**Supplemental Materials and Methods**

**Animal husbandry:** Breeding pairs of Tsk2/+ mice were obtained from Dr. Paul Christner at Jefferson University and housed at Drexel University College of Medicine. Tsk2/+ mice were serially backcrossed the C57Bl/6J (B6) background. Recombinant B6.Tsk2/+ mice were bred to B6.chr 1-A/J mice (Jackson Laboratory, Bar Harbor, ME) and the resulting B6.Tsk2/+ F1 mice were backcrossed to B6.chr 1-A/J mice. Wild-type littermates were used as controls.

**DNA isolation:** Tail snips were taken from weanlings, and DNA was isolated from the tissue using a GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO), following the manufacturer's protocols. DNA was diluted to a working concentration of 40 ng/μl.

**Microsatellite and SNP typing:** 120 ng of DNA per reaction was used when amplifying microsatellites by polymerase chain reaction (PCR). PCR products were separated by electrophoresis on a 3% agarose gel. .100 ng of DNA per reaction was used for SNP typing, as previously described (Bunce *et al.*, 1995), to map the boundaries of the recombinations. Microsatellite *D1Mit233*, sequence forward 5'-TAGACCCATCACTTTCCAAG-3' and reverse 5'-ACTGGCTAAAGTA TCCTAGAAAGGG-3' was run at an annealing temperature of 49 $^{o}$C. Microsatellites *D1Mit235* sequence forward 5'-CACCTGGCTAAGAGACCATACC-3' and reverse 5'-GCCTCCACTACCACCATCTC-3'; a microsatellite in the *Glutaminase* gene (*Gls*) sequence forward 5'-TGTGCACTT GAGAATTTTGCTT-3' and reverse 5'-CCCACATACTGGACCTACCC-3'; and *D1Mit18* sequence forward 5'-TCTGGTTCCAGGCTTGATTC-3'and reverse 5'-TCACAAGTGA GGCTCCAGG-3' were run at an annealing temperature of 50 $^{o}$C.

**SNP typing:** Specific locations of polymorphisms between B6 (which is very similar to 101/H) and A/J were determined using Mouse Genome Informatics (www.informatics.jax.org), and primers for SNP typing were designed using Primer 3 online software (http://frodo.wi.mit.edu) and synthesized by Integrated DNA Technologies (www.idtdna.com). PCR products were separated by electrophoresis on a 1.5% agarose gel.

**Complementation analysis with Col3a1$^{-/+}$ mice:** Col3a1$^{-/+}$ mice (bearing one allele containing *Col3a1* and one allele where *Col3a1* has been knocked out) were received as a generous gift from Dr. Xianhua Piao at Harvard Medical School. Tsk2/+ mice were crossed to Col3a1$^{-/+}$ mice to verify that the SNP in *Col3a1* is *Tsk2*, as Tsk2/Tsk2 homozygous mice are not viable, and if *Tsk2* is located in *Col3a1*, then *Tsk2/Col3a1*$^{-}$ mice will not be viable. The resulting first generation of the cross was genotyped by PCR for Tsk2/+ using the microsatellites listed above and primers specific to *Col3a1* or the inserted neomycin cassette (see supplemental material). ***In vitro* assessment of fibrogenesis by *Col3a1*$^{Tsk2}$:** We constructed a plasmid harboring the Col3a1$^{Tsk2}$ allele by introducing the Tsk2 T→ A mutation into a wild-type Col3a1 clone (pCMV6-Kan/Neo; Origene). Skin explants of newborn mice from [Col3a1KO/+ x Col3a1KO/+] litters were harvested and cultured *in vitro* and genotyped for KO/KO homozygosity One *Col3a1*-KO line was transfected with plasmid containing 5 μg of either the *Col3a1*$^{Tsk2}$ or *Col3a1*$^{WT}$ gene, using the calcium-phosphate co-precipitation method for 3 h and then treated with 10% glycerol/PBS for 2 min, washed and cultured for 48 h as previously described (Artlett *et al.*, 1998). Supernatants were retained and cell lysates were harvested directly from the dish at 48 hours.

