Autosomal recessive chondrodysplasia with severe short stature caused by a biallelic

COL10A1 **variant**

Noor ul Ain[1], Outi Makitie[2,3,4], Sadaf Naz[1]*

1. School of Biological Sciences, University of the Punjab, Lahore, Pakistan

2. Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

3. Folkhälsan Institute of Genetics, Helsinki, Finland

4. Department of Molecular Medicine and Surgery and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden

***Correspondence to**

Dr. Sadaf Naz, School of Biological Sciences, University of the Punjab

Quaid-i-Azam Campus, Lahore, 54590, Pakistan

naz.sbs@pu.edu.pk

ABSTRACT

Background Heterozygous mutations in *COL10A1* underlie metaphyseal chondrodysplasia, Schmid type (MCDS), an autosomal dominant skeletal dysplasia.

Objective To identify the causative variant in a large consanguineous Pakistani family with severe skeletal dysplasia and marked lower-limb deformity.

Methods Whole-Exome sequencing was completed followed by Sanger sequencing to verify segregation of the identified variants. *In silico* variant pathogenicity predictions and amino acid conservation analyses were performed.

Results A homozygous c.133 C>T (p.Pro45Ser) variant was identified in *COL10A1* in all six severely affected individuals (adult heights 119 cm-130 cm, mean~-6.33 SD). The individuals heterozygous for the variant had mild phenotype of short stature (adult heights 140 cm-162 cm, mean~-2.15 SD) but no apparent skeletal deformities. The variant was predicted to be pathogenic by *in silico* prediction tools and was absent from public databases and hundred control chromosomes. Pro45 is conserved in orthologues and is located in the non-collagenous 2 domain of COL10A1, variants of which have never been associated with skeletal dysplasia.

Conclusions This first report of individuals with a homozygous variant in *COL10A1* defines a new type of autosomal recessive skeletal dysplasia. The observations in *COL10A1* variant carriers suggest a phenotypic overlap between the mildest forms of MCDS and idiopathic short stature.

INTRODUCTION

Metaphyseal chondrodysplasia, Schmid type (MCDS; OMIM 156500) is an autosomal dominant disorder. It is characterized by short stature, long bone deformities such as genua vara and coxa vara, and radiographic signs of methaphyseal dysplasia. Fifty heterozygous mutations in *COL10A1* encoding type X collagen have been reported to cause MCDS (Human Gene Mutation Database; http://www.hgmd.cf.ac.uk/ac, accessed in June 2017). Type X collagen is a homotrimer non-fibrillar collagen, composed of three alpha-1(X) chains of 680 amino acids. It is present in the extracellular matrix and is exclusively produced by the hypertrophic chondrocytes during the endochondral ossification.[1] The protein plays a role in distribution of proteoglycans and vesicles in bone matrix. COL10A1 is comprised of a signal peptide (amino acids 1-18) at the N-terminal followed by a non-collagenous 2 (NC2) region (19-56), a triple-helical region (57-519) and a non-collagenous 1 (NC1) region (520-680) at the C-terminal.[1]

All except three mutations[2 3] known to cause MCDS affect the NC1 region. Haploinsufficiency is one of the proposed mechanisms for the disorder since a 50% reduction of COL10A1 is observed in bone cartilage of MCDS patients who are heterozygous for some nonsense and frameshift mutations.[4 5] Dominant-negative effects of COL10A1 nonsense and missense variants have also been shown to be responsible for causing the disorder.[6] The NC1 region contains motifs required for trimerization of COL10A1 and a loss of NC1 leads to improper trimerization of collagen,[7] resulting in accumulation of these misfolded proteins in the cells. This leads to endoplasmic stress and apoptosis in hypertrophic chondrocytes during endochondral ossification.[8 9] However, mutations that result in a change of amino acid at the exterior of NC1 region (p.Asp617Lys and p.Gly618Val) allow trimerization but hinder supramolecular assembly and interaction of

COL10A1 with endoplasmic reticulum and proteins in extracellular matrix.[10][11] In addition, mutations of residues which introduce a charge deep inside the hydrophobic NC1 domain affect the trimerization process and produce unstable COL10A1 which is degraded in the endoplasmic reticulum[8] resulting in haploinsufficiency. Some NC1 missense variants, for example (p.Tyr598Asp) have been demonstrated to give rise to formation of abnormal disulfide bonds by exposing sulfhydryl groups of cysteine which are normally buried inside protein in the normal configuration.[8]

To date, all patients reported with pathogenic variants in *COL10A1* have been heterozygous. In the present study, we describe a large consanguineous family with a novel missense variant of *COL10A1* leading to a pronounced phenotype with severe short stature and lower-limb deformities in homozygous individuals while family members harbouring a heterozygous variant were only mildly affected. This is the first disease-causing variant in *COL10A1* located in the NC2 domain of the protein. Further, to the best of our knowledge, this is the first report of individuals with a homozygous variant in *COL10A1*.

