Published in final edited form as:
*Connect Tissue Res*. 2012 ; 53(3): 267–276. doi:10.3109/03008207.2011.636160.

# Characterization of Tissue-Specific and Developmentally Regulated Alternative Splicing of Exon 64 in the *COL5A1* Gene

**Anna L. Mitchell**[a,b,*], **LuAnn M. Judis**[a], **Ulrike Schwarze**[c], **Polina M. Vaynshtok**[b], **Mitchell L. Drumm**[a,b], and **Byers Byers**[c,d]

[a]Department of Genetics, Case Western Reserve University, Cleveland, OH 44106, USA

[b]Department of Pediatrics, Case Western Reserve University, Cleveland, OH 44106, USA

[c]Department of Pathology, University of Washington, Seattle, WA 98195, USA

[d]Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA, 98195, USA

## Abstract

The *COL5A1* gene, a member of the clade B fibrillar collagen gene family, was recently shown to contain two alternatively spliced exons (64A and 64B) that encode 23 amino acids in the carboxyl-terminal propeptide. The two are identical in length, very similar in sequence, and used in a mutually exclusive fashion because of the small intron that separates them. Each *COL5A1* allele uses both exons, but a given transcript will contain only one of the two exons. The sequences in other species are highly conserved at the amino acid level. The expression profile of the two isoforms was determined from analysis of RNA levels in a panel of murine tissues. While both isoforms were found in all tissues studied, actively proliferating tissues (liver, lung) used isoform B more often, while a less mitotically-active tissue, brain, had a higher proportion of exon 64A. The high degree of conservation between the two exons is consistent with a regional genomic duplication. The presence of the two isoforms as far back as pufferfish (tetraodon) implies an important functional significance. The exact role determined by the two sequences is not known, but involvement in the determination of chain composition of mature type V collagen or regulation of cell activity are possible, given the differences in tissue distribution.

### Keywords

Type V collagen; alternative splicing; tissue distribution; developmental profile

## INTRODUCTION

Collagens are the major protein constituents of the extracellular matrix. There are 29 types of collagen encoded by some 40 genes, each type having unique properties and functions.

[*]Corresponding author: Case Western Reserve University, 2109 Adelbert Road, Biomedical Research Building 719, Cleveland, OH 44106-4948; phone +1 (216) 368-4919; FAX +1 (216) 844-7497; alm43@case.edu.

Declaration of Interest
The authors have no conflicts of interest to report.

The fibrillar subgroup of collagens, which have triple helical domains of more than 1000 amino acids in which the Gly-X-Y triplet repeat is retained, can be divided into three clades named A, B, and C [1]. Clade A includes types I, II, and III collagen, as well as the α2 chain of type V, clade B has type XI collagen and the α1 and α3 chains of type V collagen, and types XXIV and XXVII collagen are in clade C. Within each clade, each chain has a similar protein structure and their genes have similar genomic organization. The fibrillar procollagens are trimers, composed of three individual proα chains, in which the central core is a triple helix. The three chains of a trimer initially associate via their carboxyl-terminal propeptide domains, and inter-chain disulfide bonds between portions of the propeptides stabilize the interaction [2]. Nucleation of the triple helix is followed by propagation of chain folding in a carboxyl-to-amino direction (elongation) [3]. Following secretion, the propeptide regions are cleaved to yield the mature collagen molecule that can then incorporate into the fibrils of the extracellular matrix.

Type V collagen is a minor fibrillar collagen that forms heterotypic fibrils with the more prevalent type I collagen and is thought to be important in two ways. First, as shown when the *COL5A1* gene is inactivated, it is crucial for the formation of the large fibrils composed of type I collagen and its absence results in embryonic lethality in the mouse [4]. Second, it is thought that type V collagen is important in the regulation of fibril diameter, as there is an inverse relationship between the amount of type V protein and the resulting fibril size [5, 6]. Type V procollagen most commonly exists as a heterotrimer, composed of two pro-α1(V) chains and one pro-α2(V) chain, encoded by the *COL5A1* (OMIM 120215) and *COL5A2* (OMIM 120190) genes respectively. Two other forms of type V collagen protein have been identified: an $\alpha1(V)_3$ homotrimer [7] and an α1(V)α2(V)α3(V) heterotrimer [8, 9]. The pro-α3(V) chain is encoded by the *COL5A3* gene (OMIM 120216).

The domain thought to be critical for selection of trimer binding partner is located in the C-propeptide of the proα chains. By homology to types I and III collagen, this region is located in the amino acids encoded by exon 65 of the *COL5A1* gene [10]. However, given that the sequence of exon 65 of *COL5A1* is a constant, this region is unlikely to be the sole factor that determines whether a type V collagen molecule is a homotrimer or one of the heterotrimers. This level of selection must come from additional factors. Two possibilities for regulation at this level seem likely—transcriptional/translational control of the production of the partner chains [i.e. pro-α2(V), pro-α3(V)] in a given tissue, or the use of splice isoforms of exon 64, described by Hoffman et al. [11], and further characterized here.

We describe the tissue and developmental alternative splice pattern of two mutually exclusive versions of exon 64, 64A and 64B, in the *COL5A1* gene. These exons encode a 23 amino acid sequence located within the carboxyl propeptide domain of the protein, just amino-terminal to the chain selectivity domain. The function and relative importance of the two isoforms are not known. We sought to learn about the potential roles of these two isoforms by characterizing their expression profile in a panel of murine tissues.

# METHODS

## Cell Culture; Isolation of DNA and RNA

Skin fibroblasts from individuals with classical EDS and a normal individual were cultured as described previously [12]. DNA and RNA were isolated using the appropriate Qiagen kit (QIAmp DNA Mini kit; RNeasy Mini kit; Qiagen, Valencia, CA). Complementary DNA was synthesized from the RNA using the Superscript kit (Gibco BRL).

Mouse genomic DNA was a gift of Dr. Melissa Parisi. Rat genomic DNA was isolated from a rat chondrosarcoma cell line (gift of Dr. James Pace).

## Amplification and sequencing of the 3' end of *COL5A1* cDNA

The 3' end of *COL5A1* cDNA was PCR-amplified using primers x60S 5'-TCCGCCTGGTCCAAAAGGTGCT -3' and 3'UTRA 5'-GCGGCTCCTTCCCTCTGTTCTCT-3'. Reactions were performed with denaturation at 95° for 2 minutes, 35 cycles of 95° for 15 sec, 63° for 20 sec, 72° for 1 min 15 sec, and a final extension of 72° for 5 minutes using an MJ Research thermal cycler (PTC-200). Products were analyzed on a 7% acrylamide gel. Bands of the resulting heteroduplex were cut from the gel and re-amplified, using the same primers and conditions. The products were purified with the QIAquick-spin columns (Qiagen, Valencia, CA), and sequenced with the Big Dye terminator method on the ABI310 (Applied Biosystems).

