Bifunctional promoter of type IV collagen COL4A5 and COL4A6 genes regulates the expression of α5 and α6 chains in a distinct cell-specific fashion

Malin SUND, Yohei MAESHIMA and Raghu KALLURI[1]

Center for Matrix Biology, Department of Medicine, DANA 514, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02115, U.S.A.

Type IV collagen is present ubiquitously in basement membranes. A bifunctional promoter regulates the expression of the $\alpha1/\alpha2$ genes, and the $\alpha3/\alpha4$ and the $\alpha5/\alpha6$ genes are also considered to be regulated by putative bifunctional promoters. Unlike the other type IV collagen chains, the $\alpha5$(IV) and $\alpha6$(IV) chains do not always co-localize and are present in distinct basement membranes. To address such dichotomy in the $\alpha5$(IV) and $\alpha6$(IV) gene regulation, we cloned a mouse genomic DNA fragment containing the promoter region between the two transcription start sites of these genes and we then placed this putative promoter sequence between the chloramphenicol acetyltransferase and Luciferase reporter genes, so that these genes would be transcribed in opposite directions in this unique construct. Glomerular endothelial cells and mesangial cells generate the kidney glomerular basement membrane, which always contains the $\alpha5$(IV) chain but not the $\alpha6$(IV) chain. In contrast, the basement membranes of Bowman's capsule and distal tubuli (produced by the tubular epithelial cells) contain the $\alpha6$(IV) chain. We demonstrate that, in response to TGF-β (transforming growth factor β), epidermal growth factor, vascular endothelial growth factor and platelet-derived growth factor, expression from the $\alpha5$(IV) gene is significantly enhanced in the glomerular endothelial cells and mesangial cells, but not expression from the $\alpha6$(IV) gene. In contrast, the expression from the $\alpha6$(IV) gene, and not that from the $\alpha5$(IV) gene, was significantly enhanced in response to growth factors in the tubular epithelial cells. Our results demonstrate that the proximal bifunctional promoter regulates the expression of the $\alpha5$(IV) and $\alpha6$(IV) genes in a cell-specific manner and offers the first demonstration of the promoter plasticity in growth factor regulation of type IV collagen genes in different tissues of the body.

Key words: basement membrane, COL4A5, COL4A6, gene structure, promoter analysis, type IV collagen chain.

INTRODUCTION

Type IV collagen is the most abundant protein component of all basement membranes and is involved in tasks ranging from ensuring structural integrity in tissues to modulating cell differentiation, cell growth and adhesion [1]. The triple-helical type IV collagen molecules are coded by six distinct genes, namely the $\alpha1$(IV) to $\alpha6$(IV) chains [1–10], and these can be classified into three pairs, i.e. COL4A1/COL4A2, COL4A3/COL4A4 and COL4A5/COL4A6. All these pairs have a highly unusual head-to-head orientation, sharing a bifunctional promoter [10–14]. How these bifunctional promoters are regulated is largely unknown, but transcription is considered to initiate from opposite DNA strands driven by the bidirectional promoters between the genes and influenced by elements within the genes [11,15–20].

The $\alpha1$(IV) and $\alpha2$(IV) chains are ubiquitously distributed and always expressed in the same locations, as are the $\alpha3$(IV) and $\alpha4$(IV) chains that display more specified expression patterns [21,22]. The $\alpha5$(IV) and $\alpha6$(IV) chains are also co-expressed in some tissues such as the skin and the smooth muscles, but display interesting differences in their locations at the kidney, where the $\alpha6$(IV) chain is not found in the GBM (glomerular basement membrane), but the $\alpha5$(IV) chain is always present in the GBM [23,24]. Interestingly, the trimer formed by $\alpha5$(IV) and $\alpha6$(IV) chains was found to be expressed in many tubular organs, raising the possibility that this trimer composition might confer extra support to basement membranes in organs that need to expand [24]. The trimer formed by the $\alpha3$(IV), $\alpha4$(IV) and $\alpha5$(IV) chains, on the other hand, is expressed in locations important for filtration such as the kidney glomeruli, lungs and the cerebral ventricles [24].

The physiological importance of correct expression of the COL4A5 and COL4A6 genes is evident from the Alport syndrome. This disease is caused by mutations of the $\alpha5$(IV) collagen chain and is characterized by progressive hereditary renal disease with haematuria and progression to end-stage renal failure in males due to structural defects in the GBM, as well as sensorineural hearing loss and ocular lesions [25–28]. In some families, the Alport syndrome co-segregates with diffuse leiomyomatosis, a benign tumour diathesis, and these patients were shown to harbour deletions encompassing parts of both COL4A5 and COL4A6 genes [8,25,29].

In this paper, we describe the characterization of the gene structure of mouse COL4A5 and COL4A6 genes, the bidirectional promoter, and identify several potential elements, which may be important for transcriptional regulation of these genes. By using a unique bidirectional promoter reporter construct, we show that the proximal promoter of the mouse COL4A5 and COL4A6 genes is sufficient to ensure cell-type-specific expression in cell lines derived from different compartments of the kidney in response to

stimulation with growth factors that have previously been shown to affect the expression of type IV collagen [30,31].

EXPERIMENTAL

Isolation and analysis of genomic DNA

A BAC (bacterial artificial chromosome) library containing the mouse COL4A5 and COL4A6 promoter region was obtained from Research Genetics (Huntsville, AL, U.S.A.). A clone containing the murine COL4A5/COL4A6 locus was isolated by screening a mouse BAC library using a 1.8 kb probe encompassing the start region of both COL4A5 and COL4A6 genes, generated by PCR. The sequences of the primers used to amplify the probe are as follows: 5′-CCAGGGTGCATGCTTGCGGCTCC-3′ (forward primer) and 5′-ACACAGGGCCAGTAAGAACC-3′ (reverse primer), and they were designed according to the published human sequences of COL4A5 and COL4A6 exon 1 [10,18]. By using the 1.8 kb fragment as a probe in Southern-blot analysis (M3 probe), a 7.3 kb BamHI fragment (C1) was isolated and subcloned into pGEM-7Z vector (Promega, Madison, WI, U.S.A.). The C1 clone was used for sequencing analysis of the promoter region of the murine COL4A5/COL4A6 genes by the fluorescence-labelled dye-terminator method using an Applied Biosystems 373A automatic sequencer. The mouse COL4A5/COL4A6 promoter sequence has been deposited in GenBank® database under the accession number AF141871.

Characterization of the exon–intron structure of the murine COL4A5 and COL4A6 genes was performed by analysing publicly available genomic sequences of the mouse X chromosome on the Ensembl Project database and software system (www.ensembl.org) maintained by the EMBL–EBI and the Sanger Institute (Hinxton, Cambridge, U.K.). The genomic sequences were compared with the previously published cDNA sequences for the mouse type IV collagen $\alpha 5$(IV) and $\alpha 6$(IV) chains [22,24,32].

