GENOMICS **45,** 304–312 (1997)
ARTICLE NO. GE974921

Structure of the Human Type XIX Collagen (*COL19A1*) Gene, Which Suggests It Has Arisen from an Ancestor Gene of the FACIT Family

Mohammed Khaleduzzaman, Hideaki Sumiyoshi, Yasuyoshi Ueki, Kazuhito Inoguchi, Yoshifumi Ninomiya, and Hidekatsu Yoshioka[1]

Department of Molecular Biology and Biochemistry, Okayama University Medical School, Okayama 700, Japan

Received April 30, 1997; accepted July 23, 1997

Type XIX collagen is a newly discovered member of the FACIT (fibril-associated collagens with interrupted triple helices) group of extracellular matrix proteins. Based on the primary structure, type XIX collagen is thought to act as a cross-bridge between fibrils and other extracellular matrix molecules. Here we describe the complete exon/intron organization of *COL19A1* and show that it contains 51 exons, spanning more than 250 kb of genomic DNA. The comparison of exon structures of *COL19A1* and other FACIT family genes revealed several similarities among these genes. The structure of exons encoding the noncollagenous (NC) 1–collagenous (COL) 1–NC 2–COL 2–NC 3–COL 3–NC 4 domain of the α1(XIX) chain is similar to that of the NC 1–COL 1–NC 2–COL 3–NC 3 domain of the α2(IX) chain except for the NC 3 domain of α1(XIX). The exons encoding the COL 5–NC 6 domain of α1(XIX) are also similar to those of the COL 3–NC 4 domain of α1(IX) chain. Previously, *COL19A1* was mapped to human chromosome 6q12–q14, where *COL9A1* is also located. Likewise, the present work shows that the mouse *Col19a1* gene is located on mouse chromosome 1, region A3, where *Col9a1* has also been mapped. Taken together, the data suggest that *COL19A1* and *COL9A1* (*Col19a1* and *Col9a1*) were duplicated from the same ancestor gene of the FACIT family. Three CA repeat markers with high heterozygosity were found in *COL19A1*. These markers may be useful for linkage analysis of age-related inheritable diseases involved in eyes and/or brain.

To date, 19 distinct types of collagen molecule, encoded by more than 30 genes, have been identified in vertebrates (van der Rest and Garrone, 1991; Kivirikko, 1993; Inoguchi *et al.*, 1995). The collagen superfamily can be classified into subgroups, based on their structures and functions. The FACIT (fibril-associated collagens with interrupted triple helices) family of collagens is one of those groups and includes types IX, XII, XIV, XVI, and XIX collagen. FACIT collagens are associated with the surface of collagen fibrils where they maintain the organization and stability of the extracellular matrix. Members of the FACIT family share some common structural features, the functional significance of which is based mainly on the study of type IX collagen (Shaw and Olsen, 1991). The structure of these molecules comprises several collagenous (COL) domains divided and flanked at both ends by noncollagenous (NC) domains. Type IX collagen is the best characterized molecule in the members. The COL domains and the central NC domains of this molecule interact with type II collagen through covalent cross-links to form fibrils (Diab *et al.*, 1996). The amino-terminal NC domain has a potential of interacting with other extracellular components.

We (Yoshioka *et al.*, 1992; Inoguchi *et al.*, 1995) and others (Myers *et al.*, 1993, 1994) have recently reported the identification of a new member of the FACIT family, named type XIX collagen, through the isolation of overlapping cDNA clones from a human rhabdomyosarcoma cell line (RD, CCL136). The predicted human polypeptide contains 1142 amino acid residues that include a 23-residue signal peptide and five COL domains divided and flanked by six NC domains. The size of the coding region is unexpectedly small compared to the overall length of the mRNA (10.4 kb) due to the presence of a very long 3′ untranslated region (>5 kb). Interestingly, the corresponding *COL19A1* gene was localized to 6q12–q14, the same region where the α1(IX) collagen gene (*COL9A1*) (Yoshioka *et al.*, 1992) and the α1(XII) collagen gene (*COL12A1*) reside (Oh *et al.*, 1992). Type IX collagen is found in tissues containing type II collagen such as hyaline cartilage and the vitreous body of the eye. On the other hand, type XII collagen is found in dense connective tissues such as tendons and ligaments associated with type I collagen fibrils.

Sequence data from this article have been deposited with the DDBJ/EMBL/GenBank Data Libraries under Accession Nos. AB004584 to AB004634 and AB004688. The symbols *COL19A1* and *Col19a1* represent the human and the mouse α1(XIX) collagen gene, respectively.

[1] To whom correspondence should be addressed at the Department of Molecular Biology and Biochemistry, Okayama University Medical School 2-5-1, Shikata-cho, Okayama 700, Japan. Telephone: +81-86-235-7128. Fax: +81-86-222-7768. E-mail: mbb@ccews2.cc.okayama-u.ac.jp.

0888-7543/97 $25.00

It is well established that mutations in different kinds of collagen molecules are the cause of numerous connective tissue disorders. They include osteogenesis imperfecta, Ehlers–Danlos syndrome (types VII, IV, and I), some forms of chondrodysplasias, Alport syndrome, and epidermolysis bullosa (dystrophic form) (Kivirikko, 1993). Until recently, there was no known mutation in a FACIT collagen. Muragaki *et al.* (1996) reported a COL9A2 mutation in a family affected with multiple epiphyseal dysplasia, symptoms of which are stiffness and pain in large joints combined with short stature and stubby fingers. Moreover, studies in mice also showed that type IX collagen plays a critical role during the process of skeletal formation. Transgenic mice with a truncated α1(IX) collagen chain were found to express mild chondrodysplasia and progressive osteoarthritis (Nakata *et al.*, 1993), and mice homozygous for a null mutation of *Col9a1* developed progressive osteoarthritis-like changes in articular cartilage (Faessler *et al.*, 1994).

As a necessary step toward elucidating α1(XIX) collagen function and pathology, here we describe the complete exon/intron organization of *COL19A1*. The study demonstrates that this gene contains 51 exons spanning more than 250 kb of DNA. We have identified three highly polymorphic markers within *COL19A1*. Additionally, we have mapped the mouse *Col19a1* locus to chromosome 1. These data may facilitate the identification of hereditary abnormalities of *COL19A1* (or *Col19a1*).

