의학박사 학위논문

Genomic analysis of spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type to identify the gene containing the causal mutations

관절이완-협지형 척추골단골간단이형성증의 원인 돌연변이 유전자 규명을 위한 유전체 분석 연구

2013 년 8 월

서울대학교 대학원

의과학과 의과학 전공

민 병 주

A thesis of the Degree of Doctor of Philosophy

관절이완-협지형 척추골단골간단이형성증의 원인 돌연변이 유전자 규명을 위한 유전체 분석 연구

Genomic analysis of spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type to identify the gene containing the causal mutations

August 2013

Major in Biomedical Sciences
Department of Biomedical Sciences
Seoul National University Graduate School
Byung-Joo Min

Genomic analysis of spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type to identify the gene containing the causal mutations

by
Byung-Joo Min

A thesis submitted to the Department of Biomedical Sciences in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Biomedical Sciences at Seoul National University College of Medicine

August 2013

Approved by Thesis Committee:

Professor _____Chairman
Professor _____Vice chairman
Professor _____
Professor _____
Professor _____

ABSTRACT

Introduction: Spondyloepimetaphyseal dysplasia with joint laxity (SEMD-JL), leptodactylic type (MIM 603546) is a monogenic rare disease classified by well-defined diagnostic criteria. It is characterized by short stature, joint laxity with dislocation, limb malalignment, and spinal deformity. It is considered an autosomal dominant trait and allelic variants underlying the disease has not been discovered.

Methods: Whole exome sequencing and protein damage prediction were applied to discover the mutations underlying SEMD-JL, leptodactylic type. PCR & Sanger sequencing methods were used to validate the variants. The RT-PCR for *KIF22* and its mouse homolog *Kid* was performed to confirm expression in bone and other types of connective tissue in both human and mice. *In silico* analysis of protein structure was also performed to predict structural damage of KIF22 by the sequence variations identified.

Results: Five novel sequence variations in the *KIF22* gene were identified in seven patients. Sanger sequencing of *KIF22* confirmed that c.443C>T (p.Pro148Ser) co-segregated with the phenotype in the affected individuals in the familial case; and c.442C>T (p.Pro148Leu) or c.446G>A (p.Arg149Gln) was present in four of five simplex individuals and were absent in unaffected parents and 505 Korean normal population. The other two variants were present in unaffected parents. *In silico* analysis indicated that Pro148 and Arg149 were important to maintain hydrogen bonding in the ATP binding and motor domains of the KIF22 protein.

Conclusions: Specific mutations of *KIF22* gene affecting amino acids 148 and 149 in motor domain are pathogenic for SEMD-JL, leptodactylic type.

Keywords: SEMD-JL, leptodactylic type; *KIF22*; NGS; Whole exome sequencing

Student number: 2010-21904

CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1

Chapter 2

GENERAL INTRODUCTION

Studies of monogenic rare diseases are of substantial value because important knowledge of specific gene and its function can be obtained by identification of their genetic bases. And the information provides the clue for novel biological pathways, disease mechanisms, and identifies targets for disease treatment related with the gene.

But there are some limiting factors: small number of affected individuals, absence of pedigree information, and the lack of genetic understanding. They make it hard to identify specific genetic variants responsible for such diseases.

Whole exome sequencing was adopted to evade these limiting factors in search for the causative gene of a monogenic rare disease, spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type. Whole exome sequencing is a targeted resequencing technology and is an efficient strategy to detect pathogenic mutations in monogenic rare diseases by selectively analyzing the protein coding sequences of the human genome. It is estimated that approximately 85% of disease-causing

mutations are located within the protein coding regions of the human genome (Choi et al., 2009).

In this study, I focused on the identification of pathogenic variants in spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type by whole exome sequencing and genetic validation of them by using other experimental approaches.

CHAPTER 1

Identification of *KIF22* mutations as a cause of spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type by whole exome sequencing

INTRODUCTION

Spondyloepimetaphyseal dysplasia is group of disorders defined by combination of abnormalities of epiphyses, metaphyses, and vertebral bodies based on clinical and radiological features (Taybi et al., 1996). Spondyloepimetaphyseal dysplasia with joint laxity (SEMD-JL), leptodactylic type (MIM603546) is a rare but distinct entity of the spondyloepimetaphyseal dysplasias group (Warman et al., 2010). It was first reported by Hall as spondyloepimetaphyseal dysplasia with large joint dislocations in 1998, differentiating it from SEMD-JL, Beighton type (MIM 271640) which is one of more commonly described form and was reported by Beighton (Hall et al., 1998). Several series of patients have been reported since spondyloepimetaphyseal dysplasia with large joint dislocations was classified and nosology of the rare disease was confirmed as SEMD with joint laxity (SEMD-JL), leptodactylic type (Superti-Furga et al., 2007).

Clinical finding of the patients of SEMD-JL, leptodactylic type are short stature, distinctive midface hypoplasia, progressive knee malalignment (genu valgum and/or genu varum), generalized ligamentous laxity, and mild spine deformity, but

intellectual development is regarded to be not impaired and radiographic findings are mild scoliosis, narrow lumber IP distances, sacral spinal dysraphism, lumbar posterior scalloping, vertebral end plate irregularity, hip subluxation/dislocation, slender femoral necks, small irregular epiphyses, metaphyseal irregularity, sclerosis, striations, delayed bone maturation, delayed patellar ossification, gracile metacarpals, small carpus, small carpal bones, and squared distal ends of middle phalanges (Hall et al., 1998). Sacral spinal dysraphism, small carpus, small carpal bones, and squared distal ends of middle phalanges are properties of SEMD-JL, leptodactylic type separated from other forms of SEMD-JL. Leptodactylic type is considered an autosomal dominant trait based on studies of three families (Hall et al., 2002). In this study, genomic analysis was performed on three familial affected individuals and an unaffected father (Figure 1) and on five simplex affected individuals with SEMD-JL, leptodactylic type to identify the disease-causing mutations. All subjects were of Korean origin and clinical features of the eight affected individuals are summarized in Table 3.

MATERIALS AND METHODS

1. Array CGH analysis

chromosomal changes were analyzed by using the SurePrint G3 Human Catalog 1M CGH microarray (Agilent Technologies, Santa Clara, CA) covering the whole genome with a resolution of approximately 3.2 Kb overall average probe spacing according to the manufacturer's protocol. Human genomic DNA obtained from a karyotypically normal male (Promega, Madison, WI) was used as reference DNA. DNA digestion, labeling, and hybridization were performed as previously reported. Reference DNA was labeled with Cy3-dCTP, and the subject's DNA was labeled with Cy5-dCTP by random priming using genomic DNA enzymatic labeling kit (Agilent Technologies). After hybridization of the microarray, the slide was washed and scanned using a DNA microarray scanner (Agilent Technologies). Agilent feature extraction and DNA analytics software were used to analyze scanned data.

