RESEARCH ARTICLE

Identification of a Novel Mutation in the *COL2A1* Gene in a Chinese Family with Spondyloepiphyseal Dysplasia Congenita

Xiangjun Huang[1], Xiong Deng[1], Hongbo Xu[1], Song Wu[2], Lamei Yuan[1], Zhijian Yang[1], Yan Yang[1], Hao Deng[1] *

1 Center for Experimental Medicine, the Third Xiangya Hospital, Central South University, Changsha, China,
2 Department of Orthopedics, the Third Xiangya Hospital, Central South University, Changsha, China

* hdeng008@yahoo.com

Abstract

Spondyloepiphyseal dysplasia congenita (SEDC) is an autosomal dominant chondrodysplasia characterized by disproportionate short-trunk dwarfism, skeletal and vertebral deformities. Exome sequencing and Sanger sequencing were performed in a Chinese Han family with typical SEDC, and a novel mutation, c.620G>A (p.Gly207Glu), in the collagen type II alpha-1 gene (*COL2A1*) was identified. The mutation may impair protein stability, and lead to dysfunction of type II collagen. Family-based study suggested that the mutation is a *de novo* mutation. Our study extends the mutation spectrum of SEDC and confirms genotype-phenotype relationship between mutations at glycine in the triple helix of the alpha-1 (II) chains of the *COL2A1* and clinical findings of SEDC, which may be helpful in the genetic counseling of patients with SEDC.

Introduction

Spondyloepiphyseal dysplasia congenita (SEDC, OMIM 183900) is a common subtype of the type II collagenopathies, a heterogeneous group of chondrodysplasias resulting from mutations in the collagen type II alpha-1 gene (*COL2A1*, OMIM 120140) [1]. The disorder was first described in 1966 with a low prevalence of 3.4/1,000,000 [2,3]. It is characterized by disproportionate short-trunk dwarfism, skeletal and vertebral deformities, including scoliosis and thoracic hyperkyphosis, coxa vara, avascular necrosis-like changes in bilateral femoral epiphyses, hip pain or waddling gait, genu valgum, and various joint diseases [3–5]. Skeletal deformities are the most commonly shared features in SEDC. Extraskeletal manifestations, including hearing loss, cleft palate and ocular anomalies, have also been reported in some cases [3,6]. To date, mutations in the *COL2A1* gene have been reported to be responsible for at least 16 subtypes of type II collagenopathies, including achondrogenesis, type II or hypochondrogenesis (ACG2/HCG, OMIM 200610), platyspondylic skeletal dysplasia, Torrance type (PLSDT, OMIM 151210), spondyloperipheral dysplasia (SPD, OMIM 271700), SEDC, spondyloepimetaphyseal dysplasia, Strudwick type (SEMDSTWK, OMIM 184250), Kniest dysplasia (OMIM

156550), Stickler syndrome, type I (STL1, OMIM 108300), Stickler syndrome, type I, nonsyndromic ocular (OMIM 609508), osteoarthritis with mild chondrodysplasia (OMIM 604864), avascular necrosis of the femoral head (ANFH, OMIM 608805), Legg-Calve-Perthes disease (LCPD, OMIM 150600), epiphyseal dysplasia, multiple with myopia and deafness (EDMMD, OMIM 132450), otospondylomegaepiphyseal dysplasia (OSMED, OMIM 215150), Czech dysplasia (OMIM 609162), spondyloepiphyseal dysplasia, Namaqualand type (SEDN), and vitreoretinopathy with phalangeal epiphyseal dysplasia (VPED) [7–9].

The aim of this study was to identify causal gene responsible for familial SEDC in a four-generation Chinese pedigree. A novel *COL2A1* gene mutation, c.620G>A (p.Gly207Glu), was identified, resulting in a change of conserved glycine in the triple helix of the alpha-1(II) chains. Substitution of the highly conserved amino acid may impair protein stability and lead to dysfunction of helical structure and type II collagen [10].