## Amplification and sequencing of genomic DNA

The region spanning exon 64 was amplified from human genomic DNA using primers in63S 5'-CTTGCTCTGGAGGCCGGAGAAGTAA-3' and in64A 5'-GCAGAGCCGCGAGCCAGCCGT-3'. Reactions were performed with denaturation at 95° for 2 minutes, 35 cycles of 95° for 20 sec, 58° for 20 sec, 72° for 40 sec, and a final extension of 72° for 5 minutes. For mouse and rat genomic DNA, primers used were in63S mus 5'-TCTGTCTGTAGAGGCAGAAGT-3' and in64A mus 5'-GTGATTAGGCCCATGCAGTAC-3'. Reaction conditions were: denaturation at 95° for 2 minutes, 35 cycles of 95° for 20 sec, 60° for 15 sec, 72° for 30 sec, and a final extension of 72° for 5 minutes. All reactions were done on an MJ Research thermal cycler (PTC-200). Products were purified with the QIAquick-spin columns (Qiagen, Valencia, CA), and sequenced with the Big Dye terminator method on the ABI310 (Applied Biosystems).

## Database comparisons

The cDNA sequences for other fibrillar collagen family members related to *COL5A1* were obtained from NCBI Entrez Nucleotide (http://www.ncbi.nlm.nih.gov/sites/entrez): *COL5A1*= NM_000093.2; *COL5A2*= NM_000393.2; *COL5A3*= NM_015719.2; *COL1A1*= NM_000088.2; *COL1A2*= NM_000089.2; *COL2A1*= NM_001844.3; *COL3A1*= NM_000090.2; *COL11A1*= NM_001854.2; *COL11A2*= NM_080680.1; *COL27A1*= NM_032888.2. These sequences were then used in BLAT searches (UCSC Genome Bioinformatics; http://genome.ucsc.edu/) to obtain the genomic sequences. In each genomic sequence, the region analogous to the exon 64 region of *COL5A1* was analyzed for evidence of genomic duplication.

The human sequences for *COL5A1* exons 64A and 64B were used in BLAT searches (UCSC Genome Bioinformatics; http://genome.ucsc.edu/) against the genomic databases for chicken, chimp, horse, lizard, opossum, X. tropicalis, and tetraodon to identify the paralogous sequences. Sequences for the different species were aligned with the Clustal W program (http://www.ebi.ac.uk/Tools/clustalw/).

The human sequences were also used to search the Ensembl database to identify the genomic sequences in 29 additional species, including: chimpanzee, gorilla, orangutan, macaque, marmoset, tarsier, small ear galago, guinea pig, kangaroo rat, rat, ground squirrel, pika, rabbit, dog, cat, little brown bat, dolphin, hedgehog, armadillo, lesser hedgehog, elephant, zebrafish, stickleback, Medaka, fugu, zebra finch, turkey, grey mouse lemur, platypus, and pig.

## Animal Tissues

Discarded C57BL/6J mouse tissues were harvested at the following timepoints: E13.5 and 2, 5, 7, 8, 15 and 22 weeks. From three embryos, brain, heart, liver, lung and small intestine were isolated. From the mice the following tissues were isolated: brain, colon, duodenum, eye, heart, ileum, jejunum, kidney, knee, liver, lung, ovary, skeletal muscle, skin, spleen, testes, and uterus. The dissected tissues were immediately frozen in liquid nitrogen and stored at −80°C until processed. The tissues were the generous gifts of Drs. Thomas Kelly, and Joseph Nadeau. All animals were cared for according to Case Western Reserve University approved protocols and Institutional Animal Care and Use Committee guidelines.

Discarded Xenopus laevis tissues were harvested from 1 year old male frogs (heart, liver, lung, small intestine) or 3 year old female frogs (brain), immediately frozen in liquid nitrogen, and stored at −80° until processed. Tissues were a gift from Dr. Keiko Tamai.

Brain, heart, liver, lung and small intestine were dissected from stage 30 and 40 chicken embryos, frozen in liquid nitrogen, and stored at −80° until processed. Tissues were a gift of Dr. Michiko Watanabe.

## RNA extraction and cDNA synthesis

RNA extractions from mouse embryonic tissues, adult frog tissues, and embryonic chicken tissues were performed using the RNeasy Micro Kit (Qiagen) according to manufacturer instructions. The frozen mouse skin tissue was first pulverized in a Mikro-dismembrator (Sartorius Stedim Biotech) at 3000 RPM for 1 minute. The resulting powder was suspended in 150 μl RLT Buffer from the RNeasy Mini Kit (Qiagen) with β-mercaptoethanol (Acros Organics) and processed according to manufacturer directions. RNA from the remaining mouse tissues was extracted with TRIzol Reagent (Invitrogen) and sonicated with the variable speed TissueTearor (BioSpec Products, Inc.) according to manufacturer instructions. RNA from the Trizol method was suspended in the RNA storage solution (Ambion). All RNA samples were stored at −80°C.

The cDNA synthesis reactions were performed with the SuperScript First–Strand Synthesis System for RT-PCR (Invitrogen) according to manufacturer directions. Oligo(dT) primers

were utilized. The template concentration ranged from 1–5 μg total RNA. For the frog and chicken samples, 5 μl RNA was used as template.

### Amplification of exon 64 region of *Col5a1* from cDNA

The region around exon 64 was amplified from cDNA using the following primer sequences: a) for mouse: col5a1 x63F mus 5' 6-FAM GCGAATACTGGGTCGATCCCA-3' (fluorescently labeled) and col5a1 x65R mus 5'- CCACGGACTGGTAGCAGTTGT-3'; b) for frog: col5a1 frog F 5' 6-FAMGGAATATTGGATTGACCCA-3' (fluorescently labeled) and col5a1 frog R 5'- CCACGGACTGGTAGCAGTTGT-3'; c) chicken: col5a1 x63F 5' 6-FAM GCGAATACTGGGTCGATCCCA-3' (fluorescently labeled) and col5a1 x65R 5'-CCACGGACTGGTAGCAGTTGT-3'. Polymerase chain reactions were performed with denaturation at 95° for 2 minutes, 20 to 40 cycles of 95° for 15 sec, 60° for 15sec, 72° for 20sec, and a final extension of 72° for 5 minutes using an MJ Research thermal cycler PTC-100. The number of cycles was maintained in the linear range to avoid heteroduplex formation and confirmed with 7% acrylamide gel. The 5' primer was fluorescently labeled with (6-FAM), permitting analysis of products on the ABI310 (Applied Biosystems).

### *Eco*RI digestion

An *Eco*RI site is present at position 49 of exon 64A only, resulting in 135 bp and 162 bp bands after digestion of the PCR product amplified from that isoform. The isoform with exon 64B remains intact. Half of the total volume of the PCR (from above) was digested with 5U *Eco*RI (Invitrogen) in 20μl at 37° for 1hour. Reactions were spiked with 25 ng pSPL3 DNA (6kb; Invitrogen) to monitor for complete digestion. Samples were visualized on a 0.7% agarose gel before analysis on the ABI310 (Applied Biosystems).

### ABI run and calculations

The digested PCR products (0.5–1μl) were analyzed on the ABI310 with 0.5μl ROX500 HD marker (Applied Biosystems) and 9ul HiDi formamide (Applied Biosystems). The areas under the peaks, representing the uncut product and the digested product, were calculated by the ABI software (GeneScan, version 3.1.2). These numbers were used to calculate the proportion of *Col5a1* message that contained the exon 64B isoform. Proportion of isoform B= area of uncut peak/ (area of uncut peak + area of digested peak).