2. Whole exome sequencing analysis

Whole exome sequencing of eight patients and the unaffected father of the familial case was performed. Genomic DNA

extracted from peripheral blood cells was sheared into 200- to 300-bp DNA fragments using Covaris (Covaris Inc.,Woburn, MA). Exome capture was performed to collect the protein coding regions of human genome DNA using the Agilent SureSelect Human All Exon Kit covering 38 Mb (1.22% of human genomic regions) as described in the manufacturer's instructions (Agilent, Santa Clara, CA). The gene sequences for this array are available in the Consensus Coding Sequence Region (CCDS) database (http://www.ncbi.nlm.nih.gov/projects/CCDS/). The exon-enriched DNA libraries were then subjected to a second sequencing library construction in preparation for Illumina GA sequencing and were sequenced using the Illumina Genome Analyzer II platform, following the manufacturer's instructions (Illumina, San Diego, CA). The human reference genome, together with its gene annotation, was downloaded from the UCSC database (http://genome.ucsc.edu/), version hg19 (NCBI Build37, 1 Feb 2009). Alignment of the patient sequences was performed using Burrows Wheeler Aligner (Li et al., 2009) after indexed, sorted and removed the duplicated reads using Samtools (Li et al., 2009) SNPs were called using

IndelRealigner in Genome Analysis Toolkit (McKenna et al., 2010) with the default parameters. Indels affecting coding sequence or splicing sites were identified using UnifiedGenotypeCaller and IndelCaller (Zanders et al., 2010). dbSNP was used for Q score calibration, and known variants were checked using dbSNP version 132. Annotation was done using RefSeq in GenBank. Sequence alignment, variant calling and gene annotation were performed by Dr. Jong-Il Kim at Department of Biomedical Sciences, Seoul National University College of Medicine and Dr. Namshin Kim at Korea Research Institute of Bioscience and Biotechnology. All changes were filtered against exome data of dbSNP, 1000 Genomes Project (February 28, 2011 releases for SNPs, the February 16, 2011 releases for indels fttp://www.1000genome.org), and 18 genomes from ethnic Korean individuals (Ju et al., 2011)

3. Bioinformatics analysis of the mutated variants

Several web sequence analysis programs were used to analyze the sequences including three known SNPs (rs67578835, rs235648, and rs2450399) and five novel variations in *KIF22*. Sequence homologies were described by using UCSC Human

blat (http://genome.ucsc.edu/cgi-bin/hgBlat). RepeatMasker was used to analyze repeat sequences around variants (http://www.repeatmasker.org/). DNA block aligner was used to investigate unknown paralogue of the sequences around the variants in *HYDIN* and *KIF22* genes (http://www.ebi.ac.uk/Tools/Wise2/dbaform.html). Polyphen II (http://genetics.bwh.harvard.edu/pph2/), SIFT (http://sift.jcvi.org/), and MutationAssessor (http://mutationassessor.org/) were used to predict protein damage induced by the variants. Finch TV program was used to analyze chromatogram of the sequence.

RESULTS

Array CGH analysis

Array-comparative genomic hybridization (CGH) analysis on three familial affected individuals, unaffected father and one simplex affected individual uncovered alterations encompassing or adjacent to 71 gene loci. The CNVs were excluded from further analysis because positions of the variants belong to introns and intergenic regions.

Whole exome sequencing analysis

Whole-exome sequencing of eight affected individuals and the unaffected father of the familial case was performed. Mean coverage depth of the exome was ranged from 39.53 to 60.83, which is of sufficient depth to interrogate the exons for mutations. Total of 5.1 Gb sequence mapped uniquely per individual as paired-end 76 bp reads to the human reference genome. More than 97% of the targeted bases were covered to pass the thresholds for variant calling at this depth of coverage (Table 1). In the familial case, 97 novel candidate variants in 86 genes were inherited from the mother to the two affected

offspring (Figure 1 and Table 2). Intersection to discover any gene with recurrent mutations between the family and the five simplex individuals was performed and 7 individuals in 8 SEMD-JL patients share variants in 2 genes, *KIF22* and *HYDIN* in dominant model. Function of *HYDIN* gene is not defined yet and the *HYDIN* has human specific paralogue, *HYDIN2* located at chromosome 1 with extraordinary homology (Martin, J et al., 2004). To remove possibility of the error induced by similarity of sequence between target region and its paralogue region, reference and mutated sequence were analyzed by using UCSC blat and DNA block aligner. Three variants in *HYDIN* gene suggested by whole exome sequening data were confirmed as reference sequences in *HYDIN2* paralugue. In case of *KIF22*, there was no paralogue that were more than 90 % similar to one another. Any sequence block, which is identical with mutated sequence region, was not detected either. *KIF22* (Kinesin family member 22 [MIM603213]) was found to harbor sequence variants in seven of eight affected individuals, including three known SNPs (rs67578835, rs235648, and rs2450399) and five novel variants (Figure 2 and Table 3). These sequence variants in *KIF22* are the strongest candidate

mutations for autosomal-dominant SEMD-JL, leptodactylic type.

Protein damage prediction

Five sequence variants in *KIF22* gene, c.442C>T (p.Pro148Ser), c.443C>T (p.Pro148Leu), c.446G>A (p.Arg149Gln), c.695G>A (p.Arg232Gln), and c.1677+124A>T passed the multiple exclusion criteria. As a complementary approach, we performed protein damage prediction on the five novel variations filtrated from whole exome sequencing results. We used MutationAssessor, Polyphen II, and SIFT to classify the variations. Two of these analysis methods Polyphen II and SIFT implicated c.442C>T (p.Pro148Ser), c.443C>T (p.Pro148Leu), c.446G>A (p.Arg149Gln), and c.695G>A (p.Arg232Gln) in *KIF22* as harmful mutations and MutationAssessor predicted the medium range possibility of functional effect. Damage prediction for c.1677+124A>T was not calculated because the variant is located in non-coding region. (Table 3)

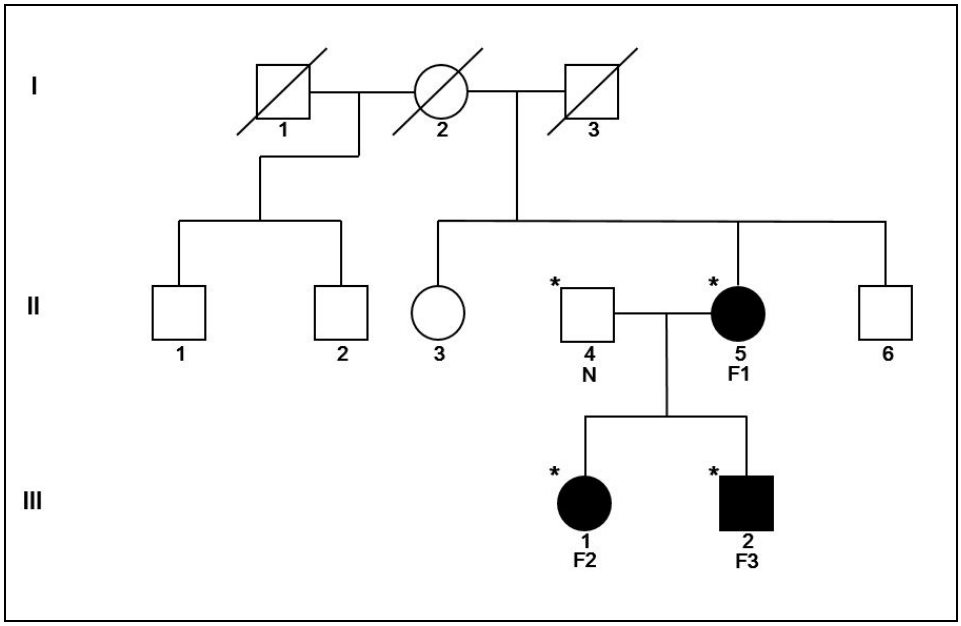


Figure 1. Pedigree of the familial case

Whole exome sequencing was performed in those family members marked with *. Affected individuals are indicated with symbols in black.