Materials and Methods

Participators and clinical evaluation

The participating individuals or their guardians provided written informed consent for this study and the approval for the research was received from the Ethics Committee of the Third Xiangya Hospital, Central South University, China. A four-generation, 16-member Chinese Han family with familial SEDC was recruited from the Third Xiangya Hospital, Central South University, China (Fig 1). Peripheral blood samples were taken from 12 family members, and detailed clinical records of medical history, physical examinations and laboratory analyses were obtained from available family members. Blood samples were also collected from 100 unrelated ethnically-matched healthy control volunteers with negative family history of this disorder (male/female: 50/50, age 32.2±8.3 years).
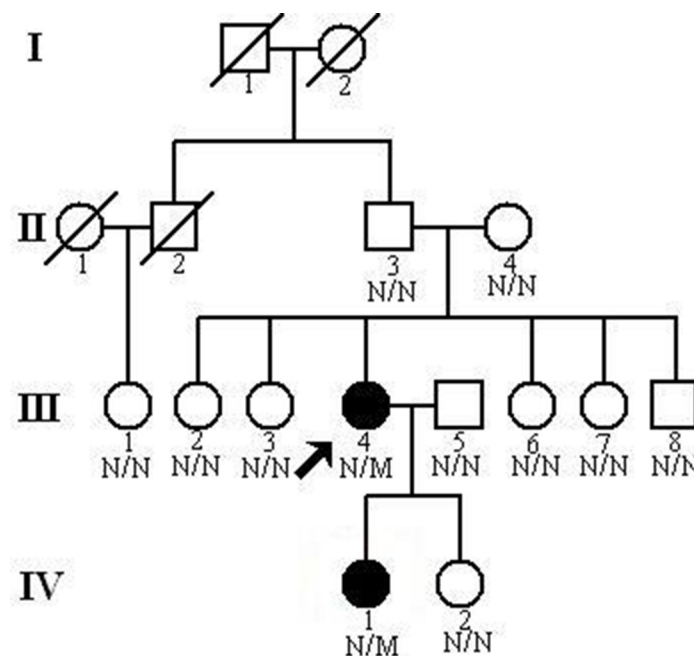

Fig 1. Pedigree of the family with spondyloepiphyseal dysplasia congenita showing affected cases (fully shaded). N: normal; M: the *COL2A1* c.620G>A (p.Gly207Glu) mutation.

doi:10.1371/journal.pone.0127529.g001

Exome capture

Extraction of genomic DNA was performed by using standard phenol-chloroform extraction method from peripheral blood lymphocytes [11]. Three micrograms (μg) of genomic DNA was used to establish the exome library. Genomic DNA of the proband (III:4) was randomly sheared by sonication and then hybridized to the Nimblegen SeqCap EZ Library to enrich exonic DNA in each library, following the manufacturer's instructions. Enriched library targeting the exome was sequenced on the Illumina HiSeq 2000 platform to generate 90-bp paired-end reads [12]. A mean exome coverage of 77.52× was obtained, allowing for the exam of selected region and sufficient depth to accurately match 99.46% of targeted exome [11].

Read mapping and variant analysis

The available human reference genomic sequences were accessed from the public online UCSC database (http://genome.ucsc.edu/), version hg19 (build 37.1). Sequences obtained from the proband were aligned using SOAPaligner (soap2.21), and single nucleotide polymorphisms (SNPs) were sifted out using SOAPsnp with the defined parameters after the deletion of duplicated reads (obtained mainly in the PCR step). Small insertions or deletions (indels) in the coding sequence and splicing sites were prosecuted [12–14]. Candidate SNPs were filtered by the following criteria: SNP quality $\geq$20 (score 20 represents 99% accuracy of a base call), sequencing depth $\geq$4, the estimated copy number $\leq$2, and the distance between two SNPs >5 [15]. The common variants and non-pathogenic candidate variants were filtered against several public databases including the Single Nucleotide Polymorphism database (dbSNP build 137, http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi), 1000 genomes project (1000genomes release_20100804, http://www.1000genomes.org/), HapMap project (2010–08_phase II + III, http://hapmap.ncbi.nlm.nih.gov/), and YanHuang (http://yh.genomics.org.cn/) project.