The sequences were analysed using the Mac Vector 6.0 (Oxford Molecular Group), BLAST programs (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/) and GCG (Oxford Molecular Group; http://www.gcg.com).

5′-RACE (5′-rapid amplification of cDNA ends)

To determine the transcription initiation sites for the mouse COL4A5 and COL4A6 genes, a 5′-RACE was performed. Total RNA was isolated from the kidneys of 6-week-old C57Bl/6J mice, as described previously [33]. Total RNA (50 μg) was used for reverse transcriptase reactions with the Superscript II RT enzyme (Invitrogen, Carlsbad, CA, U.S.A.). The following gene-specific primers were end-labelled with [γ-^{32}P]ATP (PerkinElmer Life and Analytical Sciences, Boston, MA, U.S.A.) and used in the primer extension reaction: MCOL4A5E1, 5′-AGCCAGGCACA-CTCCACG-3′ and MCOL4A6E1, 5′-AGAAACTGTCCTACCA-AC-3′. The cycling conditions were: 30 cycles of 95 °C for 30 s, 50 °C for 30 s and 70 °C for 60 s. The resulting PCR products were separated on a 6 % polyacrylamide gel and sequenced.

Generating the bifunctional promoter construct

A 1829 bp genomic fragment containing the full-length mouse COL4A5/A6 bidirectional promoter, 224 bp of exon 1 from the COL4A6 gene and 121 bp of exon 1 from the COL4A5 gene was generated by PCR from the C1 clone by using the following primers 5′-CTGCTCCGAAGCTTGGTCCCAGGAGAC-3′ and 5′-CAGACTCGAGTGACATAATCTTGAAGG-3′. The underlined sequences denote the HindIII and XhoI restriction enzyme sites that were included for subsequent cloning purposes. A 700 bp fragment containing the CAT (chloramphenicol acetyltransferase) gene was generated by PCR using the pCAT3 (Promega) plasmid as a template with the following primers: 5′-GCACCCGGGAT-GGAGAAAAAAATCACTGGATATACC-3′ and 5′-CGTGGT-ACCGCCCCGCCCTGCCACTCATCGCAG-3′, and introducing the underlined XmaI and KpnI restriction enzyme digestion sites for subsequent cloning. The generated PCR fragments were subsequently subcloned into the pGL®3 Luciferase Reporter vector (Promega). The CAT-$\alpha 5$(IV)/$\alpha 6$(IV)-Luciferase reporter construct was designed to have CAT gene expression under the control of the COL4A5 promoter (upper strand) and Luciferase gene expression under the COL4A6 promoter (lower strand). Thus the CAT and Luciferase genes in the construct are transcribed from opposite strands and are inverted in relation to each other (Figure 1A). A negative control (CAT-no promoter-Luciferase) containing only the CAT and Luciferase genes in opposite directions but lacking the promoter element was used to ensure the specificity of the obtained results. Both constructs were sequenced to verify correct cloning results.

Cell lines and transfection

All media and reagents used for cell culture were purchased from Gibco BRL (Gaithersburg, MD, U.S.A.) unless otherwise indicated. The source and establishment of the rat GEC (glomerular endothelial cells), mouse MC (mesangial cells) and mouse distal TEC (tubular epithelial cells) have been described previously [34–37]. These cell lines were used to study the regulation of the bidirectional COL4A5/A6 promoter. For GEC and MC, the cells were used between passage 5 and 8, whereas for TEC, the passage number was higher than 30. All cells were initially grown in Dulbecco's modified Eagle's medium, supplemented with 10 % (v/v) FCS (fetal calf serum), 100 units/ml penicillin, 100 units/ml streptomycin and 250 ng/ml Fungizone (BioWhittaker, Walkersville, MD, U.S.A.). For transfection, 0.08×10^6 cells were plated on six-well plates and grown to 90 % confluence. At this stage, 1.5 μg of the plasmid CAT-$\alpha 5$(IV)/$\alpha 6$(IV)-Luciferase was transfected using the Lipofectamine® 2000™ system (Invitrogen) according to the manufacturer's instructions. The plasmids pCAT®3-Control (Promega), pRL®-SV40 (Promega) and the reporter construct lacking a promoter sequence were transfected and used as positive controls for the CAT and Luciferase activities and as negative control for promoter activity respectively.

After transfection, cells were serum-starved for 24 h in K1 medium (50 % Dulbecco's modified Eagle's medium and 50 % Ham's F12) and then replaced with either a medium containing 10 % FCS (CAT and Luciferase controls, promoter control and untransfected cell lysate control) or K1 medium supplemented with only the analysed cytokine. The growth factors were used at the following concentrations: 100 ng/ml, recombinant human EGF (epidermal growth factor); 3 ng/ml, TGF-β (transforming growth factor β); 10 ng/ml, PDGF (platelet-derived growth factor); and 10 ng/ml, mouse VEGF (vascular endothelial growth factor; R&D Systems, Minneapolis, MN, U.S.A.). Cells were lysed 72 h after transfection and then assayed for reporter gene activity.

Luciferase and CAT assays

Luciferase activity was assayed using the Dual-Luciferase® Reporter Assay system (Promega) according to the manufacturer's instructions. CAT activity was measured with the CAT Enzyme Assay System (Promega) according to the manufacturer's instructions and using liquid-scintillation counting. The increase in expression of the CAT and Luciferase genes in response to the analysed growth factors were compared with the cells transfected
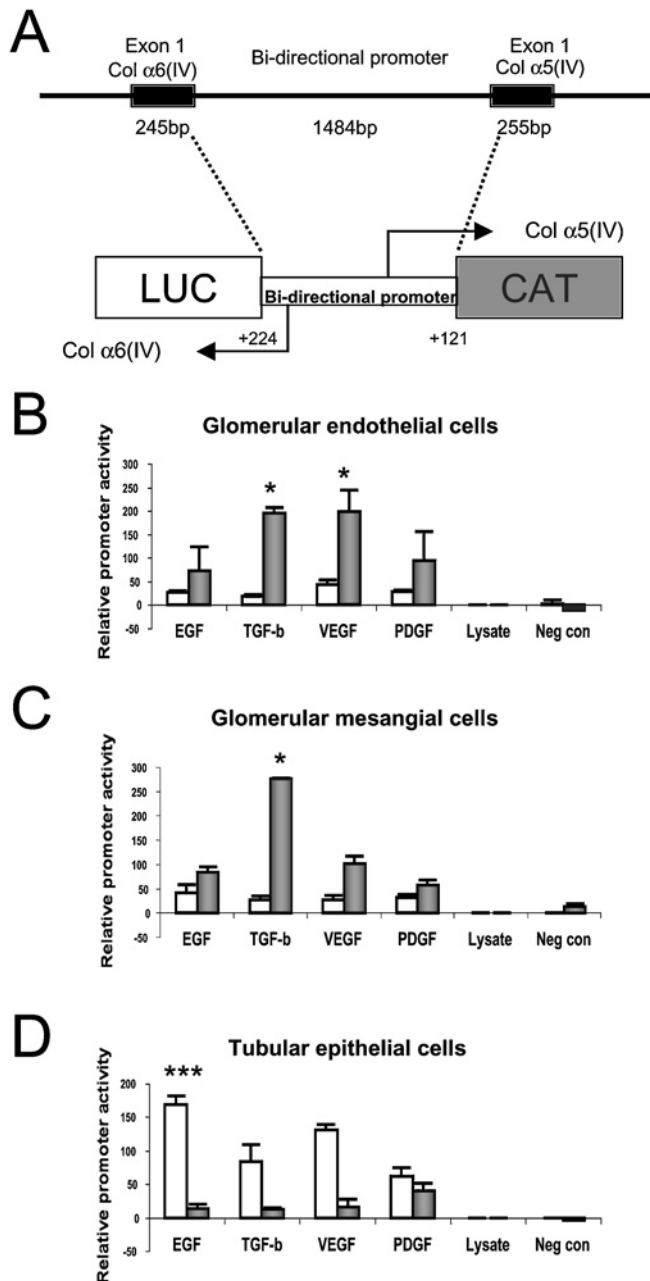
Figure 1   Bifunctional promoter construct (A) and analysis of COL4A5/COL4A6 gene expression (B–D)