MATERIALS AND METHODS

Isolation and characterization of human genomic clones. A human genomic library (HL1111j; Clontech) was screened with cDNAs encoding various portions of the human α1(XIX) collagen chain (Inoguchi *et al.*, 1995). The screening was performed according to standard procedures (Sambrook *et al.*, 1989). Positive clones were characterized by genomic mapping and by Southern blot hybridization with different probes. Isolation of clones coding exons 2, 12, 15, 17, 43, and 44 was performed using exon-specific probes made from HY 67 or KI 40 cDNA clone (Inoguchi *et al.*, 1995) as templates by PCR (polymerase chain reaction) procedure. The nucleotide sequences of primers for generating exon-specific probes are exon 2, (forward) 5′-ATCCGTGGCCGTTCACATGG-3′ and (reverse) 5′-GTCCCTAACGGTCACGGAAG-3′; exon 12, (forward) 5′-GGTGAACAAGGAGAAAAAGG-3′ and (reverse) 5′-ATTTTCTCCATTAAGGCCAG-3′; exon 15, (forward) 5′-GGTTCCCTGGGGATACAAGG-3′ and (reverse) 5′-CCTCTGACCCTCTTTTCCAG-3′; exon 17, (forward) 5′-GGACCTCCTGGAATACAAGG-3′ and (reverse) 5′-CTTGTTATCCTTGTTATAAT-3′; and exon 43/44, (forward) 5′-GGAAAACCTGGTGCCCCAGG-3′ and (reverse) 5′-TGGCTTTCCTGAGGGTCCTT-3′.

PCR was performed using 10 ng of cDNA template in a 50-μl volume containing 2 units of *Tth* DNA polymerase (Toyobo, Osaka) under the following conditions: 94°C for 1 min, 55°C for 1 min, and 70°C for 1 min for 30 cycles. The PCR products were subcloned into TA vectors (Invitrogen). The screening was performed under low-stringency conditions using exon-specific probe. Hybridization was carried out at 50°C overnight in a mixture containing 5× SSC (1× SSC is 0.15 *M* NaCl, 0.015 *M* sodium citrate, pH 6.8), 1% *N*-laurylsarcosine, 50 μg/ml salmon sperm DNA, and ^{32}P-labeled probes. The washing was performed at 50°C with 6× SSC for 30 min.

Suitable restriction fragments of the genomic clones were subcloned into pBluescript II vector for further characterization. DNA sequencing was carried out with the Sanger dideoxynucleotide chain termination method (Sambrook *et al.*, 1989). The ends of subcloned fragments were sequenced with universal forward and reverse primers for pBluescript II vector. In addition, the deletion fragments generated with mung bean nuclease deletion kit (Takara, Inc., Ootsu, Japan) or specific primers for insert fragments were useful for the sequence for the internal portions of the fragments. The sequence samples were prepared using the AmpliTaq cycle sequencing kit (Perkin–Elmer, Inc.) following the manufacturer's protocol. The samples were applied to a 373S automatic sequencer (Perkin–Elmer, Inc.). All reported sequences were confirmed by sequencing of both strands.

Allele detection using microsatellites. CA repeat sequences were found in the three genomic DNA fragments mentioned above. To identify family by polymorphism, primers for PCR were designed at the flanking regions of the CA repeats (Table 1). PCR was carried out according to previously described procedures (Oohashi *et al.*, 1995). Genomic DNAs were isolated from 25 male and 25 female unrelated Japanese individuals using standard procedures (Sambrook *et al.*, 1989).

Chromosomal mapping of Col19a1. A mouse genomic clone, gME 1, containing exons 5 and 6 of *Col19a1* (H. Sumiyoshi *et al.*, unpublished data) was used for chromosomal mapping of *Col19a1*. This clone was isolated from the mouse genomic library (a gift from Dr. F. Ramirez, Mount Sinai School of Medicine, New York, NY) using a cDNA clone, ME 1, encoding a portion of the NC 6 domain of the mouse α1(XIX) collagen chain as a probe (Sumiyoshi *et al.*, 1997). The DNA was purified using the Qiagen plasmid kit (Qiagen, Inc., Chatsworth, CA).

Mouse chromosomes were prepared according to the published procedure (Feng *et al.*, 1994). Briefly, lymphocytes were isolated from mouse spleen and cultured at 37°C in RPMI 1640 medium supplemented with 15% fetal calf serum, 3 μg/ml concanavalin A, 10 μg/ml lipopolysaccharide, and 5×10^{-5} *M* mercaptoethanol. After 44 h, the cultured lymphocytes were treated with 0.18 mg/ml bromodeoxyuridine for an additional 14 h. The synchronized cells were washed and recultured at 37°C for 4 h in α-minimal essential medium with thymidine (2.5 μg/ml). Chromosome slides were made by the conventional method used for human chromosome preparation (hypotonic treatment, fixation, and air drying). The genomic probe was biotinylated with dATP using the BioNick labeling kit (Gibco BRL) at 15°C for 1 h (Heng *et al.*, 1992).

The procedure for fluorescence *in situ* hybridization (FISH) detection was performed according to the previously published protocol (Heng *et al.*, 1992; Heng and Tsui, 1993). Briefly, slides were baked at 55°C for 1 h. After RNase A treatment, the slides were denatured in 70% formamide in 2× SSC for 2 min at 70°C followed by dehydration with ethanol. Probes were denatured at 75°C for 5 min in a hybridization mix consisting of 50% formamide and 10% dextran sulfate and mouse cot I DNA and prehybridized for 15 min at 37°C. Probes were loaded on the denatured slides. After overnight hybridization, slides were washed and signal was detected as well as amplified using published method (Heng *et al.*, 1992). FISH signals and 4,6-diamino-2-phenylindole (DAPI) banding pattern were recorded separately by taking photographs, and the assignment of the FISH mapping data with chromosomal bands was achieved by superimposing FISH signals with DAPI-banded chromosomes (Heng and Tsui, 1993).

RESULTS AND DISCUSSION

Isolation of COL19A1 Clones

We previously isolated cDNA clones covering the entire human α1(XIX) collagen chain, as well as four genomic clones, termed KIG 3, KIG 12, KIG 23, and KIG 13 (see Fig. 1) (Inoguchi *et al.*, 1995). These genomic clones contained the 5′ flanking region of *COL19A1* and the first exon, as well as the 3′ region containing

TABLE 1

DNA Sequence of CA Repeats, Flanking Primers, and Loci

Clone	Locus	CA strand primer	Repeat sequence	GT strand primer
MKG 1	Intron 2	CAT GTA AAT TTG AGA ATC AG	$(CA)_{13}$	TCT AGA TTC AGA GCA ATG CC
MKG 5	Intron 5	CCT GTG TTA CTA ATA CAG AG	$(CA)_{12}$	ATG GGG CAT TCT GGT GAA TG
MKG 64	Intron 12	CTA CTA TTA CTA CTA GGG TC	$(CA)_6CC(CA)_3CC(CA)_{15}$	CTA GCT CCC TAA TGC CCT AG

