THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 15, Issue of April 13, pp. 11791–11797, 2001

Printed in U.S.A.

Regulation of the Paired Type IV Collagen Genes *COL4A5* and *COL4A6*

ROLE OF THE PROXIMAL PROMOTER REGION*

Received for publication, August 16, 2000, and in revised form, November 21, 2000
Published, JBC Papers in Press, November 28, 2000, DOI 10.1074/jbc.M007477200

Yoav Segal, Liyan Zhuang, Eric Rondeau‡, Jean-Daniel Sraer‡, and Jing Zhou§

From the Renal Division, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115 and ‡INSERM U489, Hôpital Tenon, Paris, France

Tissue-specific expression patterns of the paired type IV collagen genes *COL4A5* and *COL4A6* form the basis for organ involvement in X-linked Alport syndrome, a disorder in which these genes are mutated. We investigated the proximal promoter region of *COL4A5* and *COL4A6* using glomerular visceral epithelial cells, in which *COL4A5* alone is transcribed; keratinocytes, in which the genes are co-transcribed; and additional model cell lines. By RNase protection assays, the intergenic region is 292 base pairs. Transcription start sites for two 5′ splice variants of *COL4A6* are 1 kilobase apart. Transient transfections with reporter gene constructs revealed that the minimal promoters for *COL4A5* and *COL4A6* are within 100 base pairs of their respective transcription start sites and are functionally distinct. In further transfection, gel shift and footprinting assays, we defined a bidirectional positive regulatory element, which functions in several cell types, but not in glomerular visceral epithelial cells selectively transcribing *COL4A5*. The existence of separate promoters for *COL4A5* and *COL4A6* permits fine control over their expression. Activation through the bidirectional element can bring about co-expression of the genes, exploiting their paired arrangement. Features of the proximal promoter region frame its roles in a hierarchy regulating type IV collagen gene expression.

Type IV collagen is a major constituent of basement membranes (1). Its lattice network is assembled from six α(IV) chains, α1(IV)–α6(IV), encoded by the genes *COL4A1* through *COL4A6* (2). In humans, these genes are arranged as head-to-head pairs *COL4A1-COL4A2*, *COL4A3-COL4A4,* and *COL4A5-COL4A6*, transcribed from opposite strands of chromosomes 13, 2, and X respectively. The functional differentiation of the individual α(IV) chains is evidenced in genetic diseases. Mutations in *COL4A5* give rise to X-linked Alport syndrome, a disorder of renal glomerular basement membrane (GBM),[1] variably associated with hearing loss and ocular abnormalities (3). Deletions involving *COL4A5* and *COL4A6* occur in Alport syndrome associated with diffuse leiomyomatosis (4). Mutations in *COL4A3* or *COL4A4* have been implicated in autosomal recessive Alport syndrome (5) and benign familial hematuria (6).

The "major" chains α1(IV) and α2(IV) are present ubiquitously in basement membranes, as $[\alpha1(IV)]_2\alpha2(IV)$ heterotrimers (1). In contrast the "minor" chains α3(IV) through α6(IV) are expressed in a tissue-restricted fashion. The α3(IV) and α4(IV) chains have been co-localized in basement membranes of the kidney, lung, choroid plexus, and neuromuscular junction (7–10). The α5(IV) chain is present in renal glomerular and distal tubular basement membranes, Bowman's capsule, and in basement membranes of the skin, trachea, eye, and neuromuscular junction (10–13). The α6(IV) chain has been localized in each of these structures except the GBM (12, 13). The importance of the discordant expression pattern in the GBM is illustrated by a recently described case of Alport syndrome associated with ectopic α6(IV) expression, and a putative regulatory mutation in the *COL4A6* promoter region (14).