(**A**) The bidirectional promoter element of the mouse COL4A5/COL4A6 gene pair spans 1484 bp between the first exons of the genes. The promoter, including 121 and 224 bp of the first exons of the α5(IV) and α6(IV) chains, was cloned between the CAT and Luciferase reporter genes. The orientation of the promoter is such that the α5(IV) promoter drives the expression of CAT (shown in grey) and the α6(IV) promoter drives the expression of Luciferase (shown in white). (**B**) Reporter gene expression in GEC in response to stimulation with growth factors. CAT expression is shown in grey and Luciferase expression in white. The expression is shown adjusted to the activity of the reporter genes for cells grown with 10 % FCS. Both VEGF and TGF-β significantly enhanced the expression from the α5(IV) promoter ($P = 0.03$ and 0.04), with moderate induction of expression by EGF and PDGF in GEC. Expression from the α6(IV) promoter could not be stimulated in the GEC by growth factors. (**C**) In MC, TGF-β significantly induces expression from the α5(IV) promoter ($P = 0.03$) and moderate enhancement is seen in response to both EGF and VEGF. Expression from the α6(IV) promoter could not be induced in the MC by growth factor stimulation. (**D**) In TEC, expression from the α6(IV) promoter was significantly enhanced in response to EGF ($P = 0.001$), but also VEGF and TGF-β stimulated the expression. PDGF induced moderate expression from the α5(IV) promoter in TEC, whereas the other cytokines did not induce expression from the α5(IV) promoter.

with promoter construct and grown with full serum supplementation.

Statistical analysis

RESULTS

Structure of the mouse COL4A5 gene

The mouse COL4A5 gene contains 53 exons and its size is approx. 209 kb and it is thus smaller than the human homologue of approx. 230–250 kb [38]. Introns 1–4 of the gene are large, their sizes ranging from 6.8 to 62.8 kb. Exons 1–3 encode for the signal peptide and the 7 S domain and also a portion of the collagenous domain. Exons 4–43 encode for the rest of the collagenous domain, and the latter part of exon 49 through 53 encodes for the N-terminal NC1 domain. The exon structure is completely conserved between the human and mouse genes, including the recently described two short 9 bp exons (41A and 41B) of the human COL4A5 gene [9,39]. The exon structure and sizes, the exon–intron boundaries and the intron sizes are shown in Table 1. Exon sizes vary from 9 to 1002 bp including the 3′-untranslated region and the NC1 domain, otherwise between 9 and 255 bp.

Structure of the mouse COL4A6 gene

The mouse COL4A6 gene contains 45 exons and the size of the gene is 304 kb. The mouse COL4A6 gene contains two very large introns at the beginning of the gene, intron 2 (146 kb) and intron 3 (93 kb). Exons 1–3 code for the signal peptide and the 7 S domain, exons 4–43 encode for the collagenous domain, and the end of exon 43 through 45 encode for the C-terminal NC1 domain. The exon structure is almost completely conserved between the human and mouse COLA6 genes. Exons 3 (human 81 bp versus mouse 84 bp), 14 (human 69 bp versus mouse 60 bp) and 19 (human 141 bp versus mouse 150 bp) are found to be of different sizes when comparing the human and mouse homologues [10]. These differences are well in line with the observed amino acid insertions and deletions between the mouse and human cDNAs for the α6(IV) chain [24]. The exon structure and sizes, the exon–intron boundaries and the intron sizes are shown in Table 2. Exon sizes vary from 36 to 951 bp when including the 3′-untranslated region and the NC1 domain, and within the collagenous domain the exon sizes range from 36 to 222 bp.

Structure of the bidirectional promoter of the mouse COL4A5/A6 genes

The bidirectional promoter spans 1484 bp between the transcription initiation sites, which were determined by 5′-RACE, of the mouse type IV collagen α5 and α6 genes (Figure 1A). The nucleotide sequence of the bidirectional promoter region for mouse COL4A5/A6 was aligned with the sequences of the published human COL4A5/A6 promoter [18], human COL4A1/A2 promoter [16], human COL4A3/A4 promoter [17] and mouse COL4A1/A2 promoter [40] (results not shown). The mouse COL4A5/A6 promoter region showed 78.7 % homology

Table 1   Exon–intron boundaries for the mouse α5(IV) collagen gene

The sizes of exons and introns are shown in the third and sixth columns respectively. Nucleotide sequences of the exon–intron boundaries are indicated by upper-case letters for exons and lower-case letters for introns. The ends of the introns are all conserved according to the GT-AG rule (shown in boldface).