Note. Primer sequences are written in the 5′ to 3′ direction.

five exons, which codes for a portion of the COL 1–NC 1–COL 2 region of α1(XIX). To complete the exon/intron organization of *COL19A1*, we isolated additional genomic clones using cDNA clones HY 67, KI 40, and KI 50 and their subfragments as probes (Inoguchi *et al.*, 1995). Restriction enzyme mapping and DNA sequencing analysis demonstrated that 11 positive clones contained the coding sequences for exons 3–11, 13, 14, 16, 18–42, and 50–51. Despite extensive screening of the genomic library, we were unable to isolated clones containing exons 2, 12, 15, 17, 43, and 44. To obtain genomic clones containing these exons, we made specific probes of 116, 54, 54, 63, and 108 bp in length for exons 2, 12, 15, 17, and 43/44, respectively. Using rather lower stringency hybridization and washing conditions, we finally isolated 5 clones (MKG 81, MKG 64, MKG 93, MKG 73, and MKG 66) that contain these exons. The alignment of the new and old genomic clones is shown in Fig. 1 in relation to the exon/intron organization of *COL19A1*. The exons that were previously designated exon +1 to +5 (Inoguchi *et al.*, 1995) correspond to exons 45–49, respectively. Likewise, the genomic fragments 3.5 and 1.48 kb in length previously analyzed by Myers *et al.* (1994) proved to contain the 3′ end of exon 36, exons 37–39, and 5′ end of exon 40.

Structure of COL19A1

Although there are several gaps between the individual genomic fragments we isolated, the data predict that the *COL19A1* gene is at least 250 kb in size and contains 51 exons. This represents the first time the

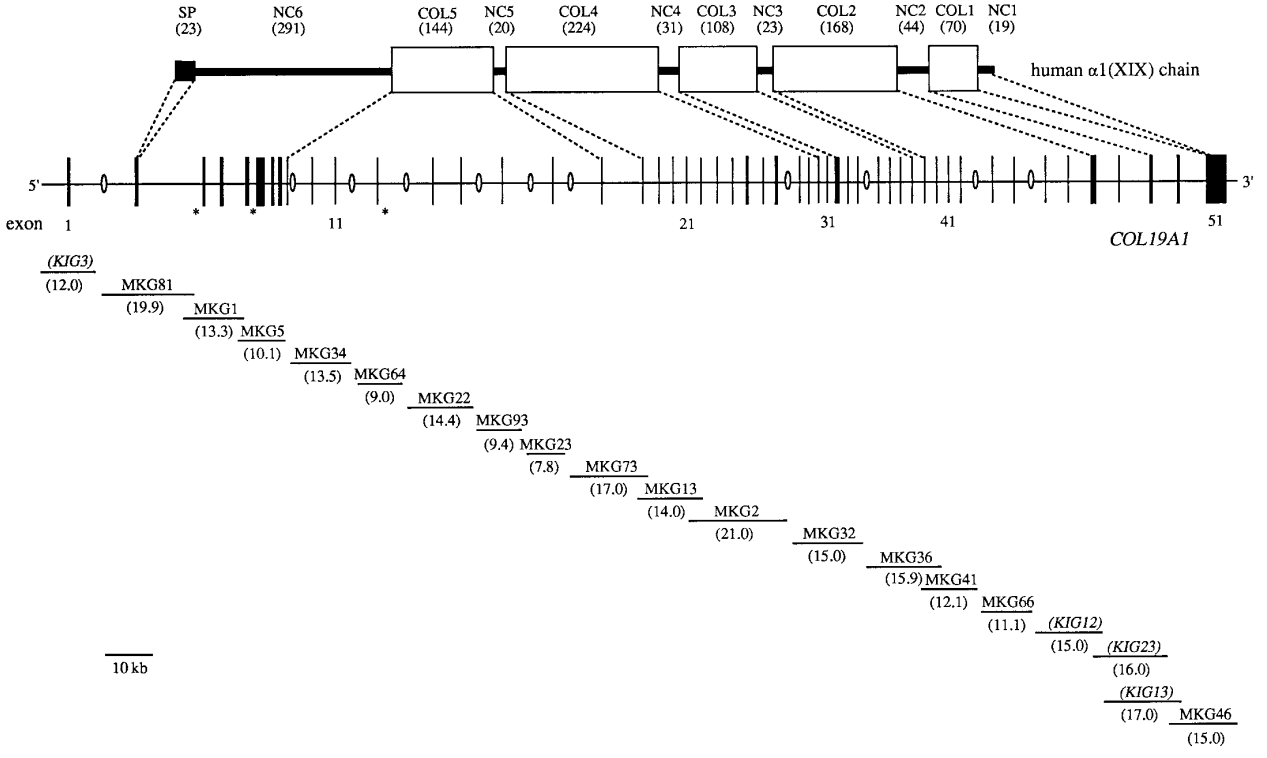

FIG. 1. Structure of the human α1(XIX) gene, *COL19A1*, and relative location of genomic clones. The *COL19A1* gene contains 51 exons, which are indicated by black boxes in the middle. The gene spans at least 250 kb in size. Oval-shaped circles among introns show that those regions are not covered by the 20 clones isolated and shown. Asterisks in introns 2, 5, and 12 indicate the locations of the CA repeats. Relative locations of 20 phage clones are shown by horizontal bars with their sizes in kilobases in parentheses underneath. Clones KIG 3, KIG, 12, KIG 23, and KIG 13 (shown in italic) were previously described (Inoguchi *et al.*, 1995). The closed box, horizontal bars, and open boxes at the top indicate a signal peptide, noncollagenous (NC), and collagenous (COL) domains, respectively. Numbers in parentheses show amino acid residues in the individual domains of the signal peptide, the COL 1 to 5, and the NC 1 to 6 domains.

organization of a human FACIT collagen gene has been determined. In fact, only the chick and mouse α2(IX) genes have been completely characterized so far. The mouse and chick α2(IX) genes span 16 and 10 kb, respectively, and contain 32 exons that are rather compact compared to other collagen genes (Ninomiya *et al.*, 1990; Perala *et al.*, 1994). In contrast to these genes, the chick α1(IX) gene, which was partially characterized with four genomic clones containing 16 exons at the 5′ end and 2 exons at the 3′ end, was estimated at approximately 100 kb in size (Ninomiya *et al.*, 1990). Like the fibrillar collagen genes whose sizes vary from 18 (*COL1A1*) (Chu *et al.*, 1984) to 750 kb (*COL5A1*) (Takahara *et al.*, 1995), the sizes of the genes belonging to the FACIT family also vary considerably.