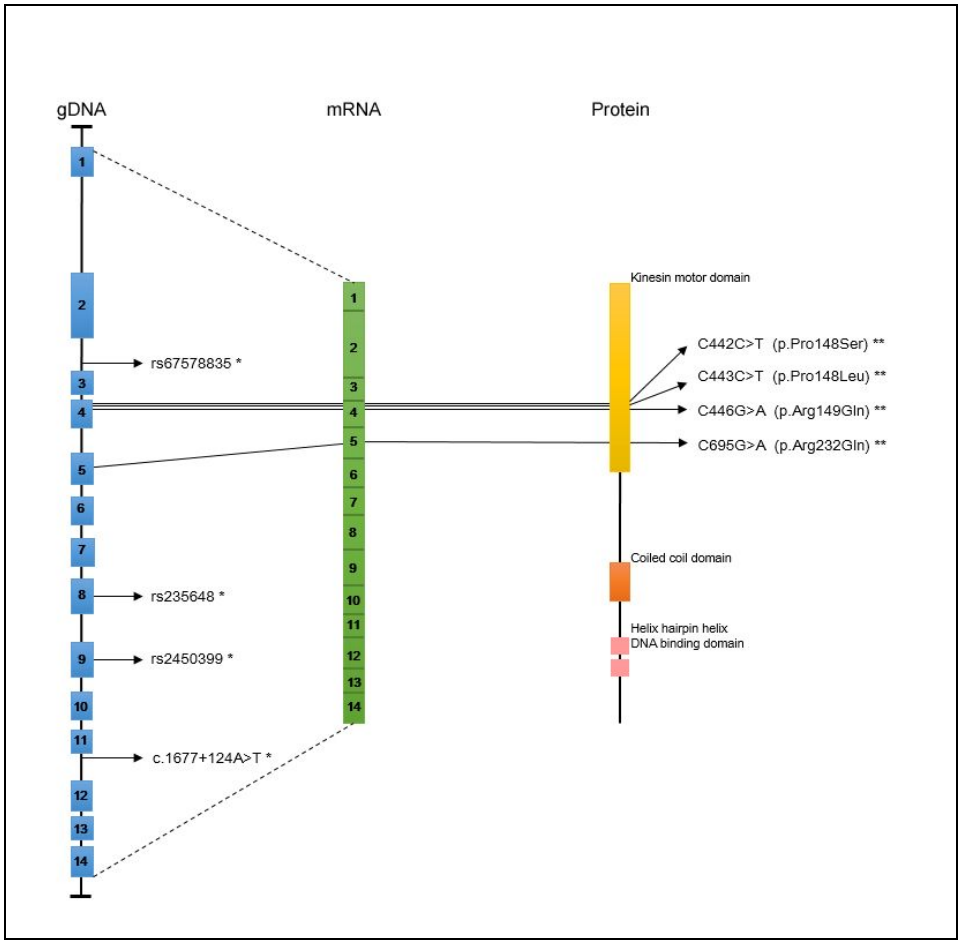


Figure 2. Genomic structure and spectrum of KIF22 variants detected in individuals with SEMD-JL, leptodactylic type

KIF22 is composed of 14 exons that encode Kinesin motor domain (yellow), coiled coil domain (orange), and helix hairpin helix DNA binding domain (pink). Arrows indicate the locations of 8 variants found in 8 affected individuals with SEMD-JL, leptodactylic type. SNP and intronic variant are indicated with *. Non-synonymous variants in CDS are indicated with **.

Table 1. Overview of whole exome sequencing performance

Case	Individual ID	Read length	Total reads	Reads mapped	% Reads mapped	Reads duplicate-removed	% Reads duplicate-removed
Normal control	N	78	59754164	58647296	98.1	56002895	93.7
Familial cases	F1	76	37494676	32957521	87.9	28624076	76.3
	F2	72	71379172	65307909	91.5	62890178	88.1
	F3	76	67181986	64748839	96.4	63007159	93.8
Simplex cases	S1	76	69004610	66005268	95.7	64391614	93.3
	S2	76	69265354	65969443	95.2	64285159	92.8
	S3	76	68166672	65125298	95.5	63592771	93.3
	S4	76	67531266	65018773	96.3	63209733	93.6
	S5	78	59765578	58713971	98.2	55986167	93.7

N: Husband of F1, father of F2 and F3

Table 2. Overview of all variants identified by whole exome sequencing in 9 individuals

		N	F1	F2	F3	S1	S2	S3	S4	S5
SNV	Total	41,664	29,223	37,304	37,334	36,451	36,246	36,199	35,832	42,591
	dbSNP	40,393	28,154	36,001	36,155	35,286	35,042	34,916	34,590	41,027
	HapMap	18,096	13,276	16,634	16,317	15,824	15,957	15,793	15,703	18,191
	1000 genomes	18,452	13,799	16,505	16,350	15,975	15,935	15,969	15,788	18,651
	Korean-specific	370	320	370	347	374	344	380	408	423
	Novel	894	728	904	822	778	842	865	804	1,105
	CDS novel	400	388	388	371	347	402	384	363	485
	CDS novel and deleterious	50	39	46	37	58	49	55	55	70
	Splice_site	41	32	32	33	43	36	25	37	34
Indel	Total	4,029	1,811	3,421	3,050	2,939	2,989	2,918	3,007	4,045
	dbSNP	3,549	1,608	2,988	2,646	2,569	2,598	2,558	2,607	3,548
	HapMap	5	2	5	2	1	2	6	5	4
	1000 genomes	17	3	14	11	16	11	15	14	21
	Korean-specific	199	101	188	160	142	144	173	151	199
	Novel	275	101	241	239	224	243	185	243	290
	CDS novel	25	10	16	23	28	31	25	26	24
	CDS novel and deleterious	4	3	4	3	6	3	3	4	2
	Splice_site	53	24	32	42	35	31	28	33	52
Complex	Complex novel and deleterious	12	8	6	7	5	7	7	5	11

Table 3. Major clinical findings in 8 individuals with SEMD-JL, leptodactylic type and summary of variants identified in *KIF22*

		Familial case			Simplex cases				
Individual ID		F1	F2	F3	S1	S2	S3	S4	S5
Gender		F	F	M	M	M	F	F	F
Clinical findings	Midface retrusion	++	++	++	++	+	+	++	++
	Radial head (sub) luxation	+	+	+	+	-	-	+	+
	Knee	Genu valgum, bilateral	Genu valgum, bilateral	Genu valgum, bilateral	Genu valgum, bilateral	Genu valgum with patellar dislocation, bilateral	Genu valgum, right > left	Windswept deformity	Genu varum, mild
	Hip	dislocation, right	None	None	None	None	None	Subluxatio, right	None,
	Spine deformity	Scoliosis	None	None	Kyphoscolios , mild	None	Scoliosis, mild	Kyphoscolios , mild	Mild kyphoscoliosi, resolved
	Laryngotra-cheomalacia	+	+	+	+	Denied	Denied	+	Denied
	Other	Strabismus	Strabismus	Strabismus	Strabismus	Strabismus	Strabismus	Strabismus	Epilepsy
Non-synonymous variants in CDS	Genomic position	29809961	29809961	29809961	29809965	•	29809965 29810441	29809962	29809962
	cDNA position	c.442C>T	c.442C>T	c.442C>T	c.446G>A	•	c.446G>A c.695G>A	c.443C>T	c.443C>T
	Protein consequences	p.Pro148Ser	p.Pro148Ser	p.Pro148Ser	p.Arg149Gln	•	p.Arg149Gln p.Arg232Gln	p.Pro148Leu	p.Pro148Leu
	PolyPhen	0.998	0.998	0.998	0.999	•	0.999 1.000	0.998	0.998
	SIFT	0.01	0.01	0.01	0.07	•	0.07 0.00	0.03	0.03
	Mutation Assessor	2.71	2.71	2.71	2.4	•	2.40 3.30	2.68	2.68
Intronic variants	Genomic position	•	•	•	•	•	•	29815510	•
	cDNA position	•	•	•	•	•	•	c.1677+124 A>T	•
SNP		rs235648 rs2450399	rs235648 rs2450399	rs235648 rs2450399	rs235648 rs2450399 rs67578835	rs235648 rs2450399	rs235648 rs2450399	rs235648 rs2450399	rs235648 rs2450399

Polyphen II and SIFT appraise mutations qualitatively as benign, possibly damaging, or probably damaging. Higher number indicates damaging and lower number indicates benign in Polyphen II. Lower number indicates damaging and higher number indicates benign in SIFT. MutationAssessor appraises mutations qualitatively as neutral, low, medium, or high. Higher number (3 to 6) indicates high and lower number (-6 to 1) indicates neutral in MutationAssessor.

DISCUSSION

Whole exome sequencing was applied to discover the mutations underlying SEMD-JL, leptodactylic type. Strong candidate of disease-causing mutation for SEMD-JL, leptodactylic type in eight Korean patients were detected in this study. Seven out of eight patients carried mutations in *KIF22*, which encodes a member of the kinesin-like protein family. The frequency of mutations in the *KIF22* gene in this cohort was 7/8 (87.5%). The mutations identified in the seven patients were localized in three amino acids in the motor domain of *KIF22*.