Multiple alignment of the protein sequence was performed across different species by the Basic Local Alignment Search Tool (http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi). Online tools including Polymorphism Phenotyping version 2 (PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/), Sorting Intolerant From Tolerant (SIFT, http://sift.jcvi.org/, score less than 0.05 is deleterious) and MutationTaster (http://www.mutationtaster.org/) were used to evaluate whether amino acid substitutions affect protein function [16–18]. Locus-specific primers used for PCR amplification and direct sequencing were designed using the online Primer3 program (http://primer3.ut.ee/). Then Sanger sequencing of PCR products was performed to validate the potential disease-causing variants with ABI3500 sequencer (Applied Biosystems Inc., Foster City, CA, USA), according to the procedures described previously [11]. Primer sequences used for the *COL2A1* gene causative variant were as follows: 5′–GCTTGGGAATCATCTGCGAC–3′ and 5′–TGGGAAATGAGAGGGAGCAG–3′.

Results

Clinical findings

Affected individuals (III:4 and IV:1) showed similar clinical and radiological abnormalities. Patients had short-trunk dwarfism, short neck, spinal and limbs deformities, coxa vara, genu valgum, waddling gait and various joint diseases, especially loss of complete bone formation of femoral head. Short-trunk dwarfism was noticed at birth in these two patients, and the growth was found to be markedly delayed. While cleft palate, ocular complications, hearing loss and other tube malformation were absent in both patients. Skeletal radiographs showed odontoid hypoplasia, atlantoaxial subluxation, scoliosis, kyphosis, lumbar lordosis and ovoid vertebral bodies, capital femoral epiphyses dysplasia, flattening of the acetabular roof, shortening of the

femoral neck and base of the ilium, and delayed fusion of the pubic and ischial bones. No SEDC-related clinical or radiological abnormalities were observed in other recruited family members.

Mutation screening

We performed exome sequencing on genomic DNA sample of the proband (III:4) of the Chinese Han family with SEDC. A total of 6.80 billion bases of sequence with a 90-bp read length from the patient were generated, and 6.62 billion bases (97.38%) passed the quality assessment. Among them, 6.19 billion bases (93.50%) were aligned to the human reference sequence and 4.33 billion bases covered the target region with a mean coverage of 77.52-fold [12]. 110,479 genetic variants, including 14,897 non-synonymous changes, were occurred at the coding sequence or the canonical dinucleotide of the splice site junctions. A filtration strategy was performed to detect the responsible mutation for the disorder, following the priorization scheme used in recent studies [11]. Given that the disorder frequency is <0.005, we excluded known variants recorded in dbSNP137, 1000 genomes project, HapMap or Yanhuang projects with MAF >0.50%. Using the above filtering strategy, the number of candidate genes was reduced by more than 93.35%. Sequence variants were also filtered by in-house database from BGI-Shenzhen. Subsequently, a missense mutation in the *COL2A1* gene, the potential disease-causing gene of SEDC, was selected for further validation.