**RNA isolation and real-time PCR**. RNA was isolated from fibroblasts using a RNA isolation

kit from Clontech (Mountain View, CA) following manufacturer's instructions. cDNAs were synthesized from 2.0 µg of total RNA using an High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA). Primers for quantitative real-time PCR (qRT-PCR) were designed and synthesized as described for SNP typing. Relative quantification of all products was measured using SYBR Green chemistry (Applied Biosystems, Foster City, CA). Expression was normalized to the geometric mean of the expression of house-keeping genes actin and β2-microglobulin, and relative expression of each gene was calculated using the ΔCt formula. The fold increase or decrease in *Col3a1*$^{Tsk2}$ or *Col3a1*$^{WT}$ transfected cells was calculated as a ratio over the expression in mock transfected controls (ΔΔCt). For primer sequences, see supplemental material.

**RNA-Seq – Sample RNA preparation:** Total RNA was prepared from three WT and four Tsk2/+ mice skin biopsies using Qiagen RNeasy Fibrous Tissue Mini Kit. The RNA integrity (RIN) was determined by Agilent Bioanalyzer Nano chip. All samples used for this study had RIN scores of 7 or greater (Erik Garrison, 2012).

**RNA-Seq:** RNA-seq sequencing libraries were prepared for the seven samples using NuGEN Ovation RNA-Seq System (NuGen, San Carlos, CA). Libraries were multiplexed and sequenced on an Illumina HiSeq 2000 platform to obtain 16.7-50.9 million 50 bp paired-end reads per sample. The raw reads were aligned to the reference mouse genome (MM9 assembly) using Tophat software with default parameters (Trapnell *et al.*, 2012a; Trapnell *et al.*, 2012b). Supplemental Figure 1 shows RNA-Seq read coverage for three interval genes.

**Variant calling and SNP identification:** In order to identify coding region and intronic genetic variation we analyzed the data using Freebayes software (Erik Garrison, 2012). The

following parameters were used: “freebayes –r 1:44200000..47100000 –f MM9.fai –b <tophat_aligned_bamfile> -v <output_SNP.vcf>” We compared the WT and Tsk2/+ mice to the MM9 reference genome strain (B6).

**454 Sequencing:** Samples were captured and amplified as described in the Roche Nimblegen sequence capture manual (Version 1.0). Based on qPCR analysis, control capture regions showed an enrichment range of 14 to 50 fold for the samples tested. Titanium general libraries were prepared from the captured DNAs from two 101/H mice and two Tsk2/+ mice using 5000 ng of DNA as described in the GS FLX Titanium, General Library Preparation Method Manual, October, 2008 (Roche Molecular Systems, Nutley, NJ). Enriched captured fragments binding to beads, titration, emulsion PCR, emulsion breaking, bead enrichment, and pico-titer plate-based pyrosequencing were performed as described in GS FLX Titanium emPCR and Sequencing Protocols, October, 2008. Sequence capture array probes were designed by Roche Nimblegen using the mouse genome sequence between 44,241,286 and 47,116,890 on chromosome 1 of mouse genome (mm9). Probes were designed corresponding to 56.3% of the linkage region, however in practice a larger area was captured due to the overhang of the larger fragments (~600bp) being used. Probes could not be designed to the remaining 43.7% due to it being composed of repetitive sequence.

Multiplexed 454 sequenced reads were assembled using Newbler v2.6 with scaffolding against the same chromosome region that the probes were derived from. Separate assemblies were created for each of the four mice by MID number, and lists of variants for each mouse were obtained from the assembler output. Variants were filtered by quality (phred scores >30), depth of coverage (>13 reads), and heterozygosity (>20% of reads differed from the reference).

Variants were mapped to exons, introns, and intergenic regions within the linkage region and set analysis between the 204 and 101 lineages were performed using custom perl scripts. Sets were examined for variants between the 204 and 101 line, and between all samples and the reference to identify heterozygous SNPs uniquely present in the tsk2 line.

**Complementation analysis with Col3a1$^{-/+}$ mice:** WT forward (common) primer 5'-CTTCTCACCCTTCTTCATCCC-3', WT reverse primer 5'-AGCCTGTTCAATCGGTACC-3', and neomycin reverse primer 5'- GCTATCAGGACATAGCGTTGG-3'. A second primer set was used to verify the knock out, WT forward (common) primer 5'-AGGGCCTTCAGAGGATTTTC-3', WT reverse primer 5'-CCATCCCCTCAGCAGTAAA-3', and the neomycin reverse 5'-GFCCAGAGGCCACTTGTGTAG-3'. Reactions were run at an annealing temperature of 63 $^{o}$C and PCR products were separated by electrophoresis on a 2% agarose gel.