Exon no.			Exon size			Intron size
1	ctctccctctctctctctta	TTTTAATCG	255	GAGGCTGCG	**gt**aagtcctccctcccgccc	62 897
2	ttccccctttatttcctat**ag**	GCCTGCCAT	60	GGAGAAAAG	**gt**aagtctgagtaaaaaaa	6891
3	acattcttgtttaattgc**ag**	GGAGAACGG	90	GGACAAAAG	**gt**atttgtttcagtagccaa	13 041
4	aataatattgtgttttc**ag**	GGTGATGAT	45	GGAATCAGA	**gt**aagcaatattcagtgtac	12 998
5	cgatttctttattattat**ag**	GGTCCTCCT	45	GGTCTTCCT	**gt**aagtatgggttttattc	83
6	tgttatatgcctttcaa**ag**	GGGATGCCA	63	GGAACCAAG	**gt**gaggttggttttttttt	3959
7	tcttcctttttaactccc**ag**	GGAGAACGT	54	GGTCCTCCA	**gt**aagtctacacttcagga	344
8	cttctttctttaataat**ag**	GGACCTCCT	27	GGTATGAAG	**gt**aagcatcacatgctaaga	2259
9	acactgttggtttcttt**ag**	GGGGAACCA	81	GGAATACAA	**gt**aagagtatggggaatttt	1807
10	ctttactccatttataac**ag**	GGCCCACCT	63	GGTTTAATG	**gt**aagttctgtaattgtat	2438
11	tccttcctgcgtcttccta**ag**	GGACCTCCT	36	GGACCAAAG	**gt**aatgttctttctgttcac	90
12	atggaaacttctctcccc**ag**	GGGAATATG	42	GGTGAAAAA	**gt**gagtaatcaaaagttggt	164
13	tatctttgttgtgtaaat**ag**	GGCGAACAA	93	GGAGATCAG	**gt**gagccctcagggaggggga	438
14	ttttctccatacctgcat**ag**	GGAGTTCCT	54	GGTCCTCCC	**gt**aagtattccaagactctt	93
15	tgaatattgatttttgc**ag**	GGTCCTCCT	57	GGCAAAAGA	**gt**aagtgttatgaccatcaa	439
16	cattgcttttgcatcctac**ag**	GGTAAACCA	45	GGAATCCCA	**gt**gagtagcaatggtccttt	4311
17	ccccatactctgttttaaa**ag**	GGTTTGGCT	54	GGTGAAAAG	**gt**aaaacttttctacattgtg	1133
18	tataacattgccttgaac**ag**	GGTCAAAAG	42	CCCGGACTT	**gt**aagttttctctcttacg	1775
19	aactccttttccataaaaa**ag**	GTAATTCCT	133	GACCTCCAG	**gt**aaaagacatgatggttat	7619
20	atatactttttgtcattt**ag**	GAACAGCAG	174	TGCCATCCA	**gt**aagccgtgctttatcct	322
21	tattttatctttaacctc**ag**	GAGATGAAA	84	GAGTAAAAG	**gt**ttgatcccaaacatcttc	9564
22	aatatttgtttttttgcc**ag**	GTGACAAAG	93	GTCCTCCAG	**gt**aaaatatcctccatatag	1259
23	ctgacctatgtgtgtcttt**ag**	GATCTCTTG	71	GGATTGCCA	**gt**aagctttgattattttat	1175
24	acttttttcttcaaactc**ag**	GGCATACCT	192	GGAGAGCCT	**gt**gagttggattaatatact	876
25	ttaaacatttccattttc**ag**	GGTGGAATT	169	GATTACCAG	**gt**atgtttaccctttgttgtc	1372
26	agtgttccatttgtttc**ag**	GTCCTAAAG	93	GTCTTCCAG	**gt**atgcaaagaatttatttt	96
27	agtgctttttttttttgt**ag**	GTGATCCTG	105	GTCCTAAAG	**gt**atgtaggatgaggtatgg	1223
28	aatggctgctatttttac**ag**	GTTTTCCAG	98	GGACCAAAG	**gt**ctgaaacaattcttaaa	1473
29	ttgtcatgtgtatgcttaa**ag**	GGAGAGCCA	151	GACCAAAAG	**gt**atggaggctcttgctact	8562
30	aactgtgctttttaaaaa**ag**	GTGATGTTG	114	GCCAACCTG	**gt**aagattacagtaaatgtg	3031
31	gttaatatgttgttaact**ag**	GCTTACATG	168	GATTTCCAG	**gt**aatgtttaataatctgtc	1129
32	gtaactgatatttttccaaa**ag**	GTGTCAAAG	90	GTCTTAAAG	**gt**aagaactgcagtttactg	1183
33	gtgtgctttgttgactttt**ag**	GTGATGATG	150	GCTTACCAG	**gt**aagtggggtggatatattc	3158
34	tttatgcttgtctttggaa**ag**	GTATTCCTG	99	GACTCCCAG	**gt**atgtaaaatacatcactg	1273
35	tatcaacattactttttct**ag**	GTCCCAAGG	90	GTTTTCCTG	**gt**gagtgatgaaaaagttaa	1820
36	acaatgcatattttcaac**ag**	GCCCTCAAG	140	GGCCCAAAG	**gt**aatcctagtctgatcatg	5543
37	gtcttтattcatgttttc**ag**	GGTGAACCT	127	GATTCCCAG	**gt**aggtgttcctttcccctt	8425
38	ctgaggtttttttattc**ag**	GAACAGCAG	81	GTCCTGTAG	**gt**aagtaagaaagacagcat	730
39	cctgtagtgttgttctt**ag**	GTGCCGGAG	99	GAGAACCAG	**gt**actatagtttttgtatgt	971
40	tatttttttttatcttaac**ag**	GTCAACCAG	51	GACTTTCTG	**gt**aaactttaaccaaacaag	1382
41	ttacactttttacttctac**ag**	GACAAAAGG	186	GGAGACCAG	**gt**atgatcgtgtgtggtatg	1651
42	gaacttccatccctaccc**ag**	GTCC	9	TCCAG	**gt**tagctccttaactccaaa	2758
43	gtgaactctgatcttccc**ag**	GTTT	9	TCAAG	**gt**tagactcttcgctggtca	2065
44	atttgttgtgcattaccaa**ag**	GTCCACCAG	134	GGACTCCCG	**gt**aagaagtgagactagata	2052
45	aatgactttattttatttt**ag**	GGTAACCCT	73	GATTCCCAG	**gt**atttgaaagagtggt	503
46	ttttatatcatcttttcca**ag**	GAATGAAAG	72	GGCCCCCAG	**gt**aagaattttttcttctcct	3599
47	ttatgtgtttgcttctta**ag**	GTCCTCCTG	129	GCCTACCAG	**gt**acctttgtaggtcatttt	250
48	atctattтctttccttgt**ag**	GTCCAATTG	99	GACAGCCAG	**gt**aaaatgactaaaacagtg	1959
49	caatgatttcgtacaaat**ag**	GTGCCCGTG	213	AAGACTTGG	**gt**aagaaaaatgacatctaa	7004
50	ttcttttcgattcatcatat**ag**	GTACGGCTG	178	CATTAGTCG	**gt**aaggcactggtttagctt	2378
51	ctcccccttttccttttacc**ag**	ATGTGCAGT	115	TTCATGATG	**gt**atgaaacacccatcttcc	347
52	cattttccttatctttttc**ag**	CATACAAGT	173	CATGTTCAA	**gt**aagtccttatggctttta	1304
53	attttttttttttatttcctat**ag**	CAAACCTCA	~1002		–	

with the human COL4A5/COL4A6 promoter sequence and exhibited 28.4, 27.9 and 28.8 % sequence homology with the human COL4A1/A2 and COL4A3/A4 as well as the mouse COL4A1/A2 promoter regions respectively. The sequences around the translation start sites of the COL4A5/A6 genes were highly conserved between human and mouse, revealing nearly 90 % similarity at the nucleotide level.