### Statistical analysis

The proportion of *Col5a1* exon 64B for each tissue was stratified by sex and age. The data from each tissue were analyzed by the 1-Way Analysis of Variance (ANOVA) statistical method, with significance defined at less than 0.05. Because there were no significant differences between males and females, data for the sexes were combined at all time points.

A t-test was done for isoform expression data from the tissues of young adult mice (7–15 weeks). There were no statistically significant differences between time points for any tissue so values were combined for each tissue.

# RESULTS

## Characterization of the mRNAs encoded by the *COL5A1* gene

When cDNA samples from primary human fibroblast cultures of subjects being screened for classical EDS disease-causing mutations were used as templates for amplification of the 3' region (exon 60-3'UTR) of the *COL5A1* cDNA, a heteroduplex was identified in every sample examined by polyacrylamide gel electrophoresis (Figure 1A), indicating a mixed sequence population. This was true even for samples from individuals whose fibroblasts stably expressed only one *COL5A1* allele as a result of a mutation that led to degradation of the product of the second allele through nonsense mediated decay (haploinsufficiency; Figure 1A, lanes 1–4).

cDNA synthesized from the cytoplasmic RNA fraction expressed by fibroblasts from haploinsufficient patients was used as a template for sequencing of the exon 64 region. Previous work had demonstrated that product from the mutant allele was not detectable in this RNA fraction [13]. Both exon 64 isoforms were detected in all patients (two examples shown in Figure 1B). Thus, each *COL5A1* allele expresses an mRNA that is alternatively spliced.

## Characterization of the alternatively spliced exons

Sequencing of the products that resulted from re-amplification of the heteroduplex identified two versions of exon 64- now called exon 64A, which is annotated in the sequence databases, and exon 64B, previously identified as intronic sequence. Both are 69 basepairs in length, but differ at 26 sites (Figure 2A), suggesting that they are the result of a regional genomic duplication of some antiquity. Eleven of the 23 amino acids encoded by the exons are identical (Figure 2A). There are two clusters of identity- one with four residues (positions 6–9 Tryptophan-Proline-Lysine-Glutamic acid [W-P-K-E]) and the other with three residues (19–21 Lysine-Arginine-Glycine [K-R-G])- that could represent regions of conserved function between the two isoforms. Most of the amino acid differences between the two isoforms represent conservative changes.

The two exons are separated by 63 basepairs in the human and 61 in the mouse and rat (intron 64A; IVS64A) (Figure 2B). The usually accepted minimum distance between two exons that allows both to be incorporated into the mRNA is thought to be about 80 nucleotides [14]. Thus, although both 64A and 64B have functional acceptor and donor sites, they cannot both be incorporated into a single transcript.

## Analysis of sequence conservation in Clade B and other fibrillar collagens

The genomic sequences of the other members of Clade B fibrillar collagens were searched, as were those of the other fibrillar collagen family members, in the region analogous to that of *COL5A1* exon 64 for evidence of duplication. None of the other clade B genes, nor the genes of the other fibrillar collagens in Clades A and C, contained duplicated sequences that would encode alternatively spliced exons in this region.

BLAT searches of other species identified the regions paralogous to *COL5A1* exons 64A and B in chicken, chimp, horse, lizard, mouse, opossum, xenopus tropicalis, and tetraodon. There is extensive homology among the sequences, as the clustal W alignment shows identical nucleotides at many positions in all nine species (Figure 3A).

The protein sequences are highly conserved, with all species except tetraodon having identical amino acid sequence for exon 64A (Figure 3B). Exon 64B is not as highly conserved, as only chicken, horse, human and mouse have the identical sequence. Interestingly, the chimp sequence differs from the human by one amino acid- serine instead of threonine at residue 13. All seven other species have a threonine at this position. The fact that the two isoforms are found in such divergent species, indicates that the regional genomic duplication occurred millions of years ago. The degree of amino acid sequence identity that has been maintained between species suggests a functional significance for the two exons.

Searches of the Ensembl database identified the regions paralogous to *COL5A1* exons 64 and B in 29 additional species (Table 1). The size of the intervening intron (IVS64A) is highly conserved, ranging from 52 to 65 base pairs in length. In all species, the intronic size is below the threshold level for inclusion of both exons in a single mRNA transcript. The evolutionary constraint to maintain alternative usage of the two exons appears to be strong.

## Tissue Expression

To gain insight into the function of the isoforms, the relative amounts of the two transcripts were measured in a panel of young adult mouse tissues from animals between 7 and 15 weeks of age. The fraction of *Col5a1* message that contained the exon 64B isoform was calculated for 17 different mouse tissues. Calculations were done for both male and female mice and data were combined because no significant differences in ratio were found between the two sexes (Figure 4A). The proportion of isoform B varied from a low of 0.06 +/− 0.03 in the brain to a high of 0.77 +/− 0.05 in the liver (Figure 4B).

Both exon 64 isoforms was expressed in adult frog brain, heart, liver, lung and small intestine, as well as embryonic chicken brain, heart, liver, lung, and small intestine (Table 1).

## Developmental Expression Profile

The tissue differences observed in mouse led us to investigate the developmental regulation of isoform expression in brain, heart, liver, lung, and small intestine from day 13.5 embryos, and animals of 2, 5, and 22 weeks of age.

A and B isoforms were identified in all five mouse embryonic tissues studied (Figure 5). The proportion of isoform 64B ranged from a low of 0.07 in brain to a high of 0.32 in the small intestine (ileum) at stage E13.5. As the mice aged, the expression profiles fell into 3 different patterns. The brain had a relatively fixed proportion of isoform B over time. For the heart and small intestine, the proportion of isoform B changed little between E13.5 and 2 weeks, but there was an increase in the relative amount of B between 2 and 5 weeks. This amount decreased slightly between 5 and 7/8 weeks before stabilizing through 22 weeks.

For the liver and lung, there was an early, sharp increase in the proportion of isoform B between E13.5 and 2 weeks, a smaller increase between 2 and 5 weeks, and a leveling off from 5 weeks through 22 weeks.

## DISCUSSION

Recently, Hoffman et al. [11] reported the sequence of the zebrafish *Col5a1* gene and noted the alternative splicing of exon 64 in that species. A high degree of conservation is maintained and the fish sequence is quite similar to that of tetraodon (pufferfish). In all species, the small intron between the two versions of exon 64 is not large enough to permit splicing and inclusion of both exons into a single message [14] so only one is included in each transcript. Expression of both exons 64A and 64B occurs from a single *COL5A1* allele, and both exons are never present in the same *COL5A1* transcript. The mechanism by which this regulation occurs is not known.

Our murine expression studies showed that while there were tissues that incorporated low levels of isoform B into the mRNA, all tissues used at least 20% isoform A (Figure 4). The tissues that used the highest proportion of isoform B (liver 77%; kidney 65%; lung 64%) are all organs involved in purifying air or blood for the body and, therefore, are susceptible to environmental exposures and the potential resulting damage. Cells within these three organs have a relatively high degree of activity (growth and division). It is interesting to note that these are three organs that are most prone to fibrosis, either idiopathic or as a result of infection and inflammation. It remains to be determined whether this predisposition to fibrosis is related to the relatively high use of the exon 64B isoform by these organs.