**Real-time PCR**. For primer sequences, the following were used: Col1a1 F: ACTGGTACATCAGCCCGAAC; COL1A1 R: CTACGCTGTTCTTGCAGTGATAG; COL3A1 F: CTGCTCGGAACTGCAGAGAC; COL3A1 R: CCACCAGTGCTTACGT; ACTIN F: CAGCTTCTTTGCAGCTCCTT; ACTIN R:CACGATGGAGGGGAATACAG; B2MG F: TCGCTCGGTGACCCTAGTCTTT; and B2MG R: ATGTTCGGCTTCCCATTCTCC. The fold increase or decrease in *Col3a1*$^{Tsk2}$ or *Col3a1*$^{WT}$ transfected cells was calculated as a ratio over the expression in mock transfected controls (ΔΔCt). Transfection efficiency of plasmids into the *Col3a1*-KO fibroblasts was determined from samples of plasmid-derived DNA taken from the cell lysates (5 µl) amplified with the Col3a1 primers.

**DNA microarray hybridization and data analysis:** RNA was isolated as above and used to create cDNA for the microarray analyses. Samples were amplified and labeled using the

Agilent Low Input Linear Amplification kit (Agilent Technologies, Santa Clara, CA) and were hybridized against Universal Mouse Reference (UMR) (Strategene, La Jolla, CA) to Agilent Whole Mouse Genome arrays (G4122F) (Agilent Technologies, Santa Clara, CA) in a common reference based design. Microarrays were hybridized and washed in accordance with manufacturer's protocols and scanned using a dual laser GenePix 4000B scanner (Axon Instruments, Foster City, CA). The pixel intensities of the acquired images were then quantified using GenePix Pro 5.1 software (Axon Instruments, Foster City, CA). All microarrays were visually inspected for defects or technical artifacts, and poor quality spots were manually flagged and excluded from further analysis.

The data were loaded to the UNC Microarray Database (UMD). Raw data is available from NCBI GEO at accession number GSEXXXX (submission in process).

Pre-processing based on GenePattern [1] modules were run with default parameters unless noted otherwise. Non-centered expression data for relevant samples were pulled down. Missing values were imputed via ImputeMissingValuesKNN module with k=5. Expression values for probes were collapsed to unique gene symbols via CollapseDataset module using Agilent mouse genome annotation file from GSEA FTP site. Expression data were median-centered by genes in Cluster 3.0 [2] and used to create a class label file to define phenotype classes (e.g. Tsk2 vs. WT) via ClsFileCreator module.

Pre-processed expression data were used to identify differentially expressed functional terms between 2 classes of samples. This was done via Gene Set Enrichment Analysis (GSEA) [3, 4] module using permutations by gene set. GSEA was run vs. the entire Gene Ontology (GO) [5] database of gene sets. In order to visualize differentially expressed GO gene sets on a single