The determined sequence of the promoter region was found to be rich in G and C, and lacked a typical TATA box or related sequences. A CCAAT box [41,42], a potential interaction site with CBF (CCAAT-binding factor), was located 341 bp upstream of the start site of exon 1 of the COL4A5 gene. A GC box, a potential interaction site with the transcription factor Sp1, was located around 177 bp upstream of the start site of exon 1 of COL4A5. Another GC box is also located 46 bp upstream of the start site of exon 1 of COL4A6. CTC boxes, specifically recognized by the transcription factor CTCBF (CTC box-binding factor) [43,44], are located 180 and 60 bp upstream of the start site of exon 1 of COL4A5, as well as approx. 1200 bp upstream of the start site of exon 1 of COL4A6. Three AP1 sites were identified within the

Table 2 Exon–intron boundaries for the mouse α6(IV) collagen gene

The sizes of exons and introns are shown in the third and sixth columns respectively. Nucleotide sequences of the exon–intron boundaries are indicated by upper-case letters for exons and lower-case letters for introns. The ends of the introns are all conserved according to the GT-AG rule (boldface).

Exon no.			Exon size			Intron size
1	cgccccgtcccccctccattc	CTTCCTCCC	245	GCACCCTGG	**gt**gagctgctgctgctgaga	195
2	caccctccctctcctttc**ag**	ATTGTGGCT	52	GCAGAATCG	**gt**aaggcttcaaaagacccc	146 074
3	atctcttttttatttaata**ag**	GGACAGAAG	84	GGAGCAAGA	**gt**gagattgcttattgcatt	93 192
4	gtctttctgtcatctttt**ag**	GGGCACCCT	135	GGAGATAAG	**gt**aaggaacatctttgatgc	1577
5	tctctcttctctcttttt**ag**	GGTCCCATT	45	GGTATTCCG	**gt**aagtgtgtacatggaaga	4593
6	cttaattttctttgaac**ag**	GGCCACCCT	117	GGACCACCT	**gt**atgttaccagatctttgc	2109
7	gttttcatatttttatcc**ag**	GGGCTGCCT	69	GGAATAAGG	**gt**aagagtttctaaagcagg	1589
8	tctctatttctatttctc**ag**	GGAGATCCT	36	GGAATCCCT	**gt**aagtcatgcttcttgtgt	3215
9	aaaaaaaattctgtctcc**ag**	GGTCCATCA	63	GGCTTACAA	**gt**aagaccaacttaaaacta	1329
10	tttcaatttcgttaatct**ag**	GGTCCCCCA	36	GGTCCTGAG	**gt**aagtatgctttgggagca	198
11	cctgtacccttgtgtcccc**ag**	GGAAATATG	42	GGAGTCAAG	**gt**aaatagattctgaataca	1183
12	tggcctcttgattctac**ag**	GGGGATGTT	93	GGTTCCAAG	**gt**agtaaatgtttactttag	1338
13	gccttggttaattttttc**ag**	GGTGAGCCA	36	GGTGTCCCG	**gt**aagtcttaatttgtaaata	7827
14	tgacttctcttttccttc**ag**	GAATTTGGA	60	GGACCTAGG	**gt**atgacccagtgtcattt	1059
15	tgtggttcttcgattcac**ag**	GGTCCCATG	45	GGGAGACAG	**gt**atttgagtaacatccata	488
16	acctatgtttctattac**ag**	GGAAAGAAG	54	GGAATTAAG	**gt**aatatcattgggaagtac	711
17	aatttatttaatattttc**ag**	GGTGAAAAG	70	TGATCTCAG	**gt**gcaattctttgttgattt	942
18	ttttttcactgtaatgtc**ag**	GTTATCCTG	108	CATTAACAG	**gt**gaaacagactaatccact	1159
19	gtgtgggccttatccttcc**ag**	GTCTTCCAG	150	GCCCACCTG	**gt**aagttattccttctttaa	1331
20	cattctcttaaactctgc**ag**	GTCGCACAT	105	GAGTAAAAG	**gt**catttttgtttgtctcctc	1226
21	gtagtctgtttttccttc**ag**	GAGACTCTG	161	GGGACGCCA	**gt**aagtcttccgtaaggtct	682
22	gtctttttgtttcattgtc**ag**	GGGCTATTT	180	GGCATTCAG	**gt**aaatggatcattcttgga	692
23	tctttttctttgttctct**ag**	GGAGATGGT	184	GAGCTCAAG	**gt**gagtcatgtaatcagata	5013
24	ccacatgttatcatttcc**ag**	GAGTCACCT	72	GATTCCCAG	**gt**atggaatgtggccagaga	254
25	atacacatttggggtcttcc**ag**	GCTCTAAGG	108	GATTGCCAG	**gt**aaggaaaagggtcatcaa	946
26	ttttgcctctttacattt**ag**	GGTTTCCTG	222	GACCCAAGG	**gt**gactcactcacttgacc	377
27	tgccacctcctttttttc**ag**	GTTTGAATG	162	GCATTTCAG	**gt**aaatccattgaaggaaga	1663
28	cctcttctataaccattc**ag**	GGCACCCTG	171	GCCAAAAGG	**gt**aagtgaaataattaggaa	731
29	aaagtcttacttttcaac**ag**	GTGAAAAGG	144	GAGAAAAAG	**gt**aaggggggtatggggagac	932
30	tctgatctgccttccttc**ag**	GAGACAGAG	126	GGGCCAGAG	**gt**gctgaggcaggggatatt	442
31	agattttgcttctttccc**ag**	GTGACAAAG	182	GGGAAAAGC	**gt**aagtttatacagattctt	411
32	gtttccatattgctttca**ag**	GGTTCACAG	64	GTTTAAAAG	**gt**aaggaacattttgtcttc	638
33	atattttcataccctgc**ag**	GAGATCAAG	75	GTTTTAAAG	**gt**agggctagtacatttttg	458
34	gcaattgattgatcttct**ag**	GTGTCAAAG	108	GCTCCCCAG	**gt**accatgctacttgaaaac	93
35	caccttggtttcacactt**ag**	GACTTCAAG	108	GACCTAAAG	**gt**atgtgtgtttgctagcaa	2520
36	ctgatgttttctctcctt**ag**	GATTCCCTG	72	GCACTCCAG	**gt**aagaaaggccctgaaggt	264
37	ccctctgcatttcttcct**ag**	GAGCCAGTA	126	GTGACCAAG	**gt**aagtaagcccaaatcagg	3616
38	attttatttattctttgc**ag**	GTTCCCCAG	117	GAGAACGAG	**gt**aaaaataagaaacccaa	304
39	cccatcttatccct**ag**	GCCGCCCAG	162	GGCTAAAGG	**gt**aagtctgttccaccagct	150
40	tttcctttggcccttacc**ag**	GCATGAGAG	99	GGCCAAGAG	**gt**gagtgcagggccaagcct	1803
41	gtttcactacctgtcttc**ag**	GCTTTTCCG	147	GTCTGCAAG	**gt**aagaatggcaccagagaa	442
42	gtgcattttctctcttac**ag**	GTTCCAAAG	117	GATTTGAAG	**gt**caggttaacaaaacttgg	2359
43	ttatgtccttttctttctc**ag**	GAGCTCCGG	192	AGGATCTGG	**gt**aagtacagctagtccttg	678
44	tgtccatcttctcttcc**ag**	GATTTGCTG	287	TTCCTCATG	**gt**gaggccttgaaccattc	1615
45	accccctcccttctgtttc**ag**	CACACTGCT	∼951	–		