The use of the two exons is regulated during development. As mouse development progresses, the proportion of isoform B increased for most tissues studied (Figure 5), except in the brain, where the relative proportions of the two isoforms remained fairly constant. These differences likely reflect varied requirements of the tissues for the two forms of the pro-α1(V) chain. The brain appears to use the exon 64A isoform almost to the exclusion of 64B, a requirement that does not change during development. In other tissues, expression of the two isoforms changes as development progresses.

The two isoforms of exon 64 in *COL5A1* are located in the region of the gene that encodes the carboxyl-terminal propeptide domain of the protein, just upstream of the chain selectivity recognition domain encoded in exon 65, a location inferred from the presence of such a sequence in other fibrillar collagen genes (Figure 2B). Given the proximity of the encoded sequences to exon 65, the possibility that the two exon 64 isoforms modify chain selection must be considered. The high degree of sequence conservation among widely divergent species implies a crucial role for these two protein variants in extracellular matrix synthesis. Additional studies to characterize matrix containing varying ratios of isoform 64A:B will be important for further elucidating the role of the variants.

The result of alterations in this region of the *COL5A1* gene is not known. To date, there have been no disease-causing mutations or SNPs identified in the sequences surrounding the two versions of exon 64. Mutations in *COL5A1* are only known to cause classical EDS. It is

possible that variants in one of the alternatively spliced exons could lead to features of classical EDS, or produce a different phenotype. As more individuals with classical EDS have their causative mutations identified, the role of the isoforms of *COL5A1* exon 64 should become more apparent.

This type of alternative splicing (the use of two mutually exclusive exons in transcripts of each allele) is not common. The glutamate receptor subunit 2 gene, *GluR2*, has two mutually exclusive exons called “flip” and “flop”. The near identity of their DNA sequences indicates that they likely arose from a duplication event. Inclusion of “flip” occurs during development, while adult tissues use “flop” [15]. These exons are separated by a 413 bp intron, which is large enough to permit inclusion of both exons in a single message. The β-tropomyosin gene contains two versions of exon 6 that are used in a mutually exclusive fashion in a tissue-specific manner. Exon 6A is used by smooth muscle cells and non-muscle tissue, while exon 6B is used by skeletal muscle [16]. In both these examples, the temporal/spatial usage of the alternative splice products has been defined. In one case it is developmentally-regulated and in the other it is tissue-specific. For the *Col5a1* exon 64 splicing, the alternative usage of the two exons does not appear to be tissue-specific, as all tissues studied utilized both isoforms, although in varying ratios. Less is known about the developmental regulation of the *Col5a1* alternative splicing, as the only embryonic stage studied was E13.5, however both isoforms were expressed at that time in the tissues studied. Because both isoforms derive from each allele, the mechanism of alternative splicing is likely to be regulated trans-acting factors that shape the pre-mRNA prior to engagement with the splicing machinery and are developmentally regulated in a tissue specific manner

## CONCLUSIONS

This identification of an alternative splicing pattern of two mutually exclusive exons that appear to result from a regional genomic duplication is novel for the fibrillar collagen family. These two isoforms do not show the same kind of tissue restriction seen in the alternative splice products of other collagen genes [17–23]. Nor do they seem to have the developmental regulation seen in some other examples of alternative splicing. Both isoforms were detected in all tissues studied, although the proportion of each varied among the different tissues. The extensive degree of amino acid identity among the *Col5a1* exon 64 sequences from different species is highly suggestive of a conserved functional role for these two variants, a role that remains to be identified.

## Acknowledgments

We thank Yongqui Doughman for dissecting the chick embryos and John Dunn for assistance with figures.

This work was supported by grants AR21557, 2 T32 GM 007454-31, AR051444 and HD043376.

## References

1. Exposito JY, Larroux C, Cluzel C, Valcourt U, Lethias C, Degnan BM. Demosponge and Sea Anemone Fibrillar Collagen Diversity Reveals the Early Emergence of A/C Clades and the Maintenance of the Modular Structure of Type V/XI Collagens from Sponge to Human. J BiolChem. 2008; 283:28226–28235.

2. Olsen BR, Hoffmann H, Prockop DJ. Interchain disulfide bonds at the COOH-terminal end of procollagen synthesized by matrix-free cells from chick embryonic tendon and cartilage. Arch Biochem Biophy. 1976; 175:341–350.
3. Bruckner P, Bachinger H, Timpl R, Engel J. Three conformationally distinct domains in the amino-terminal segment of type III procollagen and its rapid triple helix leads to and comes from coil transition. Eur J Biochem. 1978; 90(3):595–603. [PubMed: 710449]
4. Wenstrup RJ, Florer JB, Brunskill EW, Bell SM, Chervoneva I, Birk DE. Type V Collagen Controls the Initiation of Collagen Fibril Assembly. J Biol Chem. 2004; 279(51):53331–53337. [PubMed: 15383546]
5. Birk DE, Fitch JM, Babiarz JP, Doane KJ, Linsenmayer TF. Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. J Cell Sci. 1990; 95(pt 4):649–657. [PubMed: 2384532]
6. Linsenmayer TF, Gibney E, Igoe F, Gordon MK, Fitch JM, Fessler LI, Birk DE. Type V Collagen: Molecular Structure and Fibrillar Organization of the Chicken α 1(V) $NH_2$-terminal Domain, a Putative Regulator of Corneal Fibrillogenesis. J Cell Biol. 1993; 121(5):1181–1189. [PubMed: 8501123]
7. Haralson MA, Mitchell WM, Rhodes RK, Kresina TF, Gay R, Miller EJ. Chinese hamster lung cells synthesize and confine to the cellular domain a collagen composed solely of B chains. Proc Natl Acad Sci USA. 1980; 77:5206–5210. [PubMed: 7001474]
8. Niyibizi C, Fietzik PP, van der Rest M. Human Placenta Type V Collagen. J Biol Chem. 1984; 259(22):14170–14174. [PubMed: 6501291]
9. Rhodes RK, Miller EJ. Evidence for the existence of an alpha1(V)alpha2(V)alpha3(V) collagen molecule in human placental tissue. Coll Relat Res. 1981; 1(4):337–343. [PubMed: 7346227]
10. Lees JF, Tasab M, Bulleid NJ. Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. EMBO J. 1997; 16(5):908–916. [PubMed: 9118952]
11. Hoffman GG, Branam AM, Huang G, Pelegri F, Cole WG, Wenstrup RM, Greenspan DS. Characterization of the six zebrafish clade B fibrillar procollagen genes, with evidence for evolutionarily conserved alternative splicing within the α1(V) C-propeptide. Matrix Biol. 2010; 29(4):261–275. [PubMed: 20102740]
12. Bonadio J, Holbrook KA, Gelinas RE, Jacob J, Byers PH. Altered triple helical structure of type I procollagen in lethal perinatal osteogenesis imperfecta. J Biol Chem. 1985; 260:1734–1742. [PubMed: 2981871]
13. Mitchell AL, Schwarze U, Jennings JF, Byers PH. Molecular Mechanisms of Classical Ehlers-Danlos Syndrome. Human Mutation. 2009; 30(6):995–1002. [PubMed: 19370768]
14. Wieringa B, Hofer E, Weissmann C. A minimal intron length but no specific internal sequence is required for splicing the large rabbit beta-globin intron. Cell. 1984; 37(3):915–925. [PubMed: 6204770]
15. Crovato TE, Egebjerg J. ASF/SF2 and SC35 regulate the glutamate receptor subunit 2 alternative flip/flop splicing. FEBS Letters. 2005; 579:4138–4144. [PubMed: 16023113]
16. Gallego ME, Gattoni R, Stevenin J, Marie J, Expert-Bezancon A. The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the β-tropomyosin alternative exon 6A. The EMBO Journal. 1997; 16:1772–1784. [PubMed: 9130721]
17. Chen Y, Sumiyoshi H, Oxford JT, Yoshioka H, Ramirez F, Morris NP. *Cis*-acting elements regulate alternative splicing of exons 6A, 6B, and 8 of the α1(XI) collagen gene and contribute to the regional diversification of collagen XI matrices. Matrix Biology. 2001; 20:589–599. [PubMed: 11731275]
18. Davies GB, Oxford JT, Hausafus LC, Smoody BF, Morris NP. Temporal and Spatial Expression of Alternative Splice-Forms of the α1(XI) Collagen Gene in Fetal Rat Cartilage. Dev Dyn. 1998; 213:12–26. [PubMed: 9733097]
19. Lui VC, Ng LJ, Nicholls J, Tam PP, Cheah KS. Tissue-Specific and Differential Expression of Alternatively Spliced a1(II) Collagen mRNAs in Early Human Embryos. Dev Dyn. 1995; 203:198–211. [PubMed: 7655082]