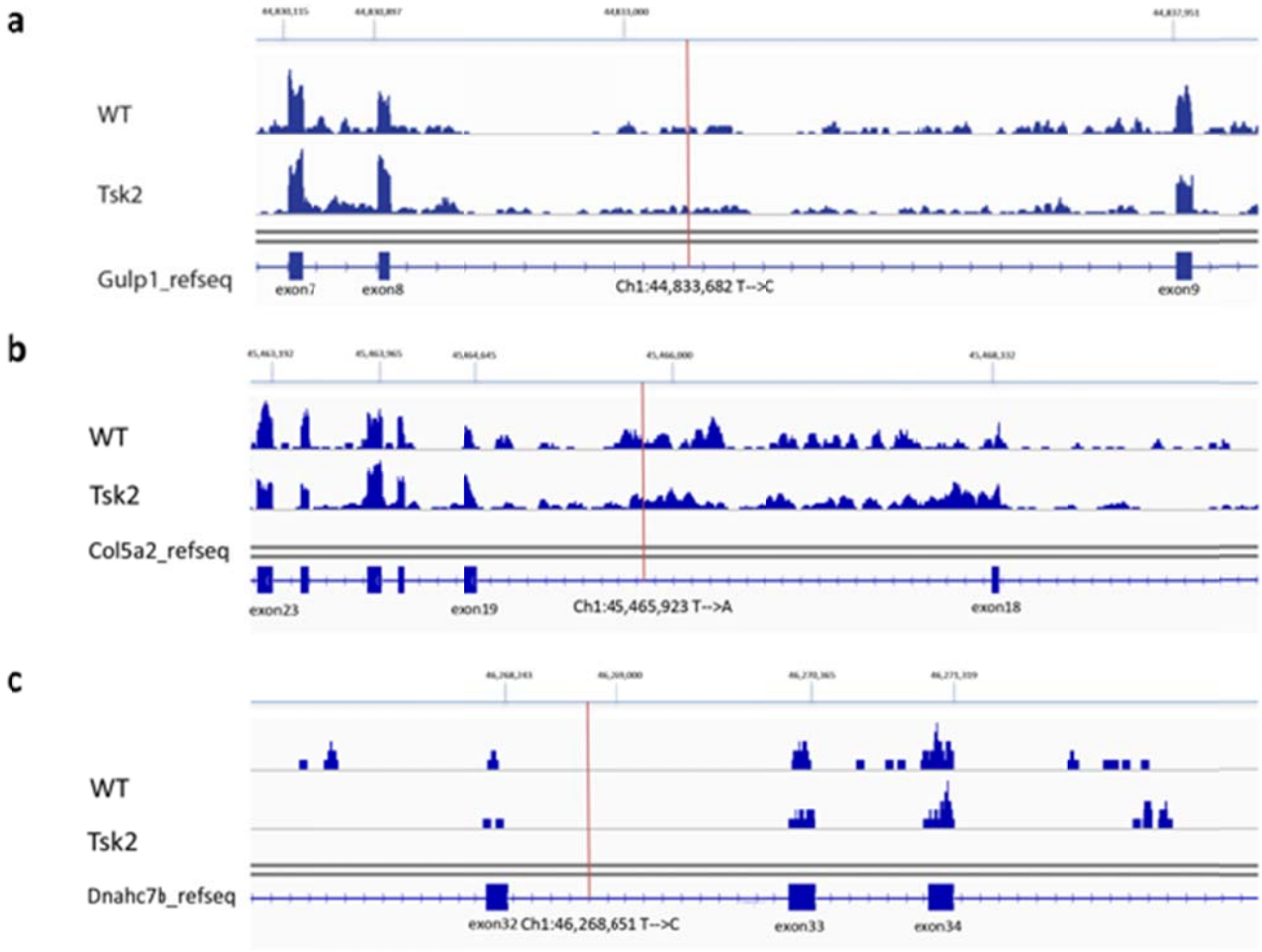
sample basis, single sample GSEA (ssGSEA) [6] was run via ssGSEAProjection module vs. GO database of gene sets. Raw ssGSEA enrichment scores were normalized by dividing by the absolute maximum ssGSEA enrichment score and median-centering rows in Cluster 3.0 thus generating ssGSEA normalized enrichment scores (NES). ssGSEA NES for significant GO gene sets (GSEA FDR<5% between 2 classes of samples) were pulled down and visualized in TreeView [7].

**Western blot analysis of *in vitro* COL3A1 and COL1A1 expression and *in vivo* COL3A1 expression:** Collagen content was determined by western blot analysis. *In vitro:* culture supernatant was collected. *In vivo*: skin was homogenized in RIPA buffer (Sigma-Aldrich, St Louis MO) using a glass homogenizer and centrifuged at 8,000 x g for 10 minutes at 4 $^{o}$C to pellet debris. Total protein was measured with a Bradford assay (Sigma-Aldrich, St Louis MO). Approximately 20 ul of culture supernatant or 75 µg of protein from skin lysate was added to reducing buffer, boiled, and then loaded onto an 8% SDS gel. After separation, proteins were transferred to a polyvinylidene fluoride membrane. The membrane was blocked in 5% nonfat milk in Tris buffered saline and then probed with goat anti-COL3A1 (#sc-8781) or goat anti-COL1A1 (#sc-28657) from Santa Cruz Biotechnology, Inc, Santa Cruz, CA, or rabbit anti-β-Actin (#4967, Cell Signaling Technologies, Boston, MA) and then probed with a secondary antibody, donkey anti-goat (#705-035-003, Jackson ImmunoResearch Laboratories, West Grove, PA) or goat anti-rabbit (#111-035-003, Jackson ImmunoResearch) respectively, and developed using SuperSignal West Dura ECL reagent (Thermo Scientific Inc, Rockford, IL). Band intensities were measured using ImageQuant TL Software (GE Healthcare Life Sciences).