After validation by Sanger sequencing and comparison with mutation from the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/), a novel variant, c.620G>A (p.Gly207-Glu), in the *COL2A1* gene was observed in the proband. The c.620G>A variant was subsequently identified in her affected daughter (IV:1) (Fig 2), and validated for the absence in unaffected family individuals and 100 unrelated normal controls. The co-segregation of variant with this disease and the bioinformation analysis suggest that this variant is likely the pathogenic mutation. The glycine at position 207 is phylogenetically conserved among various species based on multiple sequence alignment (Fig 3). PolyPhen-2 analysis predicted to be probably damaging with a score of 1.00 on HumVar database (sensitivity: 0.00; specificity: 1.00). The SIFT prediction also showed a damaging effect with a score of 0.00. MutationTaster
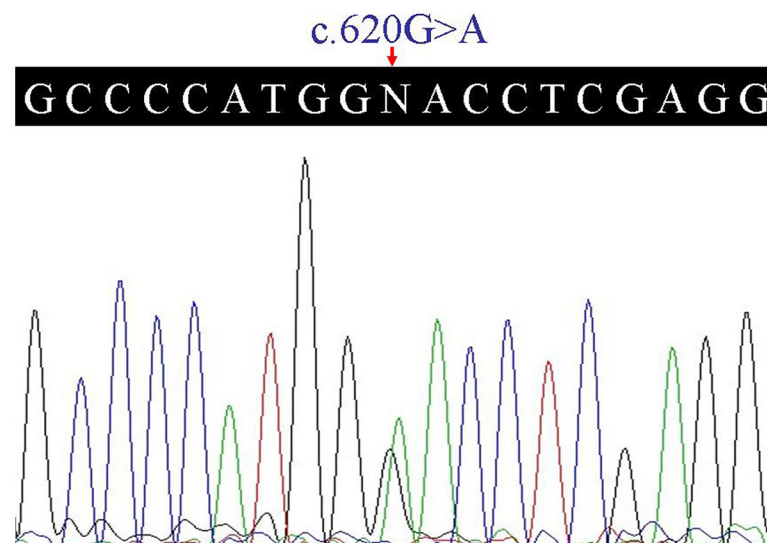

Fig 2. Heterozygous c.620G>A (p.Gly207Glu) mutation in the *COL2A1* gene.

doi:10.1371/journal.pone.0127529.g002

p.Gly207

Human	V M Q G P M G P M **G** P R G P P G P A G
Chimpanzee	V M Q G P M G P M **G** P R G P P G P A G
Rhesus monkey	V M Q G P M G P M **G** P R G P P G P A G
Dog	V M Q G P M G P M **G** P R G P P G P A G
Cattle	V M Q G P M G P M **G** P R G P P G P A G
Horse	V M Q G P M G P M **G** P R G P P G P A G
Sheep	V M Q G P M G P M **G** P R G P P G P A G
Mouse	V M Q G P M G P M **G** P R G P P G P A G
Rat	V M Q G P M G P M **G** P R G P P G P A G
Chicken	V M Q G P M G P M **G** P R G P P G P T G

Fig 3. Conservation analysis of the COL2A1 p.Gly207 amino acid residue.

doi:10.1371/journal.pone.0127529.g003

predicted that the alteration was disease-causing with a probability value close to 1 indicating the high security of prediction. Computer-based protein analysis indicates that the variant in the *COL2A1* gene was likely deleterious and the disease-causing mutation in our family.

Discussion

SEDC, a rare but differentiable skeletal malformation caused by mutations in the *COL2A1* gene, is usually evidenced at birth [19]. It is characterized by short-trunk dwarfism, abnormal epiphyses, and platyspondyly [20,21]. SEDC, including both mild and severe coxa vara (SEDC-M and SEDC-S), is the most severe form of hypogenesis of the femoral head. SEDC can be distinguished from other diseases by its characteristic manifestations, including platyspondyly, odontoid hypoplasia, normal hands and feet, and extraskeletal features, including myopia, sensor neural hearing loss, bifid uvula, cleft palate and hypospadias [6,7]. It is pathologically characterized by PAS-positive cytoplasmic inclusions and cystic spaces in resting cartilage [22]. Though SEDC is an autosomal dominant skeletal disorder, most cases of SEDC are sporadic [6]. Clinical variability exists in the SEDC, even the same point mutation can cause different manifestations [20].