Any specific Mendelian disease caused by *KIF22* has not been reported and role of *KIF22* in the development of the skeletal system has not been explained yet. However, variations of other kinesin family members are reported as pathogenic mutations responsible for certain phenotype. Loss of mouse *Kif24* from cycling cells resulted in aberrant cilia assembly but did not promote growth of abnormally long centrioles (Kobayashi, T et al., 2011). *Kif3a* deficiency in mouse causes abnormal topography of hedgehog signaling, growth plate dysfunction, and nonphysiologic responses and processes in perichondrial tissues, including ectopic cartilage formation and

excessive intramembranous ossification (Koyama, E et al., 2007). Mutations in *KIF7* were reported in human Joubert syndrome and knockdown of *KIF7* caused problems in cilia formation, centrosomal duplication, and Golgi network (Dafinger, C et al., 2011). It is likely that the specific mutations of *KIF22* identified in individuals with SEMD-JL, leptodactylic type could interfere with hedgehog signaling as in *Kif3a* knock-out mice resulting in osteochondrodysplasia phenotype.

We did not detect any mutation of *KIF22* in a simplex case, S2 by either whole exome sequencing or aCGH. This individual showed a relatively mild clinical phenotype, including moderate short stature and equivocal midface retrusion, suggesting the possibility of a mutation in another gene related to *KIF22*.

As a result, *KIF22* ranked highly among the candidate genes estimated by the strategy for reducing the pool of candidate genes and predicting the protein damage. The mutations were detected in motor domain of *KIF22* and the mechanism lead to disease is likely related to motor dysfunction of *KIF22*, but additional studies will be needed to validate the candidate variants and exhibit causal mutations of SEMD-JL, leptodactylic type.

CHAPTER 2

Validation of disease causal mutations and identification of familial specific polymorphism of *KIF22* in spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type

INTRODUCTION

KIF22 gene is a member of kinesin-like protein family. KIF22 protein encoded by *KIF22* is a microtubule-dependent molecular motor transporting organelles within a cell and moving chromosomes during cell division. *Kid*, which is the mouse form of *KIF22*, mediates anaphase/telophase chromosome compaction, preventing formation of multinucleated blastomeres. Since the formation of multinucleated cells may disturb the proper zygotic gene expression, leading to embryonic death, the role of *Kid* is especially important during the very early stages of development. The motor domain and DNA-binding domain are critical for *Kid* localization with microtubules in between adjacent chromosomes during anaphase (Ohsugi et al., 2008).

We analyzed the exomes of three patients, the unaffected father of familial case, and five unrelated individuals with spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type by whole exome sequencing and 8 variations including three known SNPs (rs67578835, rs235648, and rs2450399) and five novel variations in the *KIF22* gene were detected in 7 of 8 patients.

To revise the false positive signal and reduce the candidate variants of the disease, multiple validation methods including Sanger sequencing and other tools are necessary. In this study, we screened 8 mutated variants in *KIF22* discovered by whole exome sequencing of three familial and five sporadic patients of SEMD-JL, leptodactylic type, and compared them to the sequences at identical region of the unaffected family members and 505 normal Korean population to identify disease causal mutation and familial specific polymorphism. The RT-PCR for human *KIF22* and *Kid*, which was a mouse form of *KIF22* and was performed to confirm expression in bone and other types of connective tissues. *In silico* analysis of protein structure was also performed to predict structural damage of KIF22 by those sequence variations.

MATERIALS AND METHODS

1. Validation of mutations using gDNA of the patients and their family members

All 14 exons of the *KIF22* gene were amplified for Sanger sequencing from genomic DNA of the eight patients, nine unaffected family members to validate mutations of *KIF22* in the patients. 12 PCR primers sets were used to amplify the 14 exons (Table 1). RBC HiYield Gel/PCR DNA Mini Kit was used to purify DNA in the PCR products. (Taipai county 220, Taiwan) Sanger sequencing was performed by Macrogen Inc. (Seoul, Korea).

2. Cohort study using gDNA samples of 505 Korean with normal phenotype by Sanger sequencing

gDNA samples of 505 Korean with normal phenotype were extracted from their whole blood samples. For cohort samples we pooled seven individuals in a single PCR reaction to detect the sequence variations using the same primer sets. RBC HiYield Gel/PCR DNA Mini Kit was used to purify DNA in the PCR products. (Taipai county 220, Taiwan) Sanger sequencing was performed by Macrogen Inc. (Seoul, Korea).

3. RNA preparation and reverse transcription polymerase chain reaction

Total RNA was extracted from various tissues from C57BL mice and human donors using the TRIzol method, and then reverse transcribed to complementary DNA using Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and oligo-(dT) 12-18 primers according to the manufacturer's protocol. The RT-PCR was performed in a reaction mixture containing Ex Taq (Takara Bio Inc., Shiga, Japan) and specific primers (Table 1). Amplicons were analyzed by agarose gel electrophoresis with size markers and visualized by staining with EtBr.

4. *In silico* analysis of protein structure

To analyze the effect of the three amino acid substitutions p.Pro148Ser, p.Pro148Leu, and p.Arg149Gln on the function of KIF22 protein, three mutant structures were generated. Appropriate three-dimensional protein structures were searched by using Protein Data Bank (PDB). Although a crystal structure of KIF22 (3BFN) is available, there are many missing structural parts, including one α-helix (from sequence number

231 to 241, NRTVGATRLNQ) that is known to be critical to ATP binding. In the absence of a complete human kinesin structure, two reference structures, 3BFN and 3KIN, were used to model the KIF22 protein structures. To model the missing structural parts, 3BFN was used as a base template structure, and 3KIN, a complete homodimer structure for the kinesin motor and neck domain from rat brain with bound ADP, to predict the missing α-helix part using MODELLER (Figure S3A). 3KIN is 36% identical to 3BFN. The amino acid substitutions were applied to the modeled structure and the side chains of the mutated residues were optimized using FoldX by calculating the lowest energy rotamer conformations. Structural analysis and visualization were performed using PyMOL. All steps for *In silico* analysis of protein structure were performed by Dr. Dongsup Kim and Mr. Taesu Chung at Korea Advanced Institute of Science and Technology.

RESULTS

Mutation validation and family study

To validate the abnormal variations including three known SNPs (rs67578835, rs235648, and rs2450399) and five novel variations detected in 7 of 8 patients by using whole exome sequencing data of previous study, several web sequence analysis programs, PCR, and Sanger sequencing method were performed. Remained gDNAs of three familial affected individuals and five simplex affected individuals, those were extracted for whole exome sequencing, were used for PCR and Sanger sequencing method. Sanger sequencing confirmed the presence of 8 sequence variations in the patients. To identifying disease causal mutation and familial specific polymorphism, we sequenced 14 exons of *KIF22* of family members with normal phenotype of the patients (Figure 1). c.442C>T (p.Pro148Ser), c.443C>T (p.Pro148Leu), and c.446G>A (p.Arg149Gln) were not existed in any of family members (Fig 2A). And the substitutions are located within a conserved domain of the protein (Fig 2B). rs235648 and rs2450399 were detected from every gDNA sample for family

study and 2 family members have rs67578835. Two novel sequence variations, c.695G>A (p.Arg232Gln), and c.1677+124A>T, were detected from specific family member of the patients. The c.695G>A (p.Arg232Gln) in affected individual S3 was inherited from her mother and the c.1677+124A>T in simplex affected individual S4 was also inherited from her mother with normal phenotype (Figure 2 and Table 2).