Table 3 Comparison of transcription factor interaction sites among promoters of α(IV) collagen chains

The number of sites identified within each promoter is indicated. Each promoter is indicated by the number of α(IV) chains it is located in between. For human α1/α2, mouse α1/α2 and human α3/α4, the first exon of each gene was included to adjust the length of the region compared to be similar to human and mouse α5/α6 promoter.

Promoter	CCAAT box	CTC box	GC box	AP1 site	AP2 site	E2F site
Human α1/α2	1	1	4	–	–	–
Mouse α1/α2	3	1	4	1	1	–
Human α3/α4	–	2	3	–	1	1
Human α5/α6	2	1	–	3	–	–
Mouse α5/α6	1	3	2	3	–	–

promoter. Additionally, we compared the potential transcription factor interaction sites among various promoter regions for the α-chains of type IV collagen (Table 3). The mouse COL4A5/ COL4A6 promoter contains two GC boxes, in contrast with the human COL4A5/COL4A6 promoter. The human COL4A1/ COL4A2 promoter has several GC boxes and, in the mouse COL4A1/COL4A2 promoter, an AP2 site was identified. For the human COL4A3/COL4A4 promoter, CCAAT box and AP1 sites were absent, whereas an AP2 site and an E2F site were present (Table 3).

Analysis of the bidirectional promoter shows cell-type-specific regulation

To study how transcription from the bidirectional promoter of the COL4A5/6 gene pair is regulated, we constructed a dual reporter gene construct in which the COL4A5 promoter drives the expression of the CAT reporter gene (upper strand) and the COL4A6 promoter drives the expression of the Luciferase reporter gene (lower strand; Figure 1A). The construct was generated to mimic the structural and functional relationship between the bidirectional promoter and the COL4A5/A6 gene regulation.

It is known from studies on the bidirectional promoter of the human COL4A5 and COL4A6 genes that this proximal promoter is sufficient to direct the expression of both genes [45]. However, it is unknown whether the proximal promoter is sufficient to regulate the expression in a cell-type-specific manner and whether differences in responses to growth factors could explain the observed differences in tissue localization. In the kidney, type IV collagen $\alpha5$ chain is expressed in the basement membranes connected to the glomeruli, the tubuli and Bowman's capsule [23]. The type IV collagen $\alpha6$ chain, on the other, is only found in the basement membrane connected to Bowman's capsule and to the distal tubuli [23]. We therefore chose three different cell lines from the kidney, MC, GEC and TEC to study the cell-type-specific regulation of expression of the $\alpha5(IV)$ and $\alpha6(IV)$ chains. To evaluate the capacity of the bifunctional promoter to regulate the expression of the $\alpha5(IV)$ and $\alpha6(IV)$ chains, we used cytokines previously established to increase the expression of type IV collagen in our *in vitro* reporter expression experiments [30,31].

In GEC, TGF-β and VEGF cause a statistically significant increase of CAT activity and thus of expression from the $\alpha5(IV)$ promoter, relative to cells grown in 10% FCS ($P = 0.04$ and 0.03; Figure 1B). Minimal CAT activity [$\alpha5(IV)$ expression] was stimulated with EGF and PDGF, although it was statistically insignificant when compared with cells grown in 10% FCS. In MC, TGF-β also caused a statistically significant up-regulation of CAT expression ($P = 0.03$; Figure 1C). EGF, VEGF and PDGF also enhanced the expression from the $\alpha5(IV)$ promoter in a statistically insignificant manner (Figure 1C). Interestingly, expression from the $\alpha6(IV)$ promoter was not significantly up-regulated in response to any of the analysed growth factors in either GEC or MC (Figures 1B and 1C). TEC stimulated with EGF displayed a statistically significant up-regulation of Luciferase activity and thus of expression from the $\alpha6(IV)$ promoter ($P = 0.001$; Figure 1D) when compared with cells grown in 10% FCS. VEGF stimulation also caused up-regulation of the $\alpha6(IV)$ promoter activity ($P = 0.09$). In TEC, expression from the $\alpha5(IV)$ promoter (CAT activity) was minimal in response to PDGF stimulation (statistically insignificant), but was negligible with the other growth factors analysed (Figure 1D).

DISCUSSION

We describe in the present study the characterization of the gene structures of the mouse type IV collagen $\alpha5$ and $\alpha6$ gene pair as well as the structure of the bidirectional promoter of these genes. The mouse COL4A5/A6 gene pair is almost 515 kb in size and thus extremely large. The human COL4A5/A6 gene pair has also been shown to be large, although estimations vary from 490 to 674 kb [38,46,47]. Both mouse COL4A5 and COL4A6 genes feature very long introns in the 5′-end. These large introns can potentially contain elements that both enhance and silence the expression of the mouse $\alpha5(IV)$ and $\alpha6(IV)$ genes, as has been shown for the human COL4A1/A2 genes [19,20]. The mouse COL4A5/A6 bidirectional promoter was cloned and characterized and the structure was compared with that of the published human COL4A5/A6, COL4A1/A2 and COL4A3/A4 as well as the mouse COL4A1/A2 promoters (Table 3). Several elements potentially important for the regulation of gene transcription were identified on both strands of the promoter.