The sizes of exons and the exon/intron boundaries are summarized in Fig. 2. The 5′ gt and the 3′ ag ends of the introns are all conserved. Like other collagen genes, the 9-bp rule is basically preserved in the exons covering five COL domains of the chain. Exon sizes vary from 27 (exon 40) to 171 bp (exon 47). Split codons are found in exons 2–5 and exons 48–51, even though the 3′ end of exon 49, exon 50, and the 5′ end of exon 51 encode a collagenous domain (COL 1 domain). Among other FACIT genes, split codons have been found, in the portions encoding the NC 4 domain of the chick α1(IX) gene and the NC 2–COL 1–NC 1 domain of *Col9a2*, chick α1(IX), and chick α1(XII). Split glycine codons are never seen in the exons encoding the central collagenous domain of the fibrillar collagen genes, but they are sometimes found in the exons encoding a collagenous domain with the imperfection(s) of Gly-X-Y or a noncollagenous domain harboring a collagenous domain. For instance, more than half of the exons encoding collagenous domains in type IV collagen α chains start with the second nucleotide of the glycine codon, which is conserved throughout evolution (Oohashi *et al.*, 1995). Likewise, these split codons are also conserved at the 5′ and 3′ ends in the FACIT family.

The first exon of the *COL19A1* gene encodes only 5′ untranslated sequence. Usually, the first exon in collagen genes includes the first Met residue. In a couple of collagen genes, including mouse and human α2(IV) (Buttice *et al.*, 1990; Hostikka and Tryggvason, 1987) and chick and human α1(X) collagen genes (Lu-Valle *et al.*, 1988; Reichenberger *et al.*, 1992), the first exon encodes only the 5′ untranslated sequence. Exon 2 contains part of the 5′ untranslated region, the putative signal peptide, and 8 amino acids of the amino-terminal portion of the NC 6 domain. The large portion of the NC 6 domain is encoded by exons 3–8, whose sizes vary from 75 (exon 3) to 276 bp (exon 6). Exon 6 is the largest among the coding exons. COL 5, with 144 amino acids, is encoded by exons 9–16 and the 5′ part of exon 17. There is a discontinuity in the Gly-X-Y repeat at the end of exon 14 due to the Ala for Gly substitution. The NC 5 domain consists of 20 amino acids and is encoded by the 3′ end of exon 17 and the 5′ end of exon 18. The COL 4 domain, with 224 amino acids, is the largest COL domain and is encoded by the 3′ end of exon 18 and exons 19–30. Exons 20, 25, and 27 contain one, two, and one interruption, respectively, of the Gly-X-Y. The NC 4 domain, consisting of 31 amino acids, is encoded by exon 31 and the 5′ end of exon 32, which contains two Cys codons. The three chains of type IX collagen have a similar amino acid sequence (Cys-Pro-X-X-Cys-Pro-X) at the end of the NC 3 domain. These Cys residues could possibly form a disulfide interaction with other chains (van der Rest *et al.*, 1985). A similar sequence (Cys-X-X-Cys) is in the same region of α1(XIX) and in the NC4 domain of the α1(XVI) chain (Pan *et al.*, 1992). This sequence may participate in interchain bridges. The COL 3 domain consists of 108 amino acids and contains an interruption of Gly-X-Y. It is encoded by the 3′ end of exon 32, exons 33–37, and the 5′ end of exon 38. The NC 3 domain (23 amino acids) is encoded by the 3′ end of exon 38 and the 5′ end of exon 39. The 45, 51, and 45-bp exons that Myers *et al.* (1994) reported correspond to exons 37–39. It should be noted that 1 nucleotide has to be added to the sequence of the 3′ end of intron 36 (Myers *et al.*, 1994). This was confirmed by additional direct sequencing of PCR products generated from three different sources of human genomic DNA as templates. The COL 2 domain (168 amino acids) is encoded by the 3′ end of exon 39 and exons 40–47. Exon 40 is 27 bp and the smallest of all exons, whereas exon 47 (171 bp) is the largest exon that encodes collagenous domain. The NC 2 domain (44 amino acids) is encoded by exon 48 and the 5′ end of exon 49. The COL 1 domain is 70 amino acids, the smallest collagenous domain, and contains two interruptions. It is encoded by the 3′ end of exon 49, exon 50, and 5′ end of exon 51. Two imperfections, namely Gly-X sequence in exon 50 and Gly-X-Y-X-Y sequence in exon 51, are conserved among FACIT genes. In exon 50, the nucleotide and the location of the imperfection matched exactly with exon 31 of *Col9a2* and the second exon from the 3′ end of the chick α1(IX) gene, showing 63 and 52% identity and 79 and 76% similarity, respectively, in amino acids. The NC 1 domain consists of 19 amino acids and is encoded by a part of exon 51. The Cys residue at the end of the COL 1 domain and another at position 5 of the NC 1 domain are conserved among chains of the FACIT. Despite this similarity, Myers *et al.* (1994) suggested that these two Cys residues in α1(XIX) chain are involved in an intrachain interaction, whereas those in the other FACIT chains participate in interchain disulfide bonds. The large portion of exon 51 encodes a long 3′ untranslated region 5.3 kb in size.

Evolutionary Similarities between COL19A1 and Collagen IX Genes

Although the overall sizes are different (more than 250 kb vs 16 kb), the exon structures of *COL19A1* and *Col9a2* are similar. The structure of exons encoding the NC 1–COL 1–NC 2–COL 2–NC 3–COL 3–NC 4