The mechanisms accounting for the specialized distribution of the minor α(IV) collagen chains are poorly understood. Expression of the α3(IV), α4(IV), and α5(IV) chains is activated in the developing renal glomerulus, as basement membranes of distinct ultrastructure and type IV collagen subunit composition are elaborated by endothelial and visceral epithelial cells (10, 12, 15–19). Developmental regulation of minor chain expression has been characterized in other tissues as well (20, 21). Special considerations relating to expression of the minor chains arise in Alport syndrome. In this condition, pathogenic mutations in one of the minor chains generally lead to loss of the remaining chains (11, 12, 22). This effect can be achieved by global downregulation at the mRNA level, as demonstrated in a recent study of a canine model with a *COL4A5* mutation (23), or by post-transcriptional mechanisms (19, 24).

It is likely that the pairing of the type IV collagen genes facilitates their coordinated regulation. The regulation of *COL4A1* and *COL4A2*, separated by 130 bp, is served by a shared bidirectional promoter, which is subject to the effects of remote *cis*-acting elements (25–34). Transcriptional mechanisms regulating the minor chains have not been elucidated. In contrast to the *COL4A1-COL4A2* promoter, which applies to obligately co-expressed isoforms, a framework for the minor chains must account for coordinate expression in several tissues, as well as the discordant expression pattern of the α5(IV) and α6(IV) chains in GBM.

Using established models for keratinocytes, in which α5(IV) and α6(IV) are co-expressed, and renal glomerular visceral epithelial (GVE) cells, in which α5(IV) is expressed selectively, we provide evidence that tissue-specific expression patterns of type IV collagen arise at the transcriptional level. By detailed investigation of the proximal promoter region, we arrive at an initial understanding of the mechanisms regulating *COL4A5*

* This work was supported by National Institutes of Health Grants DK02419 and DK48317. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Renal Division, Brigham and Women's Hospital, Harvard Medical School, HIM 522, 77 Ave. Louis Pasteur, Boston, MA 02115. Tel.: 617-525-5860; Fax: 617-525-5861; E-mail: zhou@rics.bwh.harvard.edu.

[1] The abbreviations used are: GBM, glomerular basement membrane(s); bp, base pair(s); GVE, glomerular visceral epithelial; RT, reverse transcription; PCR, polymerase chain reaction; nt, nucleotide(s).

and *COL4A6* transcription. We demonstrate the existence of two promoters for *COL4A6*, which are functionally separable from the *COL4A5* promoter, and report the identification of a positive regulatory element, which can effect coordinated regulation. Finally, on the basis of findings in these and other cell lines, we frame potential roles for the proximal promoter region, in a hierarchy required to explain cell-specific transcription of this gene pair.

EXPERIMENTAL PROCEDURES

Cell Culture—Human GVE cells from the 56/10A1 line (35) were maintained in a 1:1 mixture of Dulbecco's modified Eagle's medium and Ham's F12 supplemented with 5 μg/ml insulin, 5 μg/ml transferrin, 5 ng/ml selenium, 10 ng/ml dexamethasone, and 1% fetal bovine serum. SCC-25 keratinocytes (ATCC CRL-1628) were maintained in a 1:1 mixture of Dulbecco's modified Eagle's medium and Ham's F12 supplemented with 0.4 μg/ml hydrocortisone and 10% fetal bovine serum. SK-N-SH brain neuroblastoma cells (ATCC HTB-11) were maintained in minimal essential Eagle's medium with Earle's balanced salts, 2 mM L-glutamine and 1 mM sodium pyruvate, supplemented with non-essential amino acids and 10% fetal bovine serum.

RNase Protection Assays—Total cellular RNA was isolated from nearly confluent monolayers, using Ultraspec RNA (Biotecx). Antisense riboprobes were transcribed using T3 or T7 RNA polymerase (Ambion) in the presence of [α-^{32}P]UTP and freed from unincorporated nucleotides using Sephadex G-50 columns (Roche Molecular Biochemicals). Assays were carried out using Ambion RPAII kits. To assess expression in model cell lines, 40-μg samples of RNA were incubated with 50,000 cpm of α5(IV) (nt 252–455, GenBank™/EBI accession U04520) and α6(IV) (nt 5353–5511, GenBank™/EBI accession U04845) riboprobes, and 500 cpm of low specific activity β-actin (nt 1464–1601, GenBank™/EBI accession NM_001101) riboprobe. Hybridizations were carried out at 60 °C overnight and digested with RNase T1 for 90 min at 37 °C. Products were analyzed on 6–8% denaturing polyacrylamide gels. Transcription start sites were determined by similar methods, using genomic subclones, shown in Fig. 3, as templates for riboprobe synthesis (36).