20. Oxford JT, Doege KJ, Morris NP. Alternative Exon Splicing within the Amino-terminal Nontriple-helical Domain of the Rat Pro-α1(XI) Collagen Chain Generates Multiple Forms of the mRNA Transcript Which Exhibit Tissue-dependent Variation. J.Biol Chem. 1995; 270(16):9478–9485. [PubMed: 7721875]

21. Ryan MC, Sandell LJ. Differential Expression of a Cysteine-rich Domain in the Amino-terminal Propeptide of Type II (Cartilage) Procollagen by Alternative Splicing of mRNA. J Biol Chem. 1990; 265(18):10334–10339. [PubMed: 2355003]

22. Sugimoto M, Kimura T, Tsumaki N, Matsui Y, Nakata K, Kawahata J, Yasui N, Kitamura Y, Nomura S, Ochi T. Differential in situ expression of α2(XI) collagen mRNA isoforms in the developing mouse. Cell Tissue Res. 1998; 292:325–332. [PubMed: 9560475]

23. Tsumaki N, Kimura T. Differential Expression of an Acidic Domain in the Amino-terminal Propeptide of Mouse Pro-α2(XI) Collagen by Complex Alternative Splicing. J Biol Chem. 1995; 270(5):2372–2378. [PubMed: 7836472]

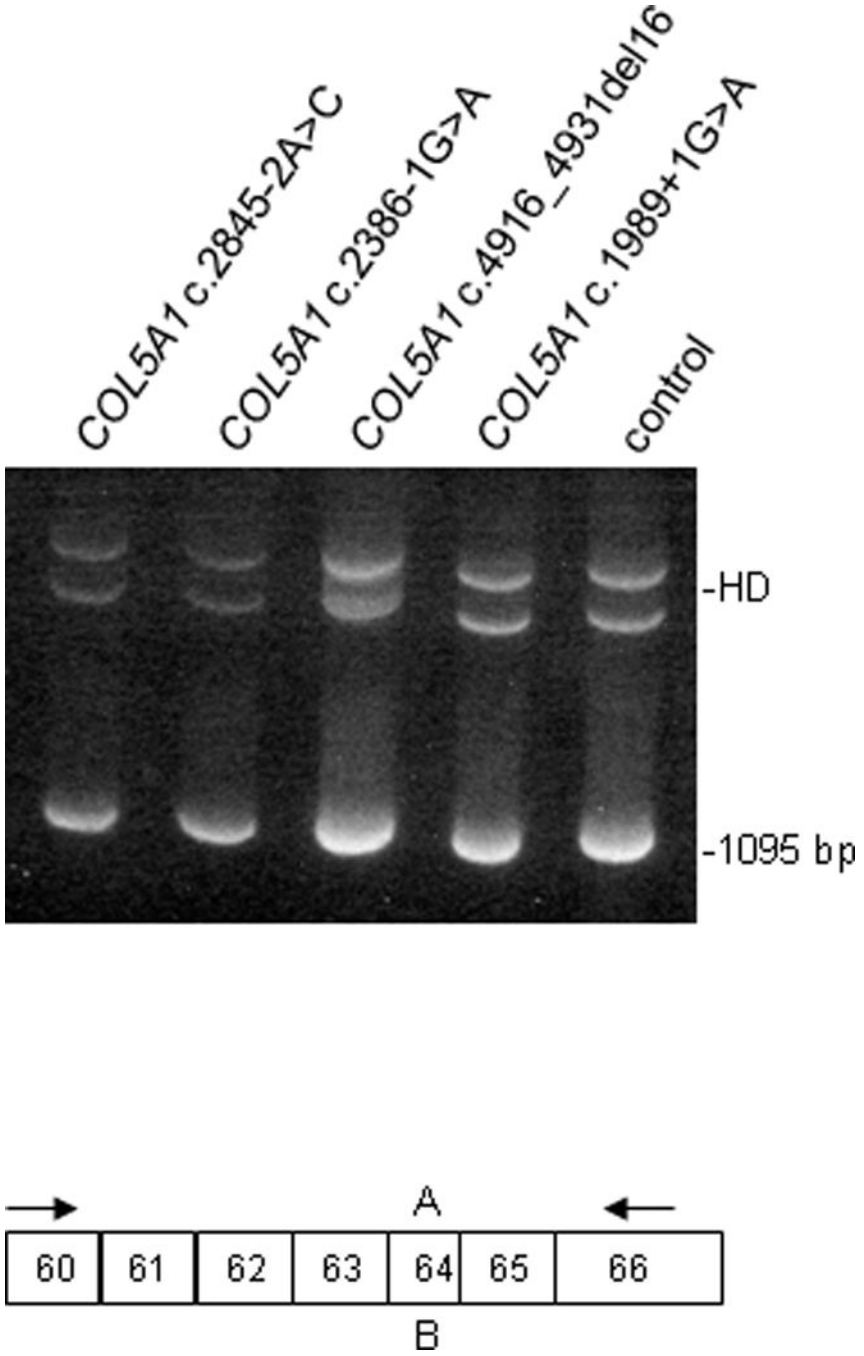



**Figure 1.**

A. Gel of PCR products from cDNA amplification of *COL5A1* with primers in exon 60 and the 3'UTR. cDNA synthesized from RNA expressed by human skin fibroblasts was used as template. The expected product is 1095 basepairs long. Lanes 1–4 are from individuals who are haploinsufficient for *COL5A1* expression and lane 5 is from a control individual. Mutations were described previously [23] and are as follows: lane 1 corresponds to P12 in previous publication; lane 2 corresponds to P10; lane 3 corresponds to P8; lane 4 corresponds to P9. A heteroduplex is present in all lanes, indicating the presence of both exon 64 isoforms. HD= heteroduplex; bp=base pairs. Arrows designate location of primers used for PCR.