MATERIALS AND METHODS

Patients

The study was approved by the Institutional Review Board at the School of Biological Sciences, University of the Punjab, Lahore, Pakistan. Family NAD-02 (figure 1A) was recruited to the study after obtaining written informed consents from the participants or their parents. A detailed history about the onset and progression of growth failure and deformities was obtained. Family history was recorded and photographs were taken. Ages and heights of all family members were

recorded and standard deviation scores for heights were calculated using WHO population-based reference values. Medical radiographs could not be obtained as all family members refused the testing. Peripheral blood samples were drawn from the participants. DNA was isolated using a standard protocol.[12]

Molecular Analysis

Whole-exome sequencing (WES) was performed for one severely affected individual IV:15, using Agilent V4 enrichment kit (Agilent Technologies, Santa Clara, CA). Paired-end reads were obtained at 50× coverage on an Illumina Hi-Seq 2000 sequencer (Otogenetics, Norcross, GA). Reads were mapped to UCSC hg19 reference human genome (http://genome.ucsc.edu/). The web based program wANNOVAR was used for further annotating variants. All heterozygous variants and variants in dbSNP database, Exome Aggregation Consortium (ExAC), 1000 Genomes and in 6500 exon sequence project with minor allele frequency (MAF) greater than 0.0099 were removed. Exonic and splice site variants were considered. Homozygous chromosomal intervals in the exome data were identified using Agilevariant mapper and intronic and synonymous variants located in these regions were also scrutinized. The disease causing potential of sorted variants was checked by using online prediction programs which included Polyphen2, Mutation taster, Mutation Assessor and fathmm. ClustalO was used for generating protein alignment of COL10A1 orthologous sequences retrieved from UCSC genome browser. Primers for the amplification of genomic regions containing the selected variants were designed using Primer3. The segregation of each variant with the phenotype was checked after PCR amplification of the specific product followed by Sanger sequencing with Big Dye Terminator V3.1 cycle Sequencing kit (Applied Biosystems).

RESULTS

Clinical Features

Family NAD-02 had eight severely affected individuals born in four consanguineous marriages (figure 1A). Six affected individuals participated in the study. All severely affected individuals had severe disproportionate short stature with short limbs and significant lower limb varus or valgus deformity but normal facial features (Table 1). Their growth was reported to be normal until the age of 6 years. However, no longitudinal height data were available. The adult heights ranged from 114 cm to 130 cm (Table 1). Lower limb deformities progressed with age in all, with older individuals complaining of joint pain in knees. None of the patients had significant scoliosis. Bowing of legs was visible but hands and feet were normal. One individual (IV:2) homozygous for the variant had genu valgum while all other severely affected individuals had genu varus. Apart from the skeletal findings, no other clinical signs, including intellectual disability, infections, respiratory problems or hearing deficits were present in any of the affected individuals. Family members who were obligate carriers and those who were subsequently found to be heterozygous for the disease-causing variant were apparently normal but had mild short statures (adult heights from 140 cm to 162 cm) (Table 1). They did not complain of joint pain or other heath related issues. None of these individuals considered themselves as affected.

Molecular Findings

Analysis of the whole exome sequencing data revealed a novel homozygous missense mutation in exon 2 of *COL10A1*, c.133 C>T (p.Pro45Ser) which fully co-segregated with the disease phenotype in the family, with the six severely affected individuals being homozygous for the variant (figures 1A, 1B, Supplementary Table 1). Analysis of data for individual 1V:15 indicated

that all coding exons located in the homozygous chromosome 6 interval (114-130 MB, according to GRCh37/hg19, data not shown), in which *COL10A1* is located, were fully covered by exome sequencing. No other potential pathogenic variant was identified in the exome data within this region of homozygosity on chromosome 6 and none of the other variants on different chromosomal intervals segregated with the phenotype (Supplementary Table 1).