To study the function of the proximal promoter of the mouse COL4A5/A6 gene pair, we generated a novel reporter gene construct in which the CAT and Luciferase reporter genes are driven by the $\alpha5(IV)$ and $\alpha6(IV)$ promoters respectively. It is evident from our results that the 1.5 kb bidirectional promoter is sufficient to drive the expression of both the genes. This has been shown to be the case also for the human COL4A5/A6 gene pair [45]. However, the responsiveness of the proximal promoter to stimulation by cytokines has not been studied previously. Our results show that the expression of the $\alpha5(IV)$ gene can be efficiently stimulated by TGF-β and VEGF in both GEC and MC, whereas expression of the $\alpha6(IV)$ gene cannot be stimulated in these cell types. On the other hand, a significant up-regulation of expression of the $\alpha6(IV)$ gene could be observed in TEC in response to EGF. TGF-β and VEGF were also potent stimulators of the expression of the $\alpha6(IV)$ gene. It is therefore evident that the proximal promoter of the mouse COL4A5/A6 gene pair can respond in a cell-type-specific manner to stimulation by the different growth factors in an *in vitro* setting.

Herzog et al. [48] analysed the human COL4A5/A6 promoter *in vivo* by generating transgenic mice in which the bidirectional promoter and up-stream sequences were used to direct the expression of the *lacZ* gene in transgenic mice. Since *lacZ* staining was only observed in certain cell layers of the upper gastrointestinal tract of the mice but not in the kidneys, lungs, skin or the other sites where the type IV collagen $\alpha5$ and $\alpha6$ chains have been localized [23,24], it was concluded that the highly specialized pattern of COL4A5/A6 expression is achieved *in vivo* by additional *cis*-acting elements interacting with the proximal promoter [48]. Our results show that the overall homology of the mouse and human COL4A5/A6 promoter region is approx. 78%. Therefore it is probable that the human COL4A5/A6 promoter might suboptimally function in transgenic mice.

In summary, our study demonstrates that the mouse type IV collagen $\alpha5$ and $\alpha6$ chains regulate their expression through a unique bifunctional promoter in a cell-specific manner. This demonstrates a unique plasticity associated with this promoter and offers an explanation as to why $\alpha5(IV)$ and $\alpha6(IV)$ chains do not always co-localize with each other in some basement membranes. The present study has broad implications to our understanding of how collagen chains might be regulated during various diseases, including organ fibrosis.

REFERENCES

1   Timpl, R. (1989) Structure and biological activity of basement membrane proteins. Eur. J. Biochem. **180**, 487–502

2   Saus, J., Wieslander, J., Langeveld, J. P., Quinones, S. and Hudson, B. G. (1988) Identification of the Goodpasture antigen as the alpha 3(IV) chain of collagen IV. J. Biol. Chem. **263**, 13374–13380

3   Gunwar, S., Saus, J., Noelken, M. E. and Hudson, B. G. (1990) Glomerular basement membrane. Identification of a fourth chain, alpha 4, of type IV collagen. J. Biol. Chem. **265**, 5466–5469

4   Hostikka, S. L., Eddy, R. L., Byers, M. G., Hoyhtya, M., Shows, T. B. and Tryggvason, K. (1990) Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome. Proc. Natl. Acad. Sci. U.S.A. **87**, 1606–1610

5   Hudson, B. G., Reeders, S. T. and Tryggvason, K. (1993) Type IV collagen: structure, gene organization, and role in human diseases. Molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. J. Biol. Chem. **268**, 26033–26036

6   Mariyama, M., Leinonen, A., Mochizuki, T., Tryggvason, K. and Reeders, S. T. (1994) Complete primary structure of the human alpha 3(IV) collagen chain. Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in human tissues. J. Biol. Chem. **269**, 23013–23017

7 Leinonen, A., Mariyama, M., Mochizuki, T., Tryggvason, K. and Reeders, S. T. (1994) Complete primary structure of the human type IV collagen alpha 4(IV) chain. Comparison with structure and expression of the other alpha (IV) chains. J. Biol. Chem. **269**, 26172–26177

8 Zhou, J., Mochizuki, T., Smeets, H., Antignac, C., Laurila, P., de Paepe, A., Tryggvason, K. and Reeders, S. T. (1993) Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in inherited smooth muscle tumors. Science **261**, 1167–1169

9 Zhou, J., Leinonen, A. and Tryggvason, K. (1994) Structure of the human type IV collagen COL4A5 gene. J. Biol. Chem. **269**, 6608–6614

10 Oohashi, T., Ueki, Y., Sugimoto, M. and Ninomiya, Y. (1995) Isolation and structure of the COL4A6 gene encoding the human alpha 6(IV) collagen chain and comparison with other type IV collagen genes. J. Biol. Chem. **270**, 26863–26867

11 Mariyama, M., Zheng, K., Yang-Feng, T. L. and Reeders, S. T. (1992) Colocalization of the genes for the alpha 3(IV) and alpha 4(IV) chains of type IV collagen to chromosome 2 bands q35-q37. Genomics **13**, 809–813

12 Griffin, C. A., Emanuel, B. S., Hansen, J. R., Cavenee, W. K. and Myers, J. C. (1987) Human collagen genes encoding basement membrane alpha1(IV) and alpha2(IV) chains map to the distal long arm of chromosome 13. Proc. Natl. Acad. Sci. U.S.A. **84**, 512–516

13 Kamagata, Y., Mattei, M. G. and Ninomiya, Y. (1992) Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha 4 chain of basement membrane collagen type IV and assignment of the gene to the distal long arm of human chromosome 2. J. Biol. Chem. **267**, 23753–23758

14 Morrison, K. E., Mariyama, M., Yang-Feng, T. L. and Reeders, S. T. (1991) Sequence and localization of a partial cDNA encoding the human alpha 3 chain of type IV collagen. Am. J. Hum. Genet. **49**, 545–554

15 Soininen, R., Huotari, M., Hostikka, S. L., Prockop, D. J. and Tryggvason, K. (1988) The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region. J. Biol. Chem. **263**, 17217–17220

16 Poschl, E., Pollner, R. and Kuhn, K. (1988) The genes for the alpha 1(IV) and alpha 2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure. EMBO J. **7**, 2687–2695

17 Momota, R., Sugimoto, M., Oohashi, T., Kigasawa, K., Yoshioka, H. and Ninomiya, Y. (1998) Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and alpha4(IV) collagen chains are arranged head-to-head on chromosome 2q36. FEBS Lett. **424**, 11–16

18 Sugimoto, M., Oohashi, T. and Ninomiya, Y. (1994) The genes COL4A5 and COL4A6, coding for basement membrane collagen chains alpha 5(IV) and alpha 6(IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. Proc. Natl. Acad. Sci. U.S.A. **91**, 11679–11683

19 Haniel, A., Welge-Lussen, U., Kuhn, K. and Poschl, E. (1995) Identification and characterization of a novel transcriptional silencer in the human collagen type IV gene COL4A2. J. Biol. Chem. **270**, 11209–11215