In this study, a heterozygous missense mutation c.620G>A (p.Gly207Glu), in the *COL2A1* gene was identified in a Chinese Han family with SEDC by exome sequencing and Sanger sequencing. Two patients (III:4 and IV:1, Fig 1) showed typical clinical features of SEDC-M, including short-trunk dwarfism, spinal and limbs deformities. Neither patient showed any abnormalities in the eye, auditory system or other tube bones. The mutation co-segregated with the disorder in the family and was absent in unaffected family members, 100 normal controls, and public databases. Multiple sequence alignment showed that the mutated glycine at position 207 is phylogenetically conserved, and computer-based analysis revealed that the mutation is likely deleterious, supporting that the p.Gly207Glu variant is likely pathogenic. Given that both parents (II:3 and II:4, Fig 1) of the proband were free of p.Gly207Glu mutation and had no clinical evidence for SEDC, the mutation was thought to be a *de novo* mutation.

The *COL2A1* gene, located on chromosome 12q13, consists of 54 exons spanning over 31.5 kb of human genomic DNA and encodes a 134.4 kDa protein with 1,487 amino acids. The collagen triple helical domain characterized by three identical polypeptide chains (Gly-X-Y) is the backbone of the COL2A1 molecule, and the alpha-chain helical region is composed of 1,014

amino acids with glycine occupying every third positions, critical to correct folding and stability of the helix [23,24]. The protein is synthesized as large procollagen chains containing N- and C-terminal amino acid sequences named propeptides [25], and the C-terminal non-collagenous domain (designated as C-propeptide), which is necessary for chain association and subsequent triple helix formation [26]. After secreted into the extracellular matrix, the propeptides are cleaved and form the mature type II collagen [25]. Type II collagen co-aggregates with collagen XI, which is covalently linked to collagen IX and interacts with small leucine-rich proteoglycans influencing collagen fibrillar structure and function [23]. The p.Gly207Glu mutation led to change of a neutral amino acid residue, glycine, to a negatively charged residue, glutamic acid, involving in the basic repeating unit of the triple helical domain, the Gly position of a Gly-X-Y triplet. The substitution may impair protein stability and thus helical structure and proper function of the protein, leading to the disorder [10,27,28].

Up to now, including the p.Gly207Glu mutation, at least 35 different mutations in the *COL2A1* gene have been described to be related to SEDC with a broad phenotypic spectrum in various ethnic groups. These mutations include missense, deletion, duplication and splicing mutations, and the p.R989C mutation appears to be a hot spot mutation with a strikingly similar phenotype [1,5,29]. The accurate relationship between the *COL2A1* gene mutations and corresponding phenotypes is far from clear. The expression profiles and characteristics of mutations can also be determined by genetic background [21].

The *COL2A1* mutations most likely produce structurally abnormal type II collagen in the extracellular matrix, hampering endochondral ossification and linear bone growth by a dominant negative mechanism [30]. The glycine mutation in the triple-helical domain may harm the intracellular transport and collagen secretion, and glycine-to-nonserine residue substitutions have been reported to create more severe phenotypes, consistent with the observation in our patients [7,31].

Conclusions

A novel *de novo* p.Gly207Glu mutation was identified in our family with SEDC, which was suggested to be the genetic cause and extends the mutation spectrum of the *COL2A1* gene. Whole exome sequencing provides a cost-effective and expedited approach to identify pathogenic mutations responsible for inherited disorders. Our finding may provide new insights and approaches into the genetic cause and diagnosis of SEDC, and may also have implications for genetic counseling and clinical intervention. More extensive studies of the *COL2A1* including in larger and diverse ethnic groups are warranted to explore the underlying pathogenic mechanism of SEDC and elucidate the potential genotype-phenotype relationship, which in turn may supplement our understanding of type II collagenopathies [3,6].

Acknowledgments

Author Contributions

Conceived and designed the experiments: XH HD. Performed the experiments: XH XD HX SW ZY YY HD. Analyzed the data: XH XD HX LY HD. Contributed reagents/materials/analysis tools: XH XD HX SW LY ZY YY HD. Wrote the paper: XH HD.