Exon	Exon-Intron boundaries		
1	gcgtccccttcccc	ACTCGCAGGG (85) CGAATTCAAG	**gt** aagctgcgcggct
2	ttttttctgttgc **ag**	ATCCGTGGCC (123) GAC AAG ACA G	**gt** atccaggccaact
3	tttctttttaaat **ag**	AA GAG TCA TGC (75) GAA GTT TCA G	**gt** aggcaatataatc
4	tgttacattttac **ag**	GT TTT GAT CTA (100) AGA GAC ACT AT	**gt** aagtaaaaaatta
5	gttttcctcccac **ag**	T AAG ATA TTT (124) ATT CCA CAG	**gt** aaagtaccattag
6	tttacatttttgc **ag**	ATT TCT ATA (276) CCT GTG GAT	**gt** aagttgtgatgtt
7	ctaatttttcat **ag**	ATT GAA CTT (81) GAT ACT AAG	**gt** aagttaatttct
8	ctttttggatttt **ag**	TGC CCA GAG (126) CCA AAC AAG	**gt** atgctagttttaa
9	tgcatttttgctc **ag**	GGA GAA GCA (63) GGA GAG CCG	**gt** aagaaaaaaaca
10	gtgtacttctttt **ag**	GGT GAA AAT (45) GGT CAA AAG	**gt** aaagagttcttga
11	attcctcttctac **ag**	GGA GAG CAA (45) GGT GAA AAG	**gt** aaatatctctttt
12	attgtttcttttt **ag**	GGT GAA CAA (54) GGA GAA AAT	**gt** aagcctaactctt
13	atattctttctac **ag**	GGT TTG AAA (54) GGA GAA AAG	**gt** attgtgtttaccc
14	tttaatctatttt **ag**	GGA GAT ACA (36) GCC TTA CCT	**gt** aagtattcttgaa
15	tctttttcctcat **ag**	GGT TCC CTG (54) GGT CAG AGG	**gt** aagtaaagctgga
16	ttctttggtttca **ag**	GGA AGA CGA (54) GGA CCA CCT	**gt** gagtaagcaatac
17	ttgtttataatac **ag**	GGA CCT CCT (63) GAT AAC AAG	**gt** atggcttcttttt
18	gtttttcaccccct **ag**	GGA AAT GAT (42) GGA GAC AAG	**gt** aatcagatttttt
19	gctttgtcttgaa **ag**	GGT GAA ACT (63) GGA GAA CCT	**gt** aagtattttagct
20	acccctttctcctt **ag**	GGA GAG CCT (36) GGA GAT AGA	**gt** aagtagatatttt
21	ttattttatttac **ag**	GGA GAA CCT (36) GGA GTA AAG	**gt** aatttcctggcat
22	cccccacgtgttt **ag**	GGT GAA CCT (54) GGA CTA AAG	**gt** atataagaaataa
23	caccttttatttt **ag**	GGT CAG CAA (54) GGA GAT GTT	**gt** atgtataatgtct
24	ctctgagttttct **ag**	GGA TTG CCA (54) GGA GAA AAG	**gt** atagtttacattt
25	ccttgttttctac **ag**	GGA GAT CCA (90) GGG GAA CCA	**gt** aagtattagccct
26	ctttttttctttt **ag**	GGA TTA GAT (45) GGG CCA AAG	**gt** atacaaatattat
27	cctttcatttcac **ag**	GGT GAA AGA (78) GGC AGA ACA	**gt** aagtgaaattcat
28	ctcatttgtactc **ag**	GGC GCC CAA (36) GGT ATT CAG	**gt** aagctatttaact
29	tccctccttttcc **ag**	GGT CCT CGA (54) GGG AAT GAT	**gt** aaggactttcttt
30	tacctctgttttc **ag**	GGA GTT CCA (54) GGT GAC CCG	**gt** atgtagacaaacc
31	tcttcttaaccac **ag**	ATT GCA CTT (42) TTG CTC AAG	**gt** actctattgctat
32	ctttatactcatt **ag**	AAT TTC TGT (105) GCC CGG AAG	**gt** gagaagcctggct
33	tcctctgtcttct **ag**	GGT GAT ATA (54) GGA CCA AAG	**gt** aagaaattctctc
34	gtattgtctttttt **ag**	GGA AGC AAA (54) GGT GAT GAG	**gt** aacagattctttt
35	aactatcttttcc **ag**	GGT CTT CAA (54) GGA CCC CCT	**gt** aagtatttgttaa
36	ttttgtgttttac **ag**	GGC TTA ATG (54) GGT GAA AAG	**gt** acaaaggaaaagc
37	atatatccc̲ttgc **ag**	GGC AGC GAC (45) GGA CCA CCT	**gt** gagttgttctagg
38	cacctctttcctta **ag**	GGT ATT CCA (51) AAA ATT AAG	**gt** atttatatttgta
39	cattttcttttga **ag**	GGA GGT GTG (45) GGT CCT CCT	**gt** aagtacagttgtt
40	atgtttctcttac **ag**	GGC CCA AAA (27) GGC CCA GTG	**gt** atgaatgttccca
41	gttcatatttaac **ag**	GGA GAG CCT (54) GGT GTA AAG	**gt** aagcacagaagtt
42	tttaaatcccaac **ag**	GGA GAT CGA (45) GGG ATG TCG	**gt** gagttcagattac
43	aatttgccttgcc **ag**	GGA AAA CCT (45) GGG GAA CCG	**gt** gagttggcagtta
44	tggatgtgtttat **ag**	GGT GAG AGA (63) GGA AAG CCA	**gt** aagtacttcttac
45	ttttttctttat **ag**	GGA ATA AAT (36) GGT GCT CAG	**gt** atgggaaatatga
46	acatctgtttttc **ag**	GGC ATC ATG (45) GGA GAA CGT	**gt** atgtatattacta
47	ttttttccttaac **ag**	GGT GAT CAG (171) GGC ATT CCG	**gt** aagtagtgctaag
48	tttttccttcttc **ag**	GCT GAT GCA (67) ATT TTT GAA G	**gt** tagattttcttaa
49	cttctatacatga **ag**	AG AGG ATG GCT (129) GGA CCC CAA G	**gt** aagtcttcaagtg
50	atatttctcactt **ag**	GC TAC AGA GGA (78) GGG ACT TCA G	**gt** aagtgggatattg
51	tttttatgtcgtt **ag**	CT CTG GGT TTG (5300) TCAGCACATT	aatgtctttaaattc

FIG. 2. Exon/intron junctions of the *COL19A1* and the sizes of the exons. The nucleotide sequences of exon/intron boundaries are shown by uppercase letters for exons and lowercase letters for introns. The 5′ gt and the 3′ ag of introns are given in boldface. Nucleotide sequences of exons are shown by a block of three letters for codon and continuous letters for untranslated region. The total numbers of nucleotides in each exon are shown in the middle in parentheses. Underlined nucleotide c in intron 36 was missing in a previous report (Myers *et al.*, 1994).

domain of α1(XIX) chain is similar to that of the NC 1–COL 1–NC 2–COL 2–NC 3 domain of α2(IX) chain. This is based on the relative sizes of exons encoding each domain, the positions of the exons encoding the portion containing Cys residues, and the locations of exons with split codons. Although the exon encoding this portion of the chick α1(IX) gene is still unavailable, we predict that it is probably similar to that of the *Col9a2*/chick α2(IX) gene. However, there is an exception in this region. No corresponding portion to the NC 3 domain of α1(XIX) was found in α2(IX) chain. Myers *et al.* (1994) pointed out that the NC 2 and the NC 4 domains of α1(XIX) chain are similar to those of α1(XVI) chain, another FACIT member, but that the NC 3 domains of both chains do not resemble each other. The portion encoding the NC 3 domain of *COL19A1* may have evolved from a different origin. In a similar region, chick α2(IX) and α3(IX) genes may

have been changed during evolution (Ninomiya *et al.*, 1990; Har-El *et al.*, 1992). Exons 19–24 in the chick α2(IX) gene have split codons, which are not present in *Col9a2*. Likewise, the chick α3(IX) chain has an interruption of Gly-X-Y in the COL 2 domain, which is not present in chick α1(IX) and α2(IX) chains.