**Reticular fiber staining:** Reticular fibers were stained using the Chandler's Precision

Reticular Fiber Stain kit (American Master*Tech, Lodi CA) according to the manufacturer's protocol.

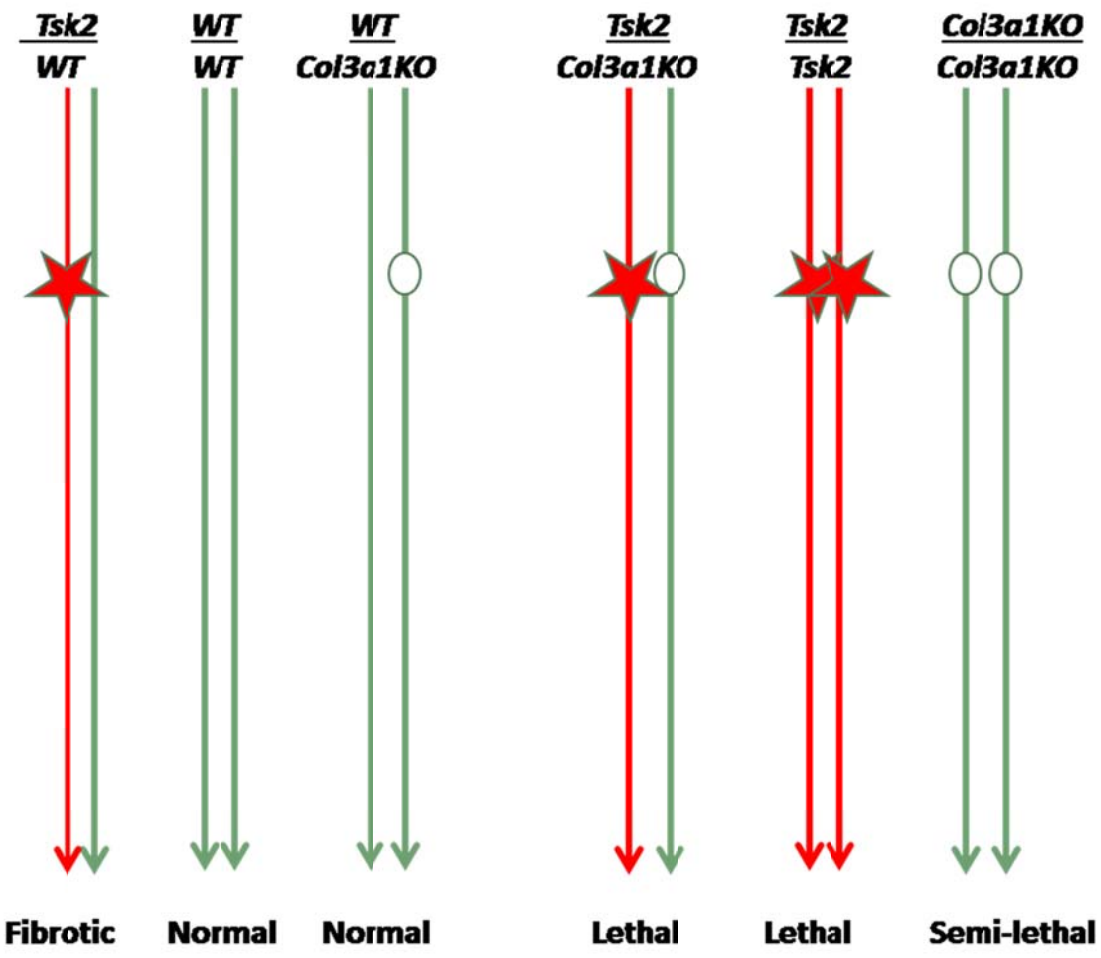
**Statistics:** A two-tailed student's t-test or a one-way ANOVA was used to determine statistical significance of collagen protein expression, as noted.




**Supplemental Figure 1.** RNA-Seq read coverage for *Gulp1*, *Col5a2* and *Dnahc7b* exons flanking the intronic *Tsk2*-specific SNPs.