Amplification of Nascent Transcripts—Transcription of *COL4A5* and *COL4A6* in model cell lines was assessed by reverse transcription-polymerase chain reaction (RT-PCR) amplification of heterogeneous nuclear RNA, using intron-directed primers (37). RT was carried out on 1-μg samples of DNase I-digested total RNA, using the Superscript Preamplification System (Life Technologies). PCR amplifications were carried out on 5% of the RT products, in the presence of [α-^{32}P]dCTP. Primers were designed on the basis of previously described genomic sequence (38). By 3′-terminal nucleotide position (GenBank™/EBI accession AL034369), these were for intron 1 of *COL4A5*: RT, 5′-CAAGAATGTTTCTGACTTGGC-3′ (72935); PCR forward, 5′-TAGTTATTTAGTTGCAGAGGCAGAG-3′ (72749); and PCR reverse, 5′-GCAATTGGATTGAAAATGGTAGGC 3′ (72913). For intron 2 of *COL4A6* these were: RT, 5′-GAGGGGGCAAAGTGTGAAG-3′ (54281); PCR forward, 5′-CTAACTATCTAAGGGGCTGTGC-3′ (54521); and PCR reverse, 5′-ATTGGCTGATGGTGAGATTTGTATC-3′ (54355). Control amplifications of β-actin mRNA were carried out using oligo(dT)-primed RT products. Primers were, by 3′-terminal nucleotide position (GenBank™/EBI accession NM_001101): PCR forward, 5′-CAGCAGTCGGTTGGAGCGAGCATC-3′ (1389); and PCR reverse, 5′-ACACGAAAGCAATGCTATCACCTC-3′ (1557). In preliminary experiments using RNA from HepG2 cells, which express both genes, we ascertained exponential phases of the PCR reactions. These ranges of cycle number were used in assaying the model cell lines. RT-PCR products of nascent transcripts were analyzed on 6% polyacrylamide gels and detected by autoradiography.

Reporter Gene Constructs and Transient Transfections—Parent constructs pα5E1, pα6E1a, and pα6E1b contain genomic restriction fragments upstream from a luciferase reporter gene, in the vector pGL3-Basic (Promega), as shown in Fig. 5. Deletion constructs were produced by excision of appropriate restriction fragments, with subsequent blunt-ending if necessary, and religation (Fig. 5).

Scanning substitution mutations MT1 through MT5, containing *Sac*II restriction sites in place of native 6-bp sequences, were introduced into pα5E1 and pα6E1a (see Fig. 6). To generate each mutant, two amplifications A and B were carried out in parallel, with pα5E1 or pα6E1a as template, so as to yield products overlapping at primer-encoded *Sac*II restriction sites. Reaction A primers (designated by 3′-terminal nucleotide position (GenBank™/EBI accession D28116) were the conserved 5′-ACTTCCAGACTAGTTGACTGA-3′ (637), and for the respective mutants: MT1, 5′-TCCCCGCGGTAGATCTATTTGTAATTGGCTTTG-3′ (738); MT2, 5′-TCCCCGCGGTAGTCCCAGTTTAGATCTATTTG-3′ (750); MT3, 5′-TCCCCGCGGTCTGAGATCCATCCGCTAAA-3′ (777); MT4, 5′-TCCCCGCGGAAGCACGGCCCCTCTGAG-3′ (791); and MT5, 5′-TCCCCGCGGACAATAAAAAGCACGGCCCC-3′ (797). Reaction B primers from the multiple cloning site of pGL3-Basic were the conserved 5′-ATCGATAGGTACCGAGCTCT-3′ for pα5E1 templates and 5′-TACCGGAATGCCAAGCTTAC-3′ for pα6E1a templates, and for the respective mutants: MT1, 5′-TCCCCGCGGGACTATTTTTTTAGCGGATGGA-3′ (789); MT2, 5′-TCCCCGCGGTAGCGGATGGATCTCAGAGG-3′ (798); MT3, 5′-TCCCCGCGGGTGCTTTTTATTGTTACTCATTA-3′ (825); MT4, 5′-TCCCCGCGGGTTACTCATTAGAAACAAATTTTG-3′ (838); and MT5, 5′-TCCCCGCGGTTAGAAACAAATTTTGGTCGGT-3′ (844). Amplification products were digested with appropriate restriction enzymes and reinserted into the parent constructs in compound ligations. Mutations were verified by direct sequencing. All plasmids were prepared for transient transfections using Qiagen plasmid kits.