The structure of the exon encoding the COL 5 domain of α1(XIX) chain seems to be similar to that of the COL 3 domain of both α1(IX) and α2(IX) chains. Likewise, the NC 6 domain of α1(XIX) resembles the NC 4 domain of α1(IX). The primary structure of the NC 6 domain of α1(XIX) chain is similar to that of the NC 4 domain of α1(IX) chain, the NC 3 domain of α1(XII) chain, the NC 3 domain of α1(XIV) chain, and the NC 11 domain of α1(XVI) chain. It has been suggested that they may have a β-sheet structure and may be involved in molecular recognition functions (Inoguchi *et al.*, 1995; Moradi-Ameli *et al.*, 1994). In addition, this domain is also conserved in the minor fibrillar collagen chains, i.e., α1(V) and α1(XI), in which this domain is called PARP (proline–arginine-rich peptide) or the thrombospondin N-terminal module (Brown and Timpl, 1995). The gene structure of the NC 6 is slightly similar to that of PARP with regard to the existence of exons encoding the portion containing conserved Cys residues (Takahara *et al.*, 1995). Probably, the portions encoding these NC domains share a common origin. The structure of the exon coding for the COL 4 and the NC 5 domains of α1(XIX) is less similar than any COL and NC domain of α(IX). Additionally, the portion encoding the COL 4 domain seems to have been generated by duplication of the 9-bp exon.

Allele Detection in COL19A1

CA repeat sequences were found in three of the genomic fragments. These CA repeats were located 1.9 kb upstream of exon 3 (i.e., intron 2), 120 bp upstream of exon 6 (i.e., intron 5), and 1.0 kb downstream of exon 12 (i.e., intron 12) (Fig. 1). These three microsatellites of *COL19A1* from 50 unrelated individuals showed polymorphic variations in a Japanese population. As shown in Table 2, the markers, COL19MK1, COL19MK2, and COL19MK3, indicated six alleles with 0.68, five alleles with 0.50, and six alleles with 0.40 heterozygosity. Expected heterozygosity and PIC (polymorphism information content) values (Botstein *et al.*, 1980) were 0.62 and 0.56, 0.42 and 0.39, and 0.45 and 0.43, respectively. The combined heterozygosity was 0.82, indicating that these markers are useful for linkage study of heritable connective tissue diseases.

Recently, mutations of *COL9A2* in humans (Muragaki *et al.*, 1996) and of *Col9a1* in mice (Nakata *et al.*, 1993; Faessler *et al.*, 1994) have been reported. Contrary to mutations in the type II (Kivirikko, 1993) or XI collagens (Li *et al.*, 1995), which are coordinately expressed, the type IX collagen mutations were not lethal but showed progressive changes in the cartilage of the aging organism. This finding suggested that type IX collagen contributes to the structural integrity of cartilage and/or maintains cartilage homeostasis. The

TABLE 2

Informativeness of New Markers for *COL19A1* Gene

Marker and allele	Size (bp)	Allele frequency[a]	Heterozygosity[b]	Expected heterozygosity[c]	PIC[d]
COL19MK1					
A1	112	0.07			
A2	110	0.04			
A3	108	0.04	0.68	0.62	0.56
A4	102	0.32			
A5	100	0.52			
A6	96	0.01			
COL19MK2					
B1	204	0.03			
B2	202	0.13			
B3	198	0.03	0.50	0.42	0.39
B4	196	0.06			
B5	194	0.75			
COL19MK3					
C1	204	0.01			
C2	202	0.08			
C3	200	0.73	0.40	0.45	0.43
C4	198	0.08			
C5	196	0.07			
C6	194	0.03			

Note. Mendelian inheritance: three markers show codominant inheritance (three families, nine meioses).

[a] Studied in 50 unrelated Japanese people.

[b] Combined observed heterozygosity 0.82.

[c] Expected heterozygosity was calculated by the formula $1 - \Sigma(P_i)^2$, where P_i is the frequency of allele i.

[d] PIC, polymorphism information content value (Botstein *et al.*, 1980).

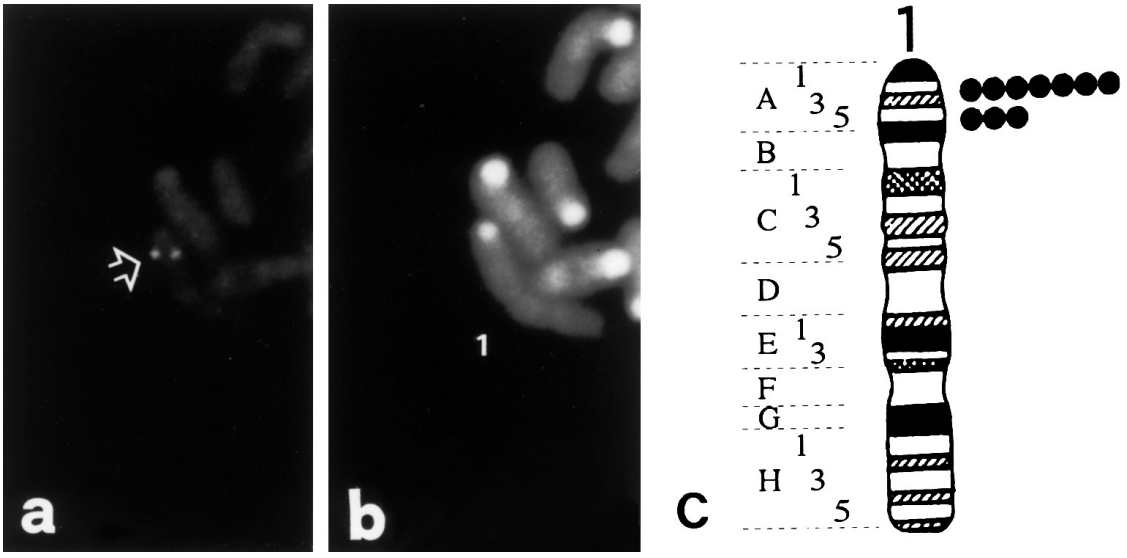

FIG. 3. FISH mapping. (**a** and **b**) The FISH signals on chromosome 1 and the same mitotic figure stained with DAPI to identify chromosome 1, respectively. (**c**) A diagram of FISH mapping, in which each dot represents the signal detected on mouse chromosome 1.