Korean population study

To analyze the frequency of the 5 novel variants, 505 gDNA samples of Korean with normal phenotype from whole blood were prepared. c.1677+124A>T in affected individual S4, which was inherited from her mother, was found in 28 chromosomes of 1010 chromosome. About 2.8% of population included in the study has the variant. The c.695G>A (p.Arg232Gln) in affected individual S3 was not found in any of control chromosomes from 505 Korean population and the remaining three sequence variations at residues Pro148 and Arg149 were not found in 1,010 control chromosomes from 505 Korean gDNAs either(Table 2).

Expression of *KIF22* in mouse and human tissues

Kid, which is mouse form of *KIF22*, expressed in various tissues including thymus, spleen, testis, and fetal liver (Tokai et al., 1996). Since SEMD-JL, leptodactylic type involves primarily the connective tissues, the expression of mouse *Kid* and human *KIF22* in various tissues, including bone and cartilage was examined by RT-PCR with gene-specific primers (Table 1). In C57BL mice, *Kid* mRNA was expressed in bone, cartilage, liver, ovary, small intestine, and spleen (Figure 3A). *KIF22* mRNA was detected in bone, cartilage, joint capsule, ligament, skin, and primary cultured chondrocytes harvested from human donors (Figure 3B).

3D structure of KIF22 and effect of mutations

A crystal structure of KIF22 (Protein Data Bank number 3BFN) was provided by the resource of Protein Data Bank. But there are many missing structural parts, including one α-helix (from sequence number 231 to 241, NRTVGATRLNQ) that is known to be critical to ATP binding (Kozielski et al., 1997). An additional reference structure, (3KIN, a homodimer structure

for kinesin motor and neck domain with bound ADP) selected to predict the structures of missing α-helix part in KIF22 with MODELER (Sali et al., 1993). The amino acid substitutions were applied to the modeled structure and the side chains of the mutated residues were optimized with FoldX (Schymkowitz et al., 2005) and the structural changes due to the mutations were investigated by the molecular dynamics (MD) simulations with GROMACS (Van Der Spoel et al., 2005). Structural analysis and visualization were performed with PyMOL. A helix-loop-helix structure adjacent to the ATP binding site spans from129 to 156 amino acid residues containing
Pro148 and Arg149. The loop structure adjacent to the ATP binding site is stabilized by hydrogen bonds among its residues, for example, Leu138 with Gly145, Pro144 with Ser140, and Gln143 and Glu142 with Ser140. Loop residues also interact with the adjacent α-helix and other residues near the loop region; for example, Gln143 with His135 and Gly145 with Arg149. In the MD simulation on the p.Pro148Leu mutation, two novel hydrogen bonds, which were absent in wild-type (Figures 4A and 4B) were found. The hydrogen bond between NH of Arg149 and C = O of Gly145 is the expected regular

hydrogen bond between ith and (i+4)th residues in the helix structure. The other hydrogen bond was formed between NH of Leu148 and C = O of Leu138. Identical results for p.Pro148Ser were obtained. With regard to the effect of the mutation, p.Arg149Gln, Arg149 forms a hydrogen bond with Gln101 and Tyr104 residues in an α-helix located near ATP binding site and also with the loop residue Pro144 (Figure 4C). The p.Arg149Gln mutation would disrupt the hydrogen bonds with the α-helix of Gln101 and Tyr104 (Figure 4D).

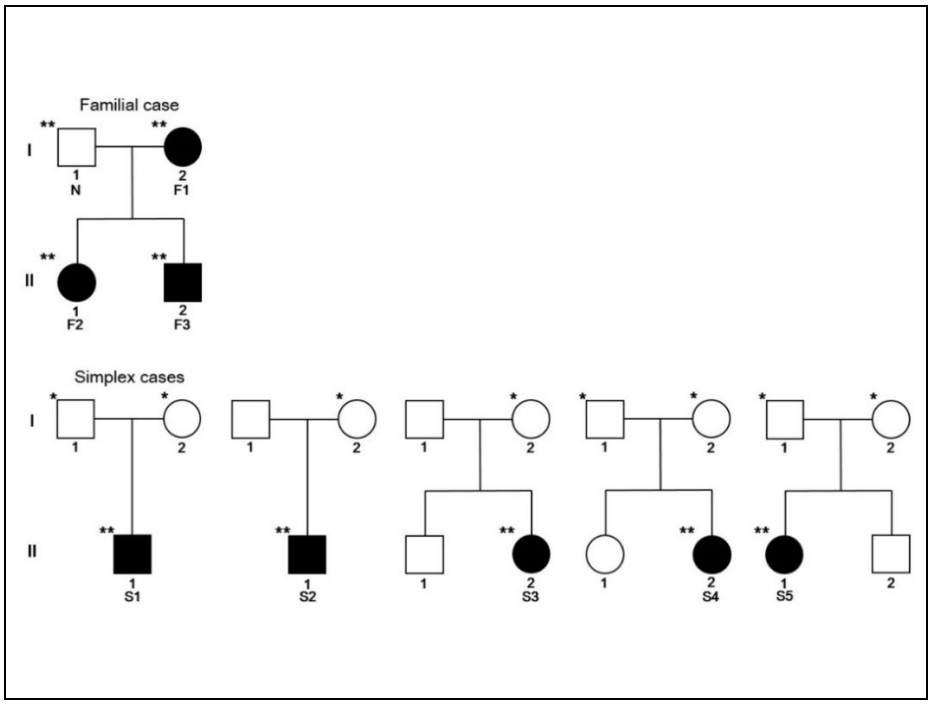
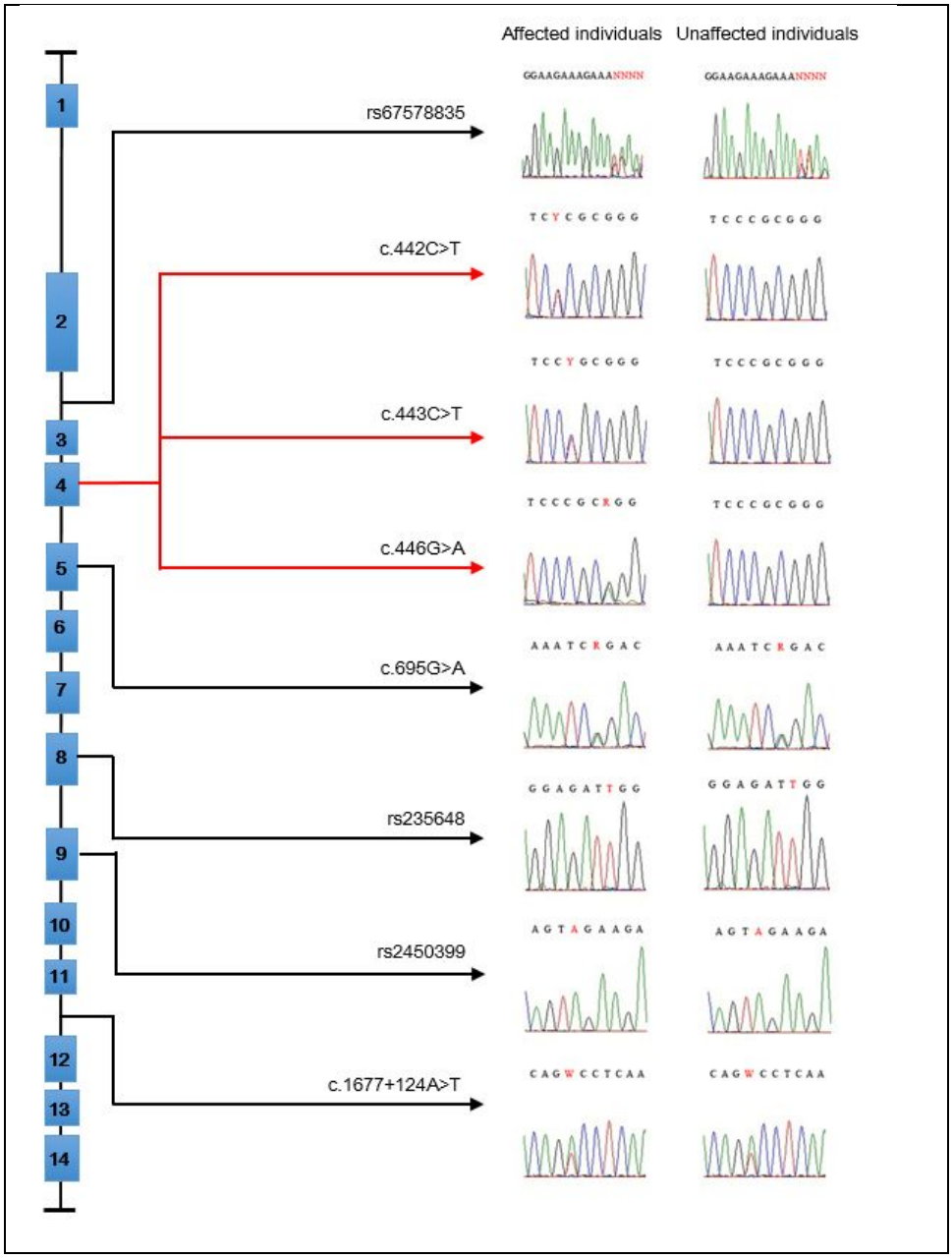