B. Sequencing chromatograms of *COL5A1* exon 64 region. cDNA synthesized from RNA expressed by skin fibroblasts from individuals haploinsufficient for *COL5A1* expression was used as template. Both exon 64 sequences were detected in this assay, although previous analysis [23] was unable to detect message from the mutant allele in these samples. *= start of exon 64. e= end of exon 64.

A.

AlaArgIleThrSer**TrpProLysGlu**Asn**Pro**GlySer**Trp**Phe**Ser**GluPhe**LysArgGly**Lys**Leu**
64A-GCC**A**G**AAT**CA**C**TTCT**TGGCCCAAAGA**AAA**C**CC**GGG**C**TCCTGGT**TC**AGT**G**AA**T**TC**AAGCG**T**GGGAAA**CTG**
64B-AGT**A**A**AAT**GG**C**CCGC**TGGCCCAAAGA**GC**A**G**CC**TTC**C**A**CCTGGT**AT**AGT**C**A**G**T**A**C**AAGCG**G**GGG**TCC**CTG**
SerLysMetAlaArg**TrpProLysGlu**Gln**Pro**SerThr**Trp**Tyr**Ser**GlnTyr**LysArgGly**Ser**Leu**

B.

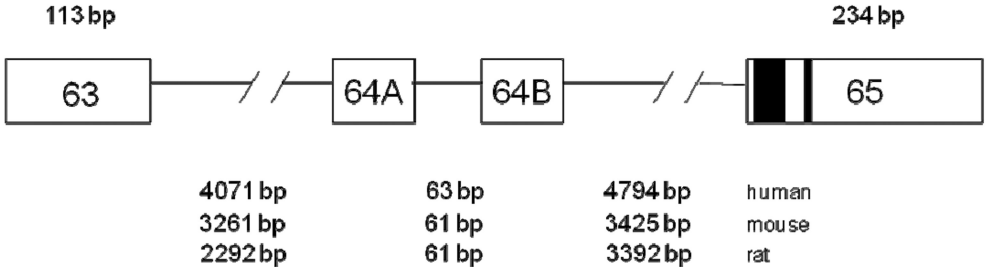



**Figure 2.**

Comparison of *COL5A1* alternatively spliced exons. A. DNA sequences of exons 64A and 64B in human. Identical basepairs are in bold. *EcoRI* site in exon 64A begins at position 49. The amino acid sequence encoded by each exon is indicated above (for exon 64A) or below (for exon 64B) the corresponding DNA sequence. Identical residues are in bold.

B. Genomic map of human and rodent *COL5A1* in the region of alternative splicing. Exons are indicated as boxes, introns as lines. Shaded regions denote the base pairs in exon 65 that encode the amino acids of the chain selectivity recognition domain. bp= base pairs.

```
                Exon64A
Human           GCCAGAATCACTTCTTGGCCCAAAGAAAACCCGGGCTCCTGGTTCAGTGAATTCAAGCGT 60
Chimp           GCCAGAATCACTTCTTGGCCCAAAGAAAACCCGGGCTCCTGGTTCAGTGAATTCAAGCGT 60
Mouse           GCCAGAATCACTTCTTGGCCCAAAGAAAACCCAGGTTCCTGGTTCAGTGAATTCAAGCGT 60
Horse           GCCAGAATCACTTCTTGGCCCAAAGAAAACCCAGGTTCCTGGTTCAGTGAATTCAAGCGT 60
Opposum         GCCAGAATTACTTCTTGGCCCAAGGAAAACCCAGGCTCCTGGTTCAGTGAATTCAAGCGT 60
Chicken         GCTAGAATTACTTCTTGGCCCAAAGGAAAACCCAGGCTCCTGGTTCAGTGAATTCAAGCGC 60
Lizard          GCTAGAATTACTTCCTGGCCCAAAGGAAAACCCTGGCTCCTGGTTCAGTGAATTCAAGCGC 60
Xenopus         GCTAGAATTACATCTTGGCCCAAGGAAAATCCAGGCTCCTGGTTCAGTGAATTCAAGCGT 60
Tetraodon       TCTAGGGTCACGTCCTGGGCAAAGGAGAGTCCTGGCTCCTGGTTCAGTGAATTCAAACGT 60
                 * **  * ** ** *** * ** ** *  ** ** ******************** **



Human           GGGAAACTGGtaaggtggcctctggcgt-ctttgcggttgtcactttaaaccc--gccca 117
Chimp           GGGAAACTGGtaaggtggcctctggcgt-ctttgcggttgtcactttaaaccc--accca 117
Mouse           GGGAAACTGGtaaggtgacaccacg-tt-ctttgcatttgtcactttaaaccc--gcc-a 115
Horse           GGTAAACTGGtaaggccgcccctggctt-ctctgggattgtcg-tttaaaccc--acccg 116
Opposum         GGTAAACTGGtaaggaagcccctcatct-ctttgcagttttggctttaaaccccaaacca 119
Chicken         GGTAAACTGGtaagatgcatctcagt---gctttcagtttt--ctttaaccc-------g 108
Lizard          GGTAAACTGGtaagaggctcttgcatttggcttgcggttggg-ctttaacct-------g 112
Xenopus         GGTAAACTGGtaagg---tccaaactttgaattccgttt----ctgtaactcc------a 107
Tetraodon       GGTAAACTGGtaagccagacctcctcta--gtcctttccctccctgaatgc------t 112
                ** ***********                          *        * *



                          Exon 64B
Human           tctcgtatcttacagAGTAAAATGGCCCGCTGGCCCAAAGAGCAGCCTTCCACCTGGTAT 177
Chimp           tctcgtatcttacagAGTAAAATGGCCCGCTGGCCCAAAGAGCAGCCTTCCAGCTGGTAT 177
Mouse           tctcctatcttgcagAGTAAAATGGCCCGGTGGCCCAAAGAACAGCCTTCCACCTGGTAT 175
Horse           tttcgtatcttccagAGTAAAATGGCCCGCTGGCCCAAAGAACAGCCTTCTACCTGGTAT 176
Opposum         tctcgtattttacagAATAAAATGGCCCGTTGGCCCAAAGAGCAGCCTTCTACCTGGTAT 179
Chicken         tgtcgtattttacagAGTAAAATGGCTCGTTGGCCCAAAGAACAGCCTTCCACATGGTAT 168
Lizard          tcccgtatttttcagAGTAAACTGGCCCGTTGGCCCAAAGAGCAGCCCGCCACGTGGTAT 172
Xenopus         tgtcatattttacagAATAAACTGGCTCGTTGGCCCAAAGAGCTGCCTGGTACATGGTAT 167
Tetraodon       tcttatattttacagAGCAAAATGGCAAAATGGCCCAAAGACTCACCGGGCACATGGTAT 172
                *    *** ** ****  *** ****    ***********    **    *  ******



Human           AGTCAGTACAAGCGGGGGTCCCTG 201
Chimp           AGTCAGTACAAGCGGGGGTCCCTG 201
Mouse           AGTCAGTACAAGCGGGGTTCCCTG 199
Horse           AGTCAGTACAAGCGAGGGTCCCTG 200
Opposum         AGTCAGTACAAGCGAGGTTCTATG 203
Chicken         AGTCAATACAAGCGGGGGTCTTTG 192
Lizard          AGTCATTTCAAACGGGGCTCCATG 196
Xenopus         AGTCATTACAAGCGGGGTTCCATG 191
Tetraodon       AGTCACTACAAGAGAGGTTCGCTG 196
                ***** * ***  * ** **  **
```