Col19a1 is widely expressed during mouse embryogenesis, but restricted to a few adult tissues such as brain, eye, and testis (Sumiyoshi *et al.*, 1997). We have previously located the *COL19A1* to chromosome 6q12–q14, close to where Stargadt disease type 3 maps as well (Stone *et al.*, 1994). This autosomal dominant disorder features central vision loss due to age-related macular degeneration. In one family, multipoint linkage analysis resulted in a lod score of 6.2 in the interval between markers D6S313 and D6S252. The gene expression and gene mapping data make *COL19A1* a strong candidate gene for congenital degenerative diseases affecting eye and/or brain. Thus, it would be intriguing to analyze *COL19A1* of affected persons in this family.

Chromosomal Mapping of Col19a1

COL19A1 and *COL9A1* reside on human chromosome 6q12–q14 (Yoshioka *et al.*, 1992). The other FACIT member, *COL12A1*, is also located on chromosome 6 (Oh *et al.*, 1992). However, *Col9a1* and *Col12a1* are located on chromosomes 1 and 9 of mouse, respectively (DeBry and Seldin, 1996). To determine how closely *Col19a1* links to *Col9a1* or *Col12a1*, chromosomal mapping of *Col19a1* was performed.

A mouse genomic DNA fragment of 14 kb containing exons 5 and 6 was used as a probe for FISH. Under the conditions used, hybridization efficiency was 94%. (Among 100 checked mitotic figures, 94 of them showed signals on one pair of the chromosomes). Since the DAPI banding was used to identify the specific chromosome, the assignment between signal from probe and mouse chromosome 1 was obtained. An example of the mapping results is presented in Fig. 3. The detailed position was further determined based on a summary of 10 photos. According to the summary, this gene is located on chromosome 1, region A3 (Fig. 3C).

Among the human genes assigned to human chromosome 6, only two mouse counterparts, *Bpag1* and *Col9a1*, are localized to mouse chromosome 1, at positions 15.0 and 16.5 cM, respectively. *Col19a1* was assigned to the same portion of mouse chromosome 1. *Col19a1* and *Col9a1* might be duplicated from an ancestor gene on that chromosomal region. However, we do not know yet whether these two genes are localized with head to head arrangement or head to tail arrangement. Among collagen genes, the former arrangement is seen in *COL4A1/COL4A2*, *COL4A3/COL4A4*, and *COL4A5/COL4A6* on chromosomes 13, 2, and X, respectively (Oohashi *et al.*, 1994), and the latter in *COL6A1/COL6A2* on chromosome 21 (Heiskanen *et al.*, 1995). A search for mouse mutants with eye and/or brain abnormalities failed to identify candidate phenotypes for *Col19a1* mutations. However, many mutations responsible for congenital degenerative eye and/or brain disease have yet to be located on the mouse genome.

CONCLUSION

In conclusion, we have elucidated the genomic organization of *COL19A1*. From data on the primary structure of α1(XIX) collagen chain, the gene structure of *COL19A1* resembles that of type IX collagen genes. These data suggest that this gene has probably arisen from an ancestor gene of the FACIT family. The fine structures of *COL19A1* and its polymorphic markers as well as gene mapping of *Col19a1* will facilitate analysis of hereditary diseases caused by the mutated gene in human and mouse.

ACKNOWLEDGMENTS

The authors are very much indebted to Dr. F. Ramirez and his graduate student, Ms. H. Zhang, with whom we started this project.

We are also grateful to Dr. Henry Heng (SeeDNA Biotech, Inc.) for his excellent mouse chromosome mapping technique and Dr. Ernst Reichenberger (Harvard Medical School) for critical comments to the manuscript.

REFERENCES

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32:** 314–331.

Brown, J. C., and Timpl, R. (1995). The collagen superfamily. *Int. Arch. Allergy Immunol.* **107:** 484–490.

Buttice, G., Kaytes, P., D'Armiento, J., Vogeli, G., and Kurkinen, M. (1990). Evolution of collagen IV genes from a 54-base pair exon: A role for introns in gene evolution. *J. Mol. Evol.* **30:** 479–488.

Chu, M.-L., de Wet, W., Bernard, M., Ding, J.-F., Morabito, M., Myers, J., Williams, C., and Ramirez, F. (1984). Human proα1(1) collagen gene structure reveals evolutionary conversation of a pattern of introns and exons. *Nature* **310:** 337–340.

DeBry, R. W., and Seldin, M. F. (1996). Human/mouse homology relationships. *Genomics* **33:** 337–351.

Diab, M., Wu, J.-J., and Eyre, D. R. (1996). Collagen type IX from human cartilage: A structural profile of intermolecular cross-linking sites. *Biochem. J.* **314:** 327–332.

Faessler, R., Schnegelsberg, P. N. J., Dausman, J., Shinya, T., Muragaki, Y., McCarthy, M. T., Olsen, B. R., and Jaenisch, R. (1994). Mice lacking α1(IX) collagen develop noninflammatory degenerative joint disease. *Proc. Natl. Acad. Sci. USA* **91:** 5070–5074.

Feng, G.-S., Shen, R., Heng, H. H. Q., Tsui, L.-C., Kazlauskas, A., and Pawson, T. (1994). Receptor-binding, tyrosine phosphorylation and chromosome localization of the mouse SH2-containing phosphotyrosine phosphatase Syp. *Oncogene* **9:** 1545–1550.

Har-El, R., Sharma, Y. D., Aguilera, A., Ueyama, N., Wu, J.-J., Eyre, D. R., Juricic, L., Chandrasekaran, S., Li, M., Nah, H.-D., Upholt, W. B., and Tanzer, M. L. (1992). Cloning and developmental expression of the α3 chain of chicken type IX collagen. *J. Biol. Chem.* **267:** 10070–10076.

Heiskanen, M., Saitta, B., Palotie, A., and Chu, M.-L. (1995). Head to tail organization of the human COL6A1 and COL6A2 genes by fiber-FISH. *Genomics* **29:** 801–803.

Heng, H. H. Q., Squire, J., and Tsui, L.-C. (1992). High-resolution mapping of mammalian genes by *in situ* hybridization to free chromatin. *Proc. Natl. Acad. Sci. USA* **89:** 9509–9513.

Heng, H. H. Q., and Tsui, L.-C. (1993). Modes of DAPI banding and simultaneous in situ hybridization. *Chromosoma* **102:** 325–332.

Hostikka, S. L., and Tryggvason, K. (1987). Extensive structural differences between genes for the α1 and α2 chains of type IV collagen despite conservation of coding sequences. *FEBS Lett.* **224:** 297–305.