References

1. Li S, Zhou H, Qin H, Guo H, Bai Y. A novel mutation in the COL2A1 gene in a Chinese family with Spondyloepiphyseal dysplasia congenita. Joint Bone Spine. 2014; 81: 86–89. doi: 10.1016/j.jbspin.2013.06.010 PMID: 23932928

2. Spranger JW, Wiedemann HR. Dysplasia spondyloepiphysaria congenita. Helv Paediatr Acta. 1966; 21: 598.

3. Xu L, Qiu X, Zhu Z, Yi L, Qiu Y. A novel mutation in COL2A1 leading to spondyloepiphyseal dysplasia congenita in a three-generation family. Eur Spine J. 2014; 23 Suppl 2: 271–277. doi: 10.1007/s00586-014-3292-0 PMID: 24736929

4. Veeravagu A, Lad SP, Camara-Quintana JQ, Jiang B, Shuer L. Neurosurgical interventions for spondyloepiphyseal dysplasia congenita: clinical presentation and assessment of the literature. World Neurosurg. 2013; 80: 431–437.

5. Mark PR, Torres-Martinez W, Lachman RS, Weaver DD. Association of a p.Pro786Leu variant in COL2A1 with mild spondyloepiphyseal dysplasia congenita in a three-generation family. Am J Med Genet A. 2011; 155A: 174–179. doi: 10.1002/ajmg.a.34354 PMID: 22106036

6. Jung SC, Mathew S, Li QW, Lee YJ, Lee KS, Song HR. Spondyloepiphyseal dysplasia congenita with absent femoral head. J Pediatr Orthop B. 2004; 13: 63–69. PMID: 15076581

7. Nishimura G, Haga N, Kitoh H, Tanaka Y, Sonoda T, Kitamura M, et al. The phenotypic spectrum of COL2A1 mutations. Hum Mutat. 2005; 26: 36–43. PMID: 15895462

8. Nagendran S, Richards AJ, McNinch A, Sandford RN, Snead MP. Somatic mosaicism and the phenotypic expression of COL2A1 mutations. Am J Med Genet A. 2012; 158A: 1204–1207. doi: 10.1002/ajmg.a.35303 PMID: 22496037

9. Kozlowski K, Marik I, Marikova O, Zemkova D, Kuklik M. Czech dysplasia metatarsal type. Am J Med Genet A. 2004; 129A: 87–91. PMID: 15266623

10. Husar-Memmer E, Ekici A, Al KA, Sticht H, Manger B, Schett G, et al. Premature osteoarthritis as presenting sign of type II collagenopathy: a case report and literature review. Semin Arthritis Rheum. 2013; 42: 355–360. doi: 10.1016/j.semarthrit.2012.05.002 PMID: 22717203

11. Guo Y, Yuan J, Liang H, Xiao J, Xu H, Yuan L, et al. Identification of a novel COL4A5 mutation in a Chinese family with X-linked Alport syndrome using exome sequencing. Mol Biol Rep. 2014; 41: 3631–3635. doi: 10.1007/s11033-014-3227-1 PMID: 24522658

12. Wang JL, Cao L, Li XH, Hu ZM, Li JD, Zhang JG, et al. Identification of PRRT2 as the causative gene of paroxysmal kinesigenic dyskinesias. Brain. 2011; 134: 3493–3501. doi: 10.1093/brain/awr289 PMID: 22120146

13. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. Bioinformatics. 2008; 24: 713–714. doi: 10.1093/bioinformatics/btn025 PMID: 18227114

14. Shi Y, Li Y, Zhang D, Zhang H, Li Y, Lu F, et al. Exome sequencing identifies ZNF644 mutations in high myopia. PLoS Genet. 2011; 7: e1002084. doi: 10.1371/journal.pgen.1002084 PMID: 21695231

15. Xiu X, Yuan J, Deng X, Xiao J, Xu H, Zeng Z, et al. A novel COL4A5 mutation identified in a Chinese Han family using exome sequencing. Biomed Res Int. 2014; 2014: 186048. doi: 10.1155/2014/186048 PMID: 25110662

16. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31: 3812–3814. PMID: 12824425

17. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010; 42: 30–35. doi: 10.1038/ng.499 PMID: 19915526

18. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010; 7: 575–576. doi: 10.1038/nmeth0810-575 PMID: 20676075

19. Unger S, Korkko J, Krakow D, Lachman RS, Rimoin DL, Cohn DH. Double heterozygosity for pseudoachondroplasia and spondyloepiphyseal dysplasia congenita. Am J Med Genet. 2001; 104: 140–146. PMID: 11746045

20. Xia XY, Cui YX, Huang YF, Pan LJ, Feng Y, Yang B, et al. Molecular prenatal diagnosis in 2 pregnancies at risk for spondyloepiphyseal dysplasia congenita. Clin Chim Acta. 2008; 387: 153–157. PMID: 17920052

21. Zhang Z, He JW, Fu WZ, Zhang CQ, Zhang ZL. Identification of three novel mutations in the COL2A1 gene in four unrelated Chinese families with spondyloepiphyseal dysplasia congenita. Biochem Biophys Res Commun. 2011; 413: 504–508. doi: 10.1016/j.bbrc.2011.08.090 PMID: 21924244

22. Turner LM, Steffensen TS, Leroy J, Gilbert-Barness E. Spondyloepiphyseal dysplasia congenita. Fetal Pediatr Pathol. 2010; 29: 57–62. doi: 10.3109/15513810903266310 PMID: 20055564

23. Kannu P, Bateman J, Savarirayan R. Clinical phenotypes associated with type II collagen mutations. J Paediatr Child Health. 2012; 48: E38–E43. doi: 10.1111/j.1440-1754.2010.01979.x PMID: 21332586

24. Jamieson SE, de Roubaix LA, Cortina-Borja M, Tan HK, Mui EJ, Cordell HJ, et al. Genetic and epigenetic factors at COL2A1 and ABCA4 influence clinical outcome in congenital toxoplasmosis. PLoS One. 2008; 3: e2285. doi: 10.1371/journal.pone.0002285 PMID: 18523590

25. Strom CM, Upholt WB. Isolation and characterization of genomic clones corresponding to the human type II procollagen gene. Nucleic Acids Res. 1984; 12: 1025–1038. PMID: 6320112

26. Desir J, Cassart M, Donner C, Coucke P, Abramowicz M, Mortier G. Spondyloperipheral dysplasia as the mosaic form of platyspondylic lethal skeletal dyplasia torrance type in mother and fetus with the same COL2A1 mutation. Am J Med Genet A. 2012; 158A: 1948–1952. doi: 10.1002/ajmg.a.35301 PMID: 22495950

27. Chan D, Taylor TK, Cole WG. Characterization of an arginine 789 to cysteine substitution in alpha 1 (II) collagen chains of a patient with spondyloepiphyseal dysplasia. J Biol Chem. 1993; 268: 15238–15245. PMID: 8325895

28. Zhang L, Qu X, Su S, Guan L, Liu P. A novel mutation in GJA3 associated with congenital Coppock-like cataract in a large Chinese family. Mol Vis. 2012; 18: 2114–2118. PMID: 22876138

29. Tiller GE, Weis MA, Polumbo PA, Gruber HE, Rimoin DL, Cohn DH, et al. An RNA-splicing mutation (G +5IVS20) in the type II collagen gene (COL2A1) in a family with spondyloepiphyseal dysplasia congenita. Am J Hum Genet. 1995; 56: 388–395. PMID: 7847372

30. Hoornaert KP, Dewinter C, Vereecke I, Beemer FA, Courtens W, Fryer A, et al. The phenotypic spectrum in patients with arginine to cysteine mutations in the COL2A1 gene. J Med Genet. 2006; 43: 406–413. PMID: 16155195

31. Freisinger P, Ala-Kokko L, LeGuellec D, Franc S, Bouvier R, Ritvaniemi P, et al. Mutation in the COL2A1 gene in a patient with hypochondrogenesis. Expression of mutated COL2A1 gene is accompanied by expression of genes for type I procollagen in chondrocytes. J Biol Chem. 1994; 269: 13663–13669. PMID: 8175802