. Gerber HP, Vu TH, Ryan AM, Kowalski J, Werb Z, Ferrara N. VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation. Nat Med 1999; 5(6):623-628.
55. Enomoto H, Inoki I, Komiya K, Shiomi T, Ikeda E, Obata K et al. Vascular endothelial growth factor isoforms and their receptors are expressed in human osteoarthritic cartilage. Am J Pathol 2003; 162(1):171-181.
56. Pufe T, Harde V, Petersen W, Goldring MB, Tillmann B, Mentlein R. Vascular endothelial growth factor (VEGF) induces matrix metalloproteinase expression in immortalized chondrocytes. J Pathol 2004; 202(3):367-374.
57. Tanaka E, Aoyama J, Miyauchi M, Takata T, Hanaoka K, Iwabe T et al. Vascular endothelial growth factor plays an important autocrine/paracrine role in the progression of osteoarthritis. Histochem Cell Biol 2005; 123(3):275-281.

58. Heon E, Priston M, Schorderet DF, Billingsley GD, Girard PO, Lubsen N et al. The gamma-crystallins and human cataracts: a puzzle made clearer. Am J Hum Genet 1999; 65(5):1261-1267.
59. Kohjima M, Noda Y, Takeya R, Saito N, Takeuchi K, Sumimoto H. PAR3beta, a novel homologue of the cell polarity protein PAR3, localizes to tight junctions. Biochem Biophys Res Commun 2002; 299(4):641-646.