Inoguchi, K., Yoshioka, H., Khaleduzzaman, M., and Ninomiya, Y. (1995). The mRNA for α1(XIX) collagen chain, a new member of FACITs, contains a long unusual 3′ untranslated region and displays many unique splicing variants. *J. Biochem. (Tokyo)* **117:** 137–146.

Kivirikko, K. I. (1993). Collagens and their abnormalities in a wide spectrum of diseases. *Ann. Med.* **25:** 113–126.

Li, Y., Lacerda, D. A., Warman, M. L., Beier, D. R., Yoshioka, H., Ninomiya, Y., Oxford, J. T., Morris, N. P., Andrikopoulos, K., Ramirez, F., Wardell, B. B., Lifferth, G. D., Teuscher, C., Woodward, S. R., Taylor, B. A., Seegmiller, R. E., and Olsen, B. R. (1995). A fibrillar collagen gene, *Col11a1*, is essential for skeletal morphogenesis. *Cell* **80:** 423–430.

LuValle, P., Ninomiya, Y., Rosenblum, N. D., and Olsen, B. R. (1988). The type X collagen gene. *J. Biol. Chem.* **263:** 18378–18385.

Moradi-Ameli, M., Deleage, G., Geourjon, C., and van der Rest, M. (1994). Common topology within a non-collagenous domain of several different collagen types. *Matrix Biol.* **14:** 233–239.

Muragaki, Y., Mariman, E. C. M., van Beersum, S. E. C., Perala, M., van Mourik, J. B. A., Warman, M. L., Olsen, B. R., and Hamel, B. C. J. (1996). A mutation in the gene encoding the α2 chain of the fibril-associated collagen IX, *COL9A2*, causes multiple epiphyseal dysplasia (EDM2). *Nature Genet.* **12:** 103–105.

Myers, J. C., Sun, M. J., D'Ippolito, J. A., Jabs, E. W., Neilson, E. G., and Dion, A. S. (1993). Human cDNA clones transcribed from an unusual high-molecular weight RNA encode a new collagen chain. *Gene (Amsterdam)* **123:** 211–217.

Myers, J. C., Yang, H., D'Ippolito, J. A., Presente, A., Miller, M. K., and Dion, A. S. (1994). The triple-helical region of human type XIX collagen consists of multiple collagenous subdomains and exhibits limited sequence homology to α1(XVI). *J. Biol. Chem.* **269,** 18549–18557.

Nakata, K., Ono, K., Miyazaki, J.-I., Olsen, B. R., Muragaki, Y., Adachi, E., Yamamura, K.-I., and Kimura, T. (1993). Osteoarthritis associated with mild chondrodysplasia in transgenic mice expressing α1(IX) collagen chains with a central deletion. *Proc. Natl. Acad. Sci. USA* **90:** 2870–2874.

Ninomiya, Y., Castagnola, P., Gerecke, D., Gordon, M. K., Jacenko, O., LuValle, P., McCarthy, M., Muragaki, Y., Nishimura, I., Oh, S., Rosenblum, N., Sato, N., Sugrue, S., Taylor, R., Vasios, G., Yamaguchi, N., and Olsen, B. R. (1990). The molecular biology of collagens with short triple-helical domains. *In* "Extracellular Matrix Genes" (L. J. Sandell and C. D. Boyd, Eds.), pp. 79–114, Academic Press, San Diego.

Oh, S. P., Taylor, R. W., Gerecke, D. R., Rochelle, J. M., Seldin, M. F., and Olsen, B. R. (1992). The mouse α1(XII)- and human α1(XII)-like collagen genes are localized on mouse chromosome 9 and human chromosome 6. *Genomics* **14:** 225–231.

Oohashi, T., Sugimoto, M., Mattei, M.-G., and Ninomiya, Y. (1994). Identification of a new collagen IV chain, α6(IV), by cDNA isolation and assignment of the gene to chromosome Xq22, which is the same locus for *COL4A5*. *J. Biol. Chem.* **269:** 7520–7526.

Oohashi, T., Ueki, Y., Sugimoto, M., and Ninomiya, Y. (1995). Isolation and structure of the *COL4A6* gene encoding the human α6(IV) collagen chain and comparison with other type IV collagen genes. *J. Biol. Chem.* **270:** 26863–26867.

Pan, T.-C., Zhang, R.-Z., Mattei, M.-G., Timpl, R., and Chu, M.-L. (1992). Cloning and chromosomal location of human α1(XVI) collagen. *Proc. Natl. Acad. Sci. USA* **89:** 6565–6569.

Perala, M., Elima, K., Metsaranta, M., Rosati, R., de Crombrugghe, B., and Vuorio, E. (1994). The exon structure of the mouse α2(IX) collagen gene shows unexpected divergence from the chick gene. *J. Biol. Chem.* **269:** 5064–5071.

Reichenberger, E., Beier, F., LuValle, P., Olsen, B. R., von der Mark, K., and Berling, W. M. (1992). Genomic organization and full-length cDNA sequence of human collagen X. *FEBS Lett.* **311:** 305–310.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shaw, L. M., and Olsen, B. R. (1991). FACIT collagens: Diverse molecular bridges in extracellular matrices. *Trends Biochem. Sci.* **16:** 191–194.

Stone, E. M., Nichols, B. E., Kimura, A. E., Weingeist, T. A., Drack, A., and Sheffield, V. C. (1994). Clinical features of a Stargardt-like dominant progressive macular dystrophy with genetic linkage to chromosome 6q. *Arch. Ophthalmol.* **112:** 765–772.

Sumiyoshi, H., Inoguchi, K., Khaleduzzaman, M., Ninomiya, Y., and Yoshioka, H. (1997). Ubiquitous expression of the α1(XIX) collagen gene (Col19a1) during mouse embryogenesis becomes restricted to a few tissues in the adult organism. *J. Biol. Chem.* **272,** 17104–17111.

Takahara, K., Hoffman, G. G., and Greenspan, D. S. (1995). Complete structural organization of the human α1(V) collagen gene (*COL5A1*): Divergence from the conserved organization of other characterized fibrillar collagen genes. *Genomics* **29:** 588–597.

van der Rest, M., Mayne, R., Ninomiya, Y., Seidah, N. G., Chretien, M., and Olsen, B. R. (1985). The structure of type IX collagen. *J. Biol. Chem.* **260:** 220–225.

van der Rest, M., and Garrone, R. (1991). Collagen family of proteins. *FASEB J.* **5:** 2814–2823.

Yoshioka, H., Zang, H., Ramirez, F., Mattei, M.-G., Moradi-Ameli, M., van der Rest, M., and Gordon, M. K. (1992). Synteny between the loci for a novel FACIT-like collagen locus (D6S228E) and α1(IX) collagen (COL9A1) on 6q12–q14 in humans. *Genomics* **13:** 884–886.