The *COL10A1*, c.133 C>T (p.Pro45Ser) mutation was absent from public databases including gnomAD (http://gnomad.broadinstitute.org/) as well as in ethnically matched 100 control chromosomes. The variant was predicted to be pathogenic or damaging by all four prediction programs and also had a high Combined Annotation Dependent Depletion (CADD) score of 14.55 (Supplementary Table 1). The Pro45 amino acid is absolutely conserved in COL10A1 orthologues (figure 1C).

DISCUSSION

Variants of COL10A1 have been shown to cause metaphyseal chondrodysplasia, Schmid type in humans and chondrodysplasia in mice. Both haploinsufficiency and dominant-negative gain of function are proposed mechanisms for MCDS. Here we describe a new form of skeletal dysplasia caused by biallelic mutations in *COL10A1* resulting in a skeletal phenotype which is reminiscent of MCDS caused by heterozygous *COL10A1* variants but with more severe symptoms. Interestingly, the carriers of (p.Pro45Ser) variant had mild short statures but no other phenotypic manifestations. One explanation for this variant causing the severe manifestations of the disorder only in the homozygous form could be due to the unusual site of the mutation (figure 1D). The variant identified in the present family is located in the NC2 region of COL10A1, in which no pathogenic variants have been identified previously. All reported

dominantly inherited mutations in MCDS patients are located in the NC1 region of *COL10A1* except for three variants.[2 3] These *COL10A1* heterozygous variants affect the signal peptide (p.Gly18Arg), (p.Gly18Glu) or the triple helical domain (p.Gly288Arg) of COL10A1 (figure 1D, Supplementary Table 2). The three variants are associated with late onset and mild forms of MCDS.[2 3]

All except one (p.Val677Glu) of the pathogenic variants of NC1 region of COL10A1 (figure 1D and Supplementary Table 2) result in the development of a severe form of the disease. The NC1 domain has been shown to play a vital role in the trimerization of collagen X.[7] Moreover, it has been proposed that the NC1 region interacts with other proteins and any alteration in the NC1 region may also lead to accumulation of its interacting proteins inside cells, adding to pathology of disease.[13]

Homozygous *Col10a1*$^{-/-}$ null mutant mice exhibit late onset extremely mild abnormalities.[14] These mutant mice develop unilateral coxa vara on ageing. The observed difference in the severity of phenotype between mice and humans was proposed to be due to the difference in the weight borne by growth plates in both species.[14] However, transgenic wild-type mice transfected with mutant *COL10A1* develop a severe phenotype with genu valgum, coxa vara and short stature similar to that found in human MCDS patients heterozygous for these mutations.[15 16] The severity of the disease in these mice is gene dosage dependent with homozygous mutant mice being more severely affected than their heterozygous littermates.[15 16] We speculate that the (p.Pro45Ser) variant causes the disease in a dosage dependent manner similar to that seen for transgenic mice with *COL10A1* mutations.

The lack of radiographs for the participants in our study hinders an unequivocal phenotypic diagnosis for both individuals heterozygous and homozygous for the variant. Nevertheless, the variant (p.Pro45Ser) in the present family results in a later onset (at around 6 years of age) of an extremely severe disorder only in individuals carrying the homozygous variant. The usual age of onset in dominantly inherited MCDS is between 1 to 4 years.[17] It has been speculated that late onset of the disease associates with mutations leading to complete nonsense-mediated RNA decay of the mutated transcripts, while the more common early onset forms are caused by mutations with incomplete nonsense-mediated decay and dominant-negative impairment of protein folding, endoplasmic reticulum stress response and altered hypertrophic chondrocyte differentiation.[18]

The heterozygous carriers of the (p.Pro45Ser) variant had short heights, but did not exhibit apparent skeletal anomalies or other associated clinical features observed in those with dominantly inherited MCDS or in the homozygous affected individuals in their own family. Although we cannot exclude the possibility of a skeletal phenotype detectable by radiography in heterozygous individuals with (p.Pro45Ser) variant, the absence of visible skeletal defects in these individuals suggest variants in COL10A1 are attractive candidates for idiopathic short stature. The phenotypic variability among patients harbouring different *COL10A1* mutations also indicates complexity in the disease mechanism.