20 Killen, P. D., Burbelo, P. D., Martin, G. R. and Yamada, Y. (1988) Characterization of the promoter for the alpha1(IV) collagen gene. DNA sequences within the first intron enhance transcription. J. Biol. Chem. **263**, 12310–12314

21 Kalluri, R., Shield, C. F., Todd, P., Hudson, B. G. and Neilson, E. G. (1997) Isoform switching of type IV collagen is developmentally arrested in X-linked Alport syndrome leading to increased susceptibility of renal basement membranes to endoproteolysis. J. Clin. Invest. **99**, 2470–2478

22 Miner, J. H. and Sanes, J. R. (1994) Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequence, distribution, association with laminins, and developmental switches. J. Cell Biol. **127**, 879–891

23 Ninomiya, Y., Kagawa, M., Iyama, K., Naito, I., Kishiro, Y., Seyer, J. M., Sugimoto, M., Oohashi, T. and Sado, Y. (1995) Differential expression of two basement membrane collagen genes, COL4A6 and COL4A5, demonstrated by immunofluorescence staining using peptide-specific monoclonal antibodies. J. Cell Biol. **130**, 1219–1229

24 Saito, K., Naito, I., Seki, T., Oohashi, T., Kimura, E., Momota, R., Kishiro, Y., Sado, Y., Yoshioka, H. and Ninomiya, Y. (2000) Differential expression of mouse alpha5(IV) and alpha6(IV) collagen genes in epithelial basement membranes. J. Biochem. (Tokyo) **128**, 427–434

25 Hudson, B. G., Tryggvason, K., Sundaramoorthy, M. and Neilson, E. G. (2003) Alport's syndrome, Goodpasture's syndrome, and type IV collagen. N. Engl. J. Med. **348**, 2543–2556

26 Flinter, F. A., Cameron, J. S., Chantler, C., Houston, I. and Bobrow, M. (1988) Genetics of classic Alport's syndrome. Lancet **2**, 1005–1007

27 Kashtan, C. E. and Michael, A. F. (1996) Alport syndrome. Kidney Int. **50**, 1445–1463

28 Kashtan, C. E. (1998) Alport syndrome and thin glomerular basement membrane disease. J. Am. Soc. Nephrol. **9**, 1736–1750

29 Antignac, C. and Heidet, L. (1996) Mutations in Alport syndrome associated with diffuse esophageal leiomyomatosis. Contrib. Nephrol. **117**, 172–182

30 Grande, J. P., Melder, D. C. and Zinsmeister, A. R. (1997) Modulation of collagen gene expression by cytokines: stimulatory effect of transforming growth factor-beta1, with divergent effects of epidermal growth factor and tumor necrosis factor-alpha on collagen type I and collagen type IV. J. Lab. Clin. Med. **130**, 476–486

31 Tsilibary, E. C. (2003) Microvascular basement membranes in diabetes mellitus. J. Pathol. **200**, 537–546

32 Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F. et al. (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc. Natl. Acad. Sci. U.S.A. **99**, 16899–16903

33 Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**, 156–159

34 Zeisberg, M., Ericksen, M. B., Hamano, Y., Neilson, E. G., Ziyadeh, F. and Kalluri, R. (2002) Differential expression of type IV collagen isoforms in rat glomerular endothelial and mesangial cells. Biochem. Biophys. Res. Commun. **295**, 401–407

35 Nitta, K., Uchida, K., Murai, K., Horita, S., Tsutsui, T., Ozu, H., Kawashima, A., Yumura, W. and Nihei, H. (1993) Role of glomerular mesangial cells in the regulation of glomerular endothelial cell growth. Nippon Jinzo Gakkai Shi **35**, 663–669

36 Wolf, G., Thaiss, F., Schoeppe, W. and Stahl, R. A. (1992) Angiotensin II-induced proliferation of cultured murine mesangial cells: inhibitory role of atrial natriuretic peptide. J. Am. Soc. Nephrol. **3**, 1270–1278

37 Strutz, F., Zeisberg, M., Ziyadeh, F. N., Yang, C. Q., Kalluri, R., Muller, G. A. and Neilson, E. G. (2002) Role of basic fibroblast growth factor-2 in epithelial-mesenchymal transformation. Kidney Int. **61**, 1714–1728

38 Srivastava, A. K., Featherstone, T., Wein, K. and Schlessinger, D. (1995) YAC contigs mapping the human COL4A5 and COL4A6 genes and DXS118 within Xq21.3-q22. Genomics **26**, 502–509

39 Martin, P. H. and Tryggvason, K. (2001) Two novel alternatively spliced 9-bp exons in the COL4A5 gene. Pediatr. Nephrol. **16**, 41–44

40 Burbelo, P. D., Martin, G. R. and Yamada, Y. (1988) Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer. Proc. Natl. Acad. Sci. U.S.A. **85**, 9679–9682

41 Maity, S. N., Golumbek, P. T., Karsenty, G. and de Crombrugghe, B. (1988) Selective activation of transcription by a novel CCAAT binding factor. Science **241**, 582–585

42 Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) A multiplicity of CCAAT box-binding proteins. Cell (Cambridge, Mass.) **50**, 863–872

43 Fischer, G., Schmidt, C., Opitz, J., Cully, Z., Kuhn, K. and Poschl, E. (1993) Identification of a novel sequence element in the common promoter region of human collagen type IV genes, involved in the regulation of divergent transcription. Biochem. J. **292**, 687–695

44 Genersch, E., Eckerskorn, C., Lottspeich, F., Herzog, C., Kuhn, K. and Poschl, E. (1995) Purification of the sequence-specific transcription factor CTCBF, involved in the control of human collagen IV genes: subunits with homology to Ku antigen. EMBO J. **14**, 791–800

45 Segal, Y., Zhuang, L., Rondeau, E., Sraer, J. D. and Zhou, J. (2001) Regulation of the paired type IV collagen genes COL4A5 and COL4A6. Role of the proximal promoter region. J. Biol. Chem. **276**, 11791–11797

46 Zhang, X., Zhou, J., Reeders, S. T. and Tryggvason, K. (1996) Structure of the human type IV collagen COL4A6 gene, which is mutated in Alport syndrome-associated leiomyomatosis. Genomics **33**, 473–479

47 Kendall, E., Evans, W., Jin, H., Holland, J. and Vetrie, D. (1997) A complete YAC contig and cosmid interval map covering the entirety of human Xq21.33 to Xq22.3 from DXS3 to DXS287. Genomics **43**, 171–182

48 Herzog, C., Zhuang, L., Gorgan, L., Segal, Y. and Zhou, J. (2003) Tissue- and developmental stage-specific activation of alpha 5 and alpha 6(IV) collagen expression in the upper gastrointestinal tract of transgenic mice. Biochem. Biophys. Res. Commun. **311**, 553–560