Figure 1. Pedigrees of the familial case and 5 simplex cases

Pedigrees showing 8 affected individuals and their family members. Affected individuals are indicated with symbols in black. ** indicates the individuals who provided gDNA to perform whole exome sequencing and PCR validation and * indicates the individuals who provided gDNA to perform PCR validation

(A)


Affected individuals
Unaffected individuals
1
2
3
4
5
6
7
8
9
10
11
12
13
14
rs67578835
c.442C>T
c.443C>T
c.446G>A
c.695G>A
rs235648
rs2450399
c.1677+124A>T

(B)

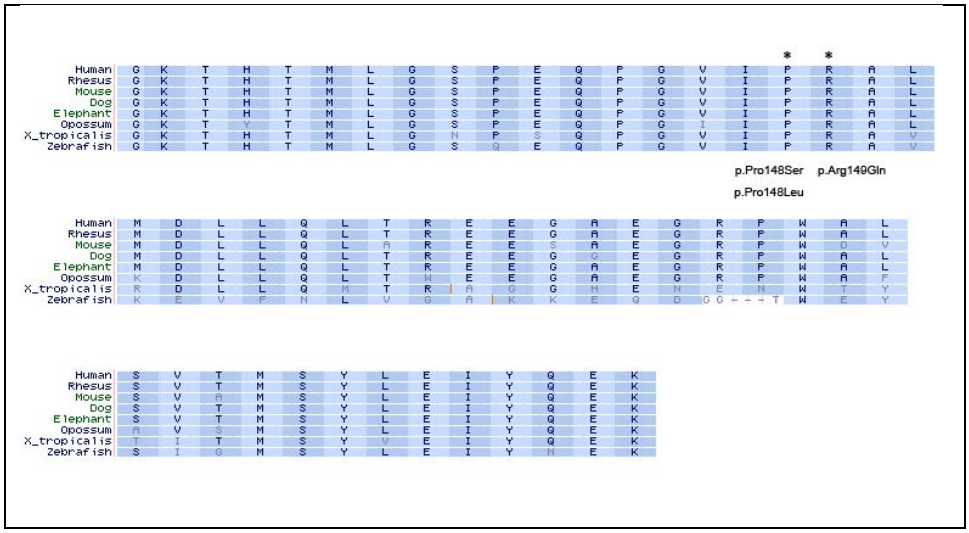
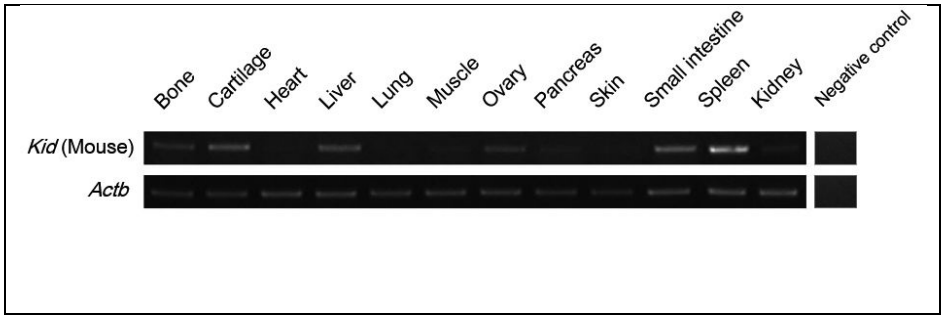


Figure 2. Sanger sequencing results of 8 variants analyzed using whole exome sequencing data and multi-species sequence alignment for exon4 of *KIF22*

Partial chromatograms surrounding the *KIF22* variant sites validating the de novo occurrence for the candidate variants causing spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type. Red arrow indicates 3 novel variants in exon 4 not detected in any unaffected individual (A). The protein sequences of human, Rhesus, mouse, dog, elephant, opossum, xenopus tropicalis and zebrafish orthologues were compared. The position of the de novo variants is labeled with an asterisk

(B).

(A)



(B)

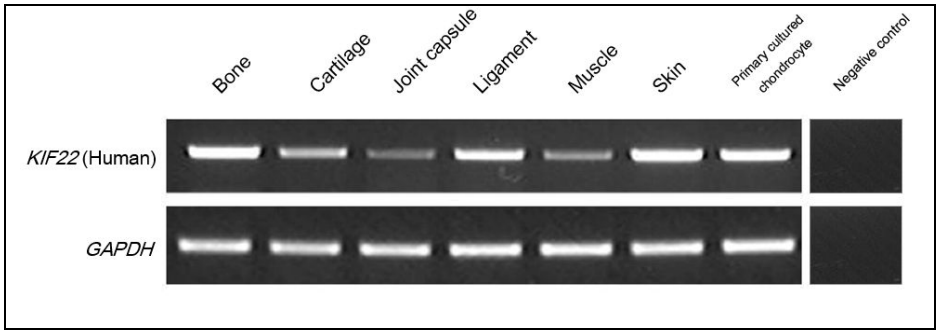
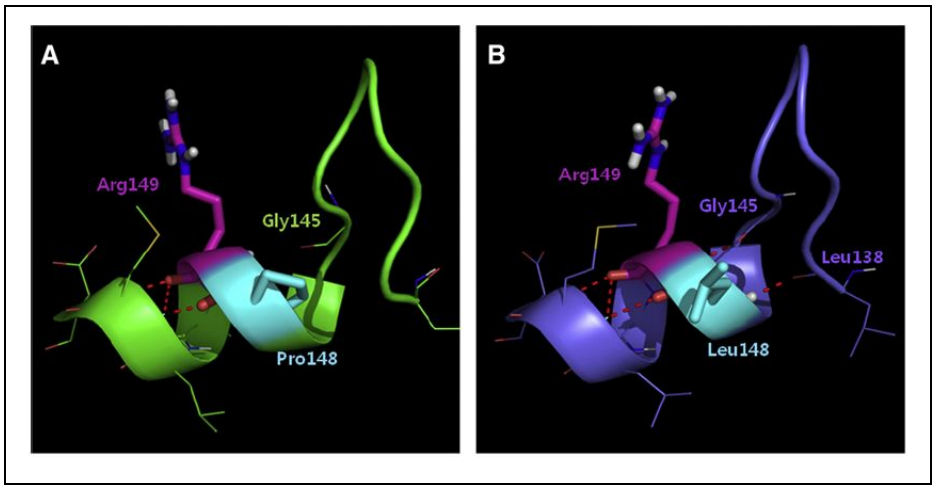


Figure 3. Expression of *Kid/KIF22* mRNA in mouse and human tissues

The expression of *Kid/KIF22* in (A) mouse and (B) human tissues was analyzed by RT-PCR and gel electrophoresis.