The pathogenic mechanism through which the missense variant (p.Pro45Ser) leads to severe short stature in our patients remains to be elucidated. The crystal structure for NC-2 domain of COL10A1 is not available, thus precluding prediction of the potential effects of the variant on

the protein. However, proline has unusual properties since it is a cyclic α-imino acid rather than an α-amino acid. Unlike other amino acids, the α-carbon atom is part of a cyclic pyrrolidine ring, which is thus introduced into the polypeptide backbone on incorporation of proline.[19] Moreover, there is no hydrogen bond donor in the amide bond. Incorporation of proline produces bends and its inclusion can serve as a helix breaker in the polypeptides. Additionally, though usually found in the *cis* state, it has been proposed that proline residues may act as molecular switches in mature proteins by transitioning between the *cis and trans* states.[20] The substitution of serine residue at position 45 may affect some of these unique properties conferred by proline. In addition, serine is an amino acid with a hydrogen bond donor. This may affect the bonding of COL10A1 or lead to its aberrant interactions with other proteins. Determination of the crystal structure of NC-2 domain is required to elucidate the precise mechanism of (p.Pro45Ser) variant's effect. The knowledge gained from *in vitro* and *in vivo* analyses of COL10A1 variants will lead to better understanding of its function in health and disease.

Web resources

1000 Genomes Project (http://www.1000genomes.org)

Agilevariant mapper (http://dna.leeds.ac.uk/agile/AgileGenotyper/)

ClustalO (http://www.ebi.ac.uk/Tools/msa/clustalo/)

dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/)

Exome Aggregation Consortium (http://exac.broadinstitute.org/)

Exome Variant Server (http://evs.gs.washington.edu/EVS/)

Fathmm (http://fathmm.biocompute.org.uk/inherited.html)

gnomAD (http://gnomad.broadinstitute.org/)

Human gene mutation database (http://www.hgmd.cf.ac.uk/)

Mutation taster (http://www.mutationtaster.org/)

Mutation Assessor (http://mutationassessor.org/r3/)

Online Mendelian Inheritance in Man, OMIM, (http://www.omim.org)

Polyphen2, (http:// http://genetics.bwh.harvard.edu/pph2/)

Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm)

UCSC Genome Browser Feb. 2009 (GRCh37/hg19), https://genome.ucsc.edu/cgi-bin/hgGateway

wANNOVAR (http://wannovar.usc.edu/)

REFERENCES

1. Bateman JF, Wilson R, Freddi S, Lamandé SR, Savarirayan R. Mutations of COL10A1 in Schmid metaphyseal chondrodysplasia. *Hum Mut* 2005;25:525-34.

2. Ikegawa S, Nakamura K, Nagano A, Haga N, Nakamura Y. Mutations in the N-terminal globular domain of the type X collagen gene (COL10A1) in patients with Schmid metaphyseal chondrodysplasia. *Hum Mut* 1997;9:131.

3. Park H, Hong S, Im Cho S, Cho T-J, Choi IH, Jin D-K, Sohn YB, Park SW, Cho H-H, Cheon J-E. Case of mild Schmid-type metaphyseal chondrodysplasia with novel sequence variation involving an unusual mutational site of the COL10A1 gene. *Eur J Med Genet* 2015;58:175-9.

4. Bateman JF, Freddi S, Nattrass G, Savarirayan R. Tissue-specific RNA surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in Schmid metaphyseal chondrodysplasia cartilage. *Hum Mol Genet* 2003;12:217-25.

5. Chan D, Weng YM, Graham HK, Sillence DO, Bateman JF. A nonsense mutation in the carboxyl-terminal domain of type X collagen causes haploinsufficiency in schmid metaphyseal chondrodysplasia. *JClin Invest* 1998;101:1490.

6. Ho MS, Tsang KY, Lo RL, Susic M, Makitie O, Chan TW, Ng VC, Sillence DO, Boot-Handford RP, Gibson G, Cheung KM, Cole WG, Cheah KS, Chan D. COL10A1 nonsense and frame-shift mutations have a gain-of-function effect on the growth plate in human and mouse metaphyseal chondrodysplasia type Schmid. *Hum Mol Genet* 2007;16:1201-15.

7. Brass A, Kadler KE, Thomas JT, Grant ME, Boot-Handford RP. The fibrillar collagens, collagen VIII, collagen X and the C1q complement proteins share a similar domain in their C-terminal non-collagenous regions. *FEBS Lett* 1992;303:126-8.

8. Wilson R, Freddi S, Chan D, Cheah KS, Bateman JF. Misfolding of collagen X chains harboring Schmid metaphyseal chondrodysplasia mutations results in aberrant disulfide bond formation, intracellular retention, and activation of the unfolded protein response. *J Biol Chem* 2005;280:15544-52.