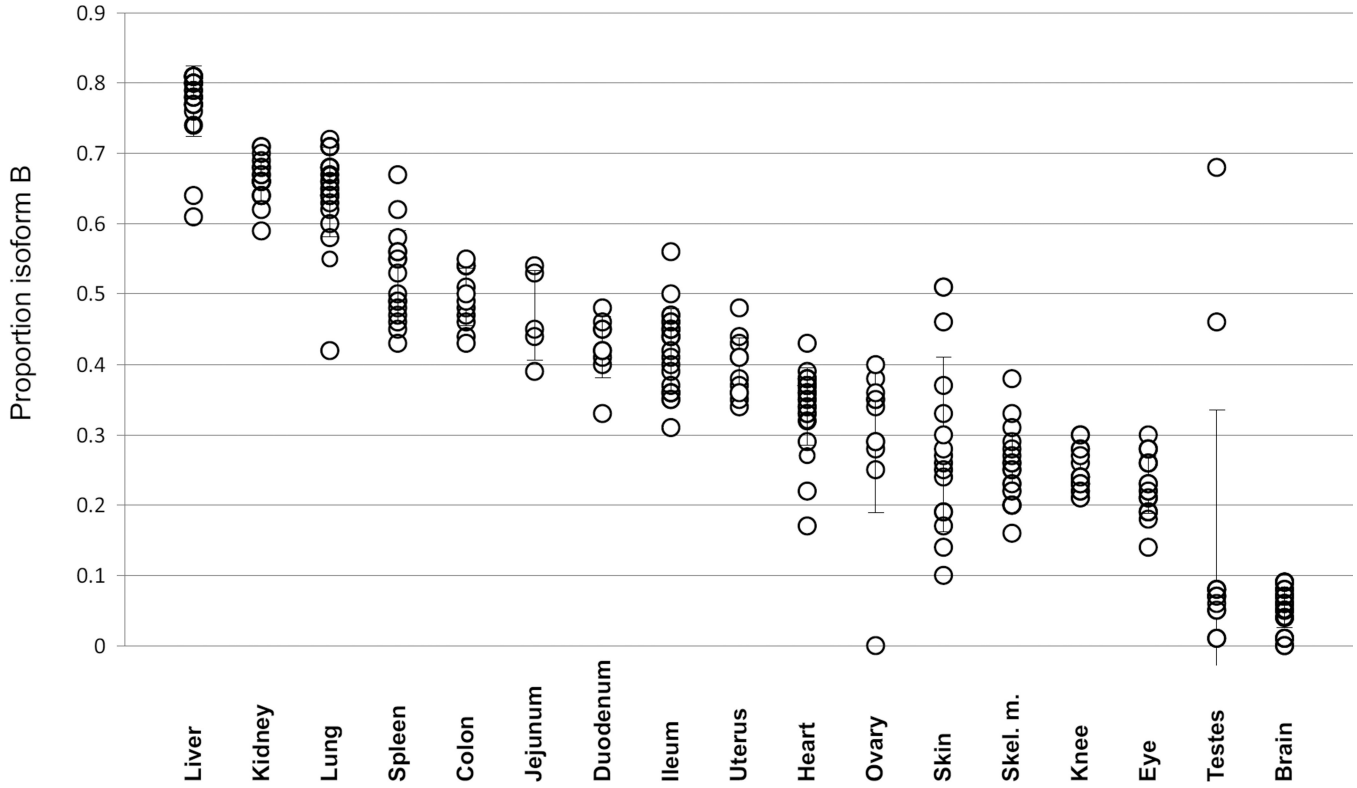
A

```
Exon 64A

Human           ARITSWPKENPGSWFSEFKRGKL 23
Chimp           ARITSWPKENPGSWFSEFKRGKL 23
Mouse           ARITSWPKENPGSWFSEFKRGKL 23
Horse           ARITSWPKENPGSWFSEFKRGKL 23
Opposum         ARITSWPKENPGSWFSEFKRGKL 23
Chicken         ARITSWPKENPGSWFSEFKRGKL 23
Lizard          ARITSWPKENPGSWFSEFKRGKL 23
Xenopus         ARITSWPKENPGSWFSEFKRGKL 23
Tetraodon       SRVTSWAKESPGSWFSEFKRGKL 23
                :*:***.**.*************

Exon64B

Human           SKMARWPKEQPSTWYSQYKRGSL 23
Chimp           SKMARWPKEQPSSWYSQYKRGSL 23
Mouse           SKMARWPKEQPSTWYSQYKRGSL 23
Horse           SKMARWPKEQPSTWYSQYKRGSL 23
Opposum         NKMARWPKEQPSTWYSQYKRGSM 23
Chicken         SKMARWPKEQPSTWYSQYKRGSL 23
Lizard          SKLARWPKEQPATWYSHFKRGSM 23
Xenopus         NKLARWPKELPGTWYSHYKRGSM 23
Tetraodon       SKMAKWPKDSPGTWYSHYKRGSL 23
                .*:*:***: *.:***::****:
```

B

**Figure 3.**

Clustal W alignment of *Col5a1* exon 64 region from nine species. A. Genomic DNA sequence alignment. Exon sequences are in upper case letters and intron 64A is in lower case. *=identical among all nine species. B. Amino acid alignment of the two exons. *= identical among all nine species; := conservative substitution; .= nonconserved change.



| tissue | mean | stdev | n |
|---|---|---|---|
| liver | 0.769 | 0.051 | 39 |
| kidney | 0.653 | 0.035 | 27 |
| lung | 0.639 | 0.067 | 39 |
| spleen | 0.543 | 0.070 | 24 |
| colon | 0.496 | 0.041 | 12 |
| jejunum | 0.470 | 0.064 | 5 |
| ileum | 0.426 | 0.064 | 36 |
| duodenum | 0.424 | 0.044 | 9 |
| uterus | 0.394 | 0.044 | 11 |
| heart | 0.359 | 0.112 | 40 |
| ovary | 0.299 | 0.109 | 11 |
| skeletal muscle | 0.288 | 0.114 | 28 |
| skin | 0.286 | 0.124 | 16 |
| knee | 0.246 | 0.029 | 18 |
| eye | 0.229 | 0.041 | 20 |
| testes | 0.126 | 0.168 | 18 |
| brain | 0.061 | 0.026 | 42 |

B

**Figure 4.**

A. Expression profile for *Col5a1* exon 64 isoforms in murine tissues. The proportion of isoform B (amount of isoform B/ total amount of *Col5a1* RNA) is graphed for the 17 tissues. Data points are from young adult mice aged 7– 15 weeks.

B. Proportion of isoform 64B in mouse tissues. The mean values for the proportion of isoform B in the 17 mouse tissues with standard deviations are indicated.