(A and B)



(C and D)

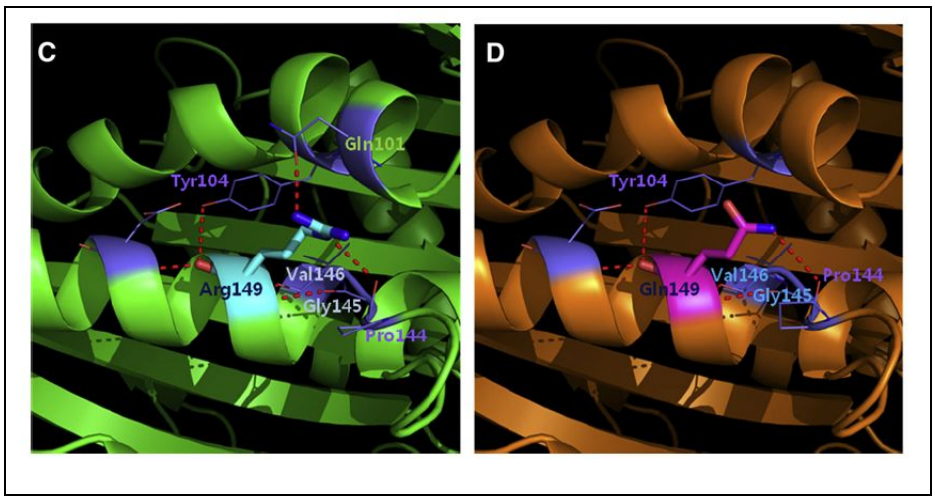


Figure 4. 3D Structure of KIF22 and the effect of mutations

(A) The wild-type structure is described in 4(A). 4(B) shows the mutant structure with Pro148Leu mutation introduces two

new hydrogen bonds; the hydrogen bond between NH of Arg149 and C = O of Gly145, and the hydrogen bond between NH of Leu148 and C = O of Leu138. Tyr104 residues in an α-helix located near ATP binding site and hydrogen bond between Arg149 and Gln101 Pro144 are described in 4(C) and 4(D) shows that predicted impact of the p.Arg149Gln mutation disrupting the hydrogen bonds with the a-helix of Gln101 and Tyr104. The red dotted line indicates a hydrogen bond.

Table 1. Primer set for the variant validation and tissue specific expression confirmation

	Gene	Position	Forward primer	Reverse primer
Variant validation	*KIF22* (Human)	Exon 1	ATTGGTGGAACCTGGGAGAG	TGGACCACACTCCGACACAT
		Exon 2	CATGACCAGGGCAGAATGAG	GAAGGCTCAGGGAACAAAGC
		Exon 3	TAGAATCCCTTACCCACCCC	CTGCAGGAGGTCCATGAGAG
		Exon 4	ACTTGCTGGAAGGGCAGAAT	GGATTACCAGGTCTCCCGAA
		Exon 5	GGCCCTTTCTGTCACCATGT	GTCCTCTGACCCAGCCAAGT
		Exon 6	GTCATGCTGTGCTCCTGGTC	AGAAGCCCAGCCTTCCTATG
		Exon 7	CTGTCTCAGGGAAGAAGGGG	AGCCTCTGTGCTGATGCTTG
		Exon 8	GCTGGGTCTGGAAATTAGGG	TCTGCTCAAGGATCTGACGG
		Exon 9	ACCCAGAGAAACTGGATGCC	GCCTGAGTGTCCTTCCCTGT
		Exon 10	CCTGGTAACCCACTCCCTTC	CCTCCCTGGGACTGGTATGT
		Exon 11	AGCTGGGGAATCAGAAATGG	ACCACATTCCAGGGGTAAGC
		Exon 12, 13, and 14	CGGGATTTTGGACACACTTG	AAGATGGCGCCCAGTGGAGT
Tissue specific expression confirmation	*KIF22* (Human)	Exon 11, 12, 13, and 14	TGAGGGGGCCCTCCGACG	TGAGGGGGCCCTCCGACG
	Kid (Mouse)	Exon 10, 11, and 12	TGCCCCTGCAACGGATTCAGAA	TGCGCCCATGTGCCAGTAACTC
	Actb	Exon 4	TACCACAGGCATTGTGATGG	TCTTTGATGTCACGCACGATT
	GAPDH	Exon 8	GCCTTCCGTGTCCCCACT	TGAGGGGGCCCTCCGACG

Actb and *GADDH*: Controls

Table 2. Eight variants detected in affected individuals, unaffected family members, and 505 Korean population.

	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6	Variant 7	Variant 8
Position	Intron 2	Exon 4	Exon 4	Exon 4	Exon 5	Exon 8	Exon 9	Intron 11
	29809674-29809677	29809961	29809962	29809965	29810441	29811319	29814234	29815510
Nucleotide mutation	c.267-21delT	c.442C>T	c.443C>T	c.446G>A	c.695G>A	c.1230C>T	c.1425G>A	c.1677+124 A>T
Amino acid change	.	p.Pro148Ser	p.Pro148Leu	p.Arg149Gln	p.Arg232Gln	p.Ile410Ile	p.Val475Val	.
Reference SNP	rs67578835	Novel	Novel	Novel	Novel	rs235648	rs2450399	Novel
Affected individuals	S1	F1 F2 F3	S4 S5	S1 S3	S3	All	All	S4
% Affected individuals (n=8)	12.50%	37.50%	25.00%	25.00%	12.50%	100%	100%	12.50%
Unaffected family members	Mother of S1 Mother of S4	None	None	None	Mother of S3	All	All	Mother of S4
% Unaffected family members (n=9)	22.20%	0.00%	0.00%	0.00%	11.10%	100.00%	100.00%	11.10%
% Korean population (n=505)	NA	0.00%	0.00%	0.00%	0.00%	NA.	NA	2.80%

NA: Not analyzed

DISCUSSION

Pathogenic mutation detection for a rare monogenic disease has significant meaning because it provides information to figure out primary role of target gene, specific pathway of target gene, and disease mechanism.

Whole exome sequencing was performed to detect the causative mutation of SEMD-JL, leptodactylic type and seven out of eight patients carried novel nucleotide variants in *KIF22*, which encodes a member of the kinesin-like protein family by data combining. Although *KIF22* was ranked highly among the candidate genes, still it is hard to classify that which variant is disease causal mutation and which one is not.

Candidates for provisional exclusion including 23,389 positions with excess heterozygosity suggestive of alignment errors were identified and 1,009 positions in which human genome reference sequence appeared to contain a minor allele in exome sequencing (Fajardo, K et al., 2011). To remove false positive signals and classify disease causal mutation, multiple validation methods were performed for five novel variants and three known SNPs (rs67578835, rs235648, and rs2450399). Four novel nonsynonymous variants were implicated as harmful

mutations by at least 2 protein damage prediction tools in previous study and 3 among 4 variants localized in two amino acids 148 and 149 were validated in only affected individual group and were absent in unaffected parents and 505 Korean population. The substitutions were located within well conserved regions in multiple sequence alignments. Another novel nonsynonymous sequence variation in affected individual S3 changing Arg232 to Gln232 was inherited from the unaffected mother, suggesting that p.Arg232Gln is a silent mutation. It is not found in any 505 Korean gDNA samples and possibly familial specific rare variant inherited by their ancestor. The novel intronic sequence variation (c.1677+124A>T) was inherited from the unaffected mother and is unlikely to be involved any functional effect and 2.8% of population included in the study has the variant. It was not filtrated by dbSNP and 1000 genome database and is possibly Korean specific SNP. For each novel silent mutation, we classified them as familial specific and regional specific single nucleotide variants. Two of the known SNPs (rs235648 and rs2450399) are synonymous amino acid substitutions and the allele frequency is very high in

the Korean population. Another SNP, rs67578835 was also frequently found in unaffected family members, too.

Kinesin binds to ATP and hydrolyzes it to provide energy for molecular motors, and many regions in kinesin are involved in ATP binding (Kozielski et al., 1997). Two new hydrogen bonds originated by p.Pro148Leu and p.Pro148Ser mutation slightly pull the loop structure away from ATP binding pocket, damaging ATP binding and probably resulting in the loss of motor function and p.Arg149Gln mutation disrupt the connection between two α-helices to move away from the ATP binding pocket.