9. Tsang KY, Chan D, Cheslett D, Chan WC, So CL, Melhado IG, Chan TW, Kwan KM, Hunziker EB, Yamada Y. Surviving endoplasmic reticulum stress is coupled to altered chondrocyte differentiation and function. *PLoS Biol* 2007;5:e44.

10. Bonaventure J, Chaminade F, Maroteaux P. Mutations in three subdomains of the carboxy-terminal region of collagen type X account for most of the Schmid metaphyseal dysplasias. *Hum Genet* 1995;96:58-64.

11. Chan D, Cole WG, Rogers JG, Bateman JF. Type X collagen multimer assembly in vitro is prevented by a Gly618 to Val mutation in the α1 (X) NC1 domain resulting in Schmid metaphyseal chondrodysplasia. *J Biol Chem* 1995;270:4558-62.

12. Grimberg J, Maguire S, Belluscio L. A simple method for the preparation of plasmid and chromosomal E. coli DNA. *Nucleic Acids Res* 1989;17:8893-.

13. Bogin O, Kvansakul M, Rom E, Singer J, Yayon A, Hohenester E. Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen X NC1 domain trimer. *Structure* 2002;10:165-73.

14. Kwan KM, Pang MK, Zhou S, Cowan SK, Kong RY, Pfordte T, Olsen BR, Sillence DO, Tam PP, Cheah KS. Abnormal compartmentalization of cartilage matrix components in mice lacking collagen X: implications for function. *J Cell Biol* 1997;136:459-71.

15. Jacenko O, LuValle P, Solum K, Olsen B. A dominant negative mutation in the alpha 1 (X) collagen gene produces spondylometaphyseal defects in mice. *Prog Clin Biol Res* 1992;383:427-36.

16. Jacenko O, LuValle PA, Olsen BR. Spondylometaphyseal dysplasia in mice carrying a dominant negative mutation in a matrix protein specific for cartilage-to-bone transition. *Nature* 1993;365:56-61.

17. Mäkitie O, Susic M, Ward L, Barclay C, Glorieux FH, Cole WG. Schmid type of metaphyseal chondrodysplasia and COL10A1 mutations—findings in 10 patients. *Am J Med Genet A* 2005;137:241-8.

18. Makitie O, Susic M, Cole WG. Early-onset metaphyseal chondrodysplasia type Schmid associated with a COL10A1 frame-shift mutation and impaired trimerization of wild-type alpha1(X) protein chains. *J Orthop Res* 2010;28:1497-501.

19. Berg JM, Tymoczko JL, Stryer L. *Biochemistry*. 5th ed. New York: W. H. Freeman and Company., 2002.

20. Lu KP, Finn G, Lee TH, Nicholson LK. Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol* 2007;3:619-29.

Legends

Figure 1: Family NAD-2 with the c.133 C>T (p.Pro45Ser) **variant, sequence conservation and schematic representation of COL10A1.**

(A) Pedigree of family NAD-2. Double lines indicate consanguineous marriages, filled symbols represent homozygous affected individuals and symbols with dots denote the heterozygous

carriers of the variant. Genotypes for *COL10A1* variant c.133 C>T are provided for all participants below their symbols.

(B) Partial chromatograms of DNA sequence of *COL10A1*. Arrows indicate the point of mutation.

(C) Clustal Omega sequence alignment of COL10A1 from diverse vertebrate species showing conservation of Proline in all orthologues. The conserved amino acids are highlighted in grey. The asterisk signs below the alignment indicate evolutionary conserved amino acids, a colon indicates highly conserved amino acids, and the periods symbolize less conserved amino acid changes.