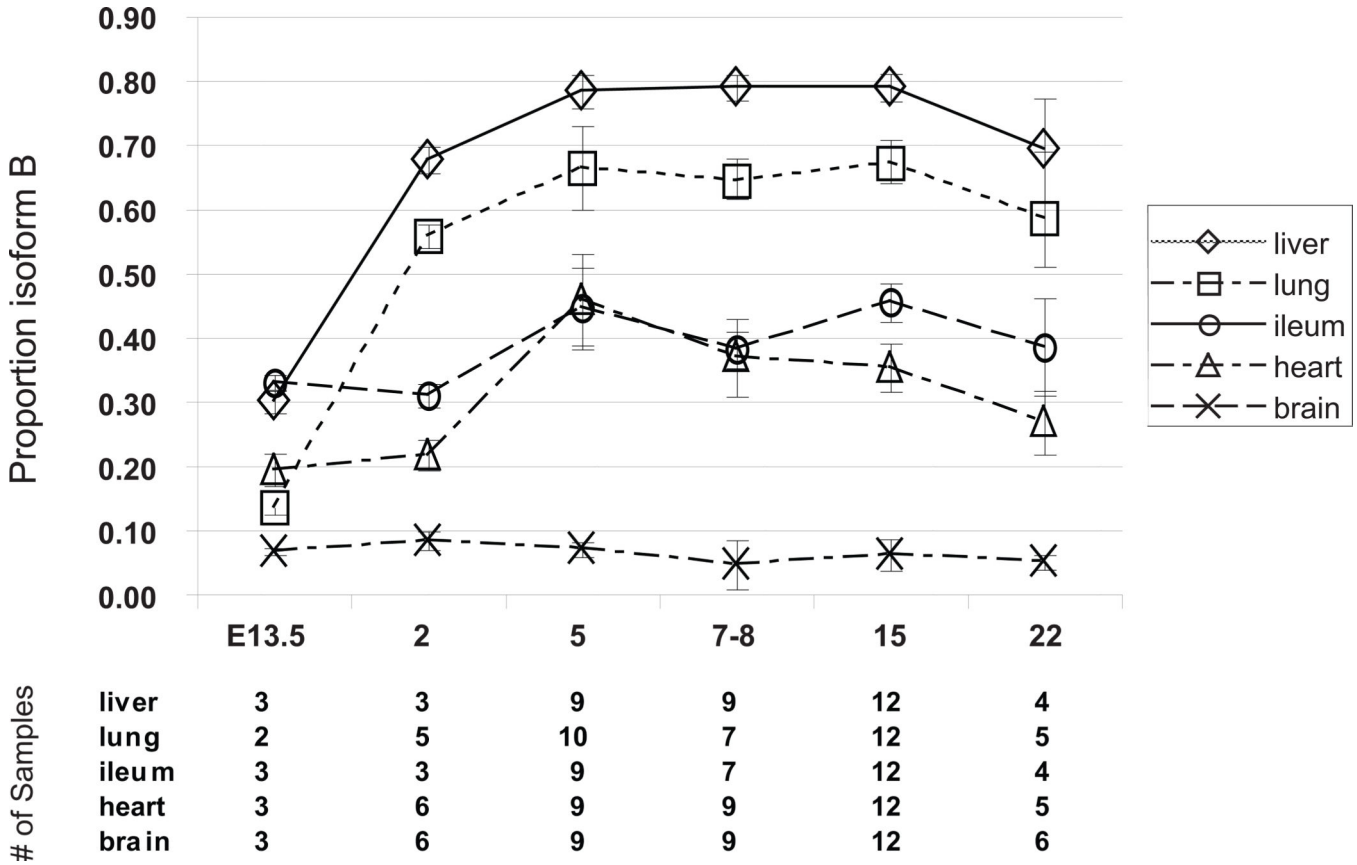



**Figure 5.**

Developmental expression profile of *Col5a1* exon 64 isoforms. The proportion of isoform B is plotted for mouse liver, lung, heart, small intestine, and brain at ages ranging from E13.5 through 22 weeks. Curve morphologies fell into three categories: 1) stable level of expression throughout all developmental time points (brain); 2) postnatal increase in proportion of B, between 2 and 5 weeks (heart and ileum); 3) dramatic early increase in isoform B usage between E13.5 and 2 weeks, with stable proportion in adult (lung and liver).

## Table 1

Intronic size and expression data for *Col5a1* exon 64 isoforms. Species with exon 64B sequence in the Ensembl or UCSC database are listed, with the size of the intervening intron (IVS64A), and the tissues in which expression of both isoforms was confirmed in our laboratory. Organisms without expression data were not studied.

| Genomic evidence for exon 64B | Intronic Size (bp) | Expression of exon 64B |
|---|---|---|
| Homo sapiens | 63 | Skin fibroblasts |
| Pan troglodytes (chimpanzee) | 63 | |
| Gorilla gorilla (gorilla) | 63 | |
| Pongo pygmaeus (orangutan) | 63 | |
| Macaca mulatta (macaque) | 63 | |
| Callithrix jacchus (marmoset) | 63 | |
| Tarsius syrichta (tarsier) | 63 | |
| Otolemur garnettii (small ear galago) | 62 | |
| Microcebus murinus (grey mouse lemur) | 62 | |
| Canis familiaris (dog) | 62 | |
| Felis catus (cat) | 62 | |
| Equus caballus (horse) | 62 | |
| Cavia porcellus (guinea pig) | 61 | |
| Dipodomys ordii (kangaroo rat) | 61 | |
| Mus musculus | 61 | Brain, colon, duodenum, eye, heart, ileum, jejunum, kidney knee, liver, lung, ovary, skeletal muscle, skin, spleen, testes, uterus |
| Rattus norvegicus | 61 | |
| Spermophilus tridecemlineatus (ground squirrel) | 63 | |
| Ochotona princeps (pika) | 61 | |
| Oryctolagus cuniculus (rabbit) | 61 | |
| Sus scrofa (pig) | 61 | |
| Tursiops truncates (bottlenose dolphin) | 61 | |
| Myotis lucifugus (microbat; little brown bat) | 54 | |
| Dasypus novemcinctus (armadillo) | 63 | |
| Loxodonta Africana (elephant) | 62 | |
| Erinaceus europaeus (hedgehog) | 62 | |
| Echinops telfairi (lesser hedgehog) | 61 | |
| Opossum | 65 | |
| Ornithorhynchus anatinus (platypus) | 65 | |
| Gallus gallus (chicken) | 54 | Embryonic brain, heart, liver, lung, small intestine |
| Meleagris gallopavo (turkey) | 54 | |
| Taeniopygia guttata (zebrafinch) | 55 | |
| Anolis carolinensis (anole lizard) | 58 | |
| Xenopus tropicalis/laevis (frog) | 53 | Brain, heart, liver, lung, small intestine |

| Genomic evidence for exon 64B | Intronic Size (bp) | Expression of exon 64B |
|---|---|---|
| Danio rerio (zebrafish) | 52 | |
| Tetraodon nigroviridis (pufferfish) | 58 | |
| Gasterosteus aculeatus (stickleback) | 60 | |
| Oryzias latipes (Japanese killifish; Medaka) | 58 | |
| Takifugu rubripes | 52 | |