In view of the *KIF22* expression observed in the bone, cartilage, joint capsule, ligament, conceivable that the phenotype of affected individuals is the result of proteomic shifts caused by the mutation of *KIF22.*

Given the mutations in *KIF22* identified in this cohort, the *KIF22* expression pattern in connective tissues, and the predicted protein structural changes caused by these mutations, specific mutations of *KIF22* gene affecting amino acids 148 and 149 in motor domain are pathogenic for SEMD-JL, leptodactylic type. The mechanism by which these mutations

lead to disease is related to deficient motor function of KIF22, but additional studies will be needed to understand the *KIF22* function in the skeletal system and to understand the molecular pathogenesis of SEMD-JL, leptodactylic type associated with *KIF22* mutation.

REFERENCES

1. Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkaloglu, A., Ozen, S., Sanjad, S., et al. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences of the United States of America 106, 19096-19101.

2. Dafinger, C., Liebau, M.C., Elsayed, S.M., Hellenbroich, Y., Boltshauser, E., Korenke, G.C., Fabretti, F., Janecke, A.R., Ebermann, I., Nurnberg, G., et al. (2011). Mutations in KIF7 link Joubert syndrome with Sonic Hedgehog signaling and microtubule dynamics. The Journal of clinical investigation 121

3. Fajardo, K., Adams, D., NISC Comparative Sequencing Program., Mason, C., Sincan, M., Tifft, C., Toro, C., Boerkoel, C., Gahl, W and Markello, T (2011). Detecting False-Positive Signals in Exome Sequencing. HUMAN MUTATION 33 609-613

4. Hall, C.M., Elcioglu, N.H., and Shaw, D.G. (1998). A distinct form of spondyloepimetaphyseal dysplasia with multiple dislocations. J Med Genet 35, 566-572.

5. Hall, C.M., Elcioglu, N.H., MacDermot, K.D., Offiah, A.C., and Winter, R.M. (2002). Spondyloepimetaphyseal dysplasia with multiple dislocations (Hall type): three further cases and evidence of autosomal dominant inheritance. J Med Genet 39, 666-670.

6. Ju, Y.S., Kim, J.I., Kim, S., Hong, D., Park, H., Shin, J.Y., Lee, S., Lee, W.C., Kim, S., Yu, S.B., et al. (2011). Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals. Nat. Genet. 43, 745-752.

7. Kobayashi, T., Tsang, W.Y., Li, J., Lane, W., and Dynlacht, B.D. (2011). Centriolar Kinesin Kif24 Interacts with CP110 to Remodel Microtubules and Regulate Ciliogenesis. Cell 145, 914-925.

8. Koyama, E., Young, B., Nagayama, M., Shibukawa, Y., Enomoto-Iwamoto, M., Iwamoto, M., Maeda, Y., Lanske, B., Song, B., Serra, R., et al. (2007). Conditional Kif3a ablation causes abnormal hedgehog signaling topography, growth plate dysfunction, and excessive bone and cartilage formation during mouse skeletogenesis. Development 134, 2159-2169.

9. Kozielski, F., Sack, S., Marx, A., Thormahlen, M., Schonbrunn, E., Biou, V., Thompson, A., Mandelkow, E.M., and Mandelkow, E. (1997). The crystal structure of dimeric kinesin and implications for microtubule-dependent motility. Cell 91, 985-994.

10. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

11. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000

Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

12. Martin, J., Han, C., Gordon, L., Terry, A., Prabhakar, S., She, X., Xie, G., Hellsten, U., Chan, Y., et al (2004). The sequence and analysis of duplication-rich human chromosome 16. NATURE 432 23-30

13. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297-1303.

14. Ohsugi, M., Adachi, K., Horai, R., Kakuta, S., Sudo, K., Kotaki, H., Tokai-Nishizumi, N., Sagara, H., Iwakura, Y., and Yamamoto, T. (2008). Kid-mediated chromosome compaction ensures proper nuclear envelope formation. Cell 132, 771-782.

15. Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234, 779-815.

16. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: an online force field. Nucleic Acids Res 33, W382-388.

17. Superti-Furga, A., and Unger, S. (2007). Nosology and classification of genetic skeletal disorders: 2006 revision. Am J Med Genet A 143, 1-18.

18. Taybi H, Lachman RS, eds. Radiology of syndronmes, mnetabolic disorders, and skeletal dysplasias. 4th ed. St Louis: Mosby, 1996:916-23.

19. Tokai, N., Fujimoto-Nishiyama, A., Toyoshima, Y., Yonemura, S., Tsukita, S., Inoue, J., and Yamamota, T. (1996). Kid, a novel kinesin-like DNA binding protein, is

localized to chromosomes and the mitotic spindle. The EMBO journal 15, 457-467.

20. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J. (2005). GROMACS: Fast, flexible, and free. J. Comput. Chem. 26, 1701-1718.

21. Warman, M.L., Cormier-Daire, V., Hall, C., Krakow, D., Lachman, R., LeMerrer, M., Mortier, G., Mundlos, S., Nishimura, G., Rimoin, D.L., et al. (2011). Nosology and classification of genetic skeletal disorders: 2010 revision. Am J Med Genet A 155A, 943-968.

22. Zanders, S., Ma, X., Roychoudhury, A., Hernandez, R.D., Demogines, A., Barker, B., Gu, Z., Bustamante, C.D., and Alani, E. (2010). Detection of heterozygous mutations in the genome of mismatch repair defective diploid yeast using a Bayesian approach. Genetics 186, 493-503.

국문 초록

서론: 관절이완을 동반하는 관절이완-협지형 척추골단골간단이형성증은 정립된 진단 기준에 의해 정의되어 있는 단일 유전자형 희귀 질환으로 단신, 관절 이완 및 탈구, 사지 부정렬, 척추 기형 등의 표현형을 가지고 있으며 상염색체 우성 유전의 경향을 따르고 있다. 현재까지 질환의 원인이 되는 유전적 돌연변이는 밝혀지지 않았다.

방법: 관절이완을 동반하는 관절이완-협지형 척추골단골간단이형성증을 유발하는 유전적 돌연변이를 찾기 위해 전체 엑솜 염기 서열 분석과 단백질 손상 예측 방법이 사용되었고 발견된 변이 확인을 위해 중합효소 연쇄 반응 및 Sanger 염기 서열 분석법이 사용되었다. 다양한 결합조직 내에서의 *KIF22* 발현 확인을 위해 역전사 중합효소 연쇄반응을 진행하였으며 구조적 변이의 확인을 위해 단백질 구조 예측 방법을 실행하였다.

결과: 7 명의 환자로부터 5 개의 새로운 염기 서열 변이를 발견하였다. c.443C>T (p.Pro148Ser) 염기 서열 변이가 가족 내 환자 3 명으로부터 공통적으로 발견되었으며 c.442C>T (p.Pro148Leu) 와 c.446G>A(p.Arg149Gln) 염기 서열 변이는 가족력이 없는 5 명의 독립된 환자 중 4 명에서 발견되었다. 세 종류의 염기 서열 변이 모두 환자 집단에서만 나타나고 가족과 505 명의 정상 한국인 집단에서는 나타나지 않는 것을 확인하였으며 단백질 구조 예측을 통해 Motor domain 내에서 ATP 결합을 저해하는 수소 결합의 생성 및 구조적 변이를 예측하였다.

결론: *KIF22* 의 Moter domain 내에 포함되는 148, 149 번째 아미노산에 특이적으로 작용하여 구조적 변이를 유발하는 유전자 돌연변이들이 관절이완을 동반하는 관절이완-협지형 척추골단골간단이형성증의 원인이 된다는 것을 밝혀내었다.

주요어: 관절이완-협지형 척추골단골간단이형성증, *KIF22*, 차세대 염기 서열 분석, 전체 엑솜 염기 서열 분석, 희귀 질환, 가족 연구

학 번: 2010-21904