(D) Schematic representation of COL10A1. The amino acid numbers corresponding to each domain are shown by integers. Vertical lines with integers show all variants of COL10A1 in the signal peptide, NC2 and the triple helical region as well as selected changes from 47 variants affecting NC1. The newly identified variant is boxed. S, signal peptide; NC2, non-collagenous domain 2; Helix, triple helical domain and NC1, non-collagenous domain 1. The amino acids for start and ends of domains are: 1-18 S, 19-56 NC2, 57-519 Helix and 520-680 NC1.
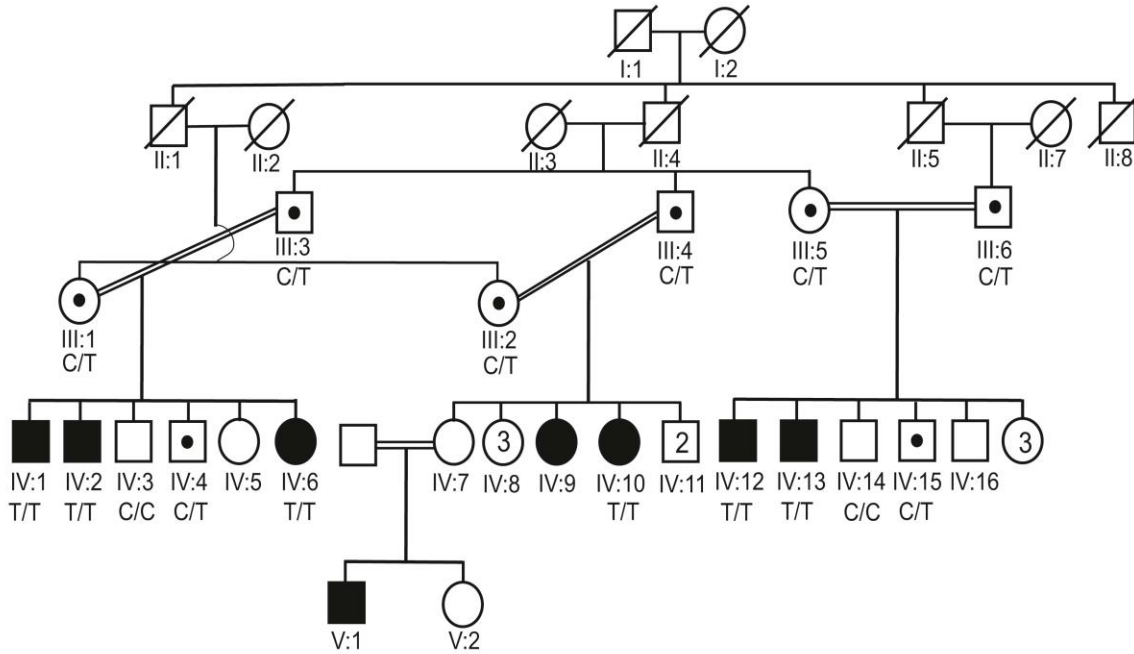
Table 1: Genotype and phenotype of individuals in family NAD-02

Name	Sex	Age (years)	Height (cm)	SD	Deformity	Zygosity	Gait
III:1	F	58	145	-2.8	No	Heterozygous	Normal
III:2	F	56	140	-3.5	No	Heterozygous	Normal
III:3	M	60	160	-2.3	No	Heterozygous	Normal
III:4	M	58	162	-2.0	No	Heterozygous	Normal
III:5	F	58	147	-2.5	No	Heterozygous	Normal
III:6	M	62	162	-2.0	No	Heterozygous	Normal
IV:1	M	23	130	-6.4	Genu varum	Homozygous	Waddling
IV:2	M	18	130	-6.1	Genu valgum	Homozygous	Waddling
IV:3	M	29	168	-1.2	No	Wild type	Normal
IV:4	M	18	160	-2.2	No	Heterozygous	Normal
IV:6	F	21	114	-7.5	Genu varum	Homozygous	Waddling
IV:11	F	16	119	-6.8	Genu varum	Homozygous	Waddling
IV:15	M	21	130	-6.4	Genu varum	Homozygous	Waddling
IV:16	M	14	122	-4.8	Genu varum	Homozygous	Waddling
IV:17	M	28	168	-1.2	No	Wild type	Normal
IV:4	M	18	150	-3.7	No	Heterozygous	Normal

SD: standard deviation, cm: centimeter

(A)

NAD-02

(B)

Wild type

C T T C A T T C C C T A C A C

Homozygous

C T T C A T T T C C T A C A C

Heterozygous

C T T C A T T Y C C T A C A C

(C)

p.Pro45Ser

```
Human      KGPLPNTKTQFFIPYTIKSKGIAVRGEQG 60
Mouse      KGPLASPKTQYFIPYAIKSKGIPVRGEQG
Chicken    KGP-----PHFLPFNVKSQGVQMRGEQG
Lizard     KGPLPNTKTQFFIPYTIKSKGIAVRGEQG
Zebrafish  -------KAPQYQPYSVKSHVVSVAGEPG
            :  *:  :**:  :  :  ** *
```

(D)