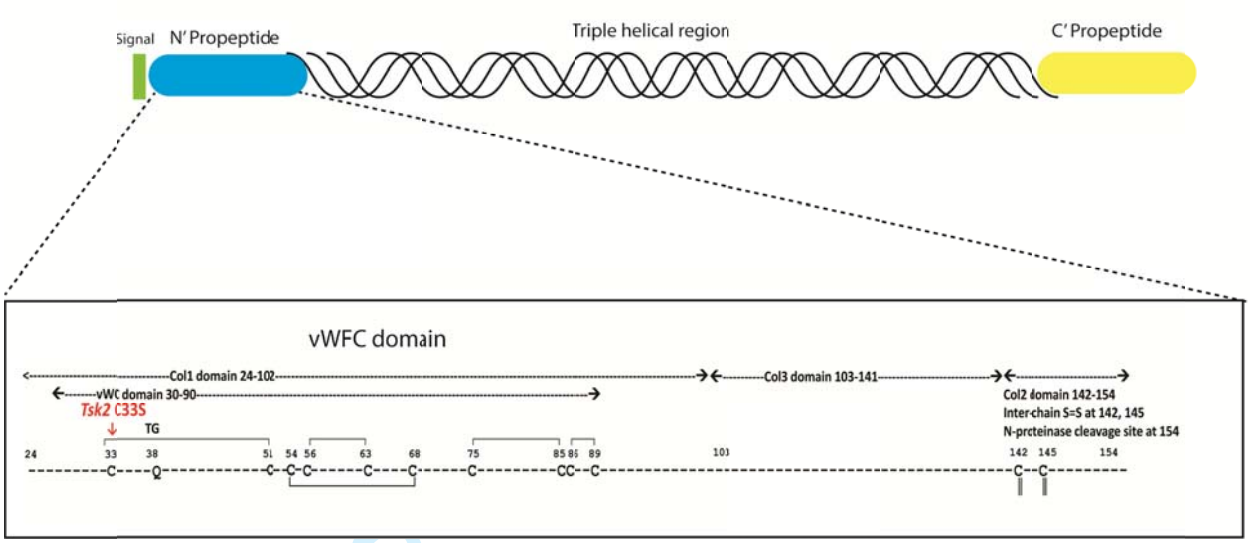
Exonic coverage is similar between WT and Tsk2/+ animals and CuffDiff analysis for differentially spliced messages did not detect differential splicing. The coverage graphs are

generated by BEDTools and visualized in Integrative Genomics Viewer. Numbers are the chromosome 1 location in nucleotides. The red bars indicate the locations of the SNPs. (**A**) Reads coverage near the *Gulp1* SNP. (**B**) Reads coverage near the *Col5a2* SNP. (**C**) Reads coverage near the *Dnahc7b* SNP.




**Supplementary Figure 2**. Complementation analysis for *Col3a1* and *Tsk2*.

The six possible genotypes generated by intercrossing (Tsk2/+ x *Col3a1*-KO/+) F1 mice are shown, with the three definitive genotypes shown on the right. If the *Tsk2* gene is identical to the *Col3a1* gene, no viable progeny will result bearing both *Tsk2* and the null allele of *Col3a1*, and this is what we observed ("lethal"). Most of the viable progeny bearing the null allele of *Col3a1* will be WT at the *Tsk2/Col3a1* locus, because approximately 90% of homozygous *Col3a1*-KO mice die before birth ("semi-lethal"), and the *Tsk2* chromosome does not rescue ("complement") the *Col3a1*-KO. The *Col3a1*-KO is indicated by the "hole" in the chromosome; the *Tsk2* gene by a star.




**Supplementary Figure 3**. Schematic representation of the mouse COL3A1 protein, with an expanded view of PIIINP.

A red arrow indicates the mutation in Cys33 in Tsk2/+ mice. The signal peptide extends from residue 1 to 23, and the Col I domain begins at amino acid 24 and extends to amino acid 102. A von Willebrand factor type c domain (VWC) is predicted between amino acids 30 and 90; the predicted transglutamination site, based on homology with human, cow, and pig Col3A1 molecules sites (Bowness *et al.*, 1987), is shown at amino acid 38 (TG). Intrachain disulphide bonds (C:C) are depicted, based on homology with other Type 3 collagens (Bruckner *et al.*, 1978). The Col3 and Col2 domains at the carboxyl terminal of the PIIINP peptide are indicated at amino acids 103-141 and 142-154, respectively.