For transient transfections, cell monolayers grown to 60–90% confluence on 60-mm dishes were incubated for 6 h with 2.5 μg of the luciferase test construct and 1.5 μg of the normalizing vector pSV-β galactosidase (Promega), in 1.5 ml of Opti-MemI medium (Life Technologies) containing 10 μg/ml Polybrene (American Bioanalytical). Cells were exposed to 30% dimethyl sulfoxide for 3 min, restored to complete medium and harvested 32 to 34 h thereafter in Reporter Lysis Buffer (Promega). Luciferase activities were determined using a Monolight 2001 luminometer (Analytical Luminescence Laboratory Inc.) with the Luciferase Assay System (Promega). They were normalized for transfection efficiency using β-galactosidase activities, determined by the β-Galactosidase Enzyme Assay System (Promega). Luciferase activities are expressed relative to paired promoterless controls, and represent mean ± S.E. for at least three transfections.

Nuclear Extracts—Nuclear extracts were prepared by standard protocols, with minor modifications (39). Ten 175-cm^2 flasks with nearly confluent monolayers were used for each preparation, and all steps were carried out at 4 °C. Cells were washed twice and scraped into phosphate-buffered saline, collected, centrifuged at 1000 × *g* for 15 min, consolidated, and centrifuged again, yielding packed cell volumes of ~0.5 ml. Following resuspension in 2.5 ml of Buffer A (10 mM KCl, 10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 0.5 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride), cells were placed on ice for 10 min, then pelleted. They were resuspended in 1 ml of Buffer A, Dounce homogenized (A type pestle), then centrifuged at 1000 × *g* for 15 min. The supernatant was removed, and the pellet was centrifuged at 30,000 × *g* for an additional 10 min, yielding a crude nuclear fraction. This was resuspended in 0.5 ml of Buffer C (420 mM NaCl, 20 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 20% glycerol), Dounce homogenized (B type pestle), then extracted in Buffer C for 30 min. Following centrifugation at 30,000 × *g* for 30 min, the supernatant was collected, and dialyzed overnight against a buffer containing 100 mM KCl, 10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, and 20% glycerol. Extracts were stored in aliquots at −80 °C, and protein concentrations were estimated using the Bio-Rad protein assay kit.

Gel Shift and Chemical Footprinting Assays—Probes for gel shift and chemical footprinting assays were prepared from 73-bp *Msl*I-*Fok*I fragments (nt 728–800, GenBank™/EBI accession D28116), which were isolated from pα6E1a or mutant pα6E1a(MT2) by multiple-step restriction digest and gel purification. Fragments were labeled at one end by fill-in of a CCCC overhang, using Klenow fragment (New England BioLabs) in the presence of [α-^{32}P]dGTP, followed by chase with excess cold dGTP. Labeled fragments were isolated from 9% non-denaturing polyacrylamide gels. Unlabeled fragments were used in competition assays.

For gel shift assays, 2,000 cpm of labeled probe was added to 8 μg of nuclear extract, 5 μg of poly(dI·dC), and 3 μg of bovine serum albumin in 20 μl of gel shift buffer containing 50 mM KCl, 20 mM HEPES, pH 7.9, 5 mM $MgCl_2$, 1 mM dithiothreitol, and 20% glycerol. Mixtures were incubated at 4 °C for 30 min, then electrophoresed on 8% non-denaturing polyacrylamide gels in 1× TGE buffer (25 mM Tris, 190 mM glycine, 1 mM EDTA).

We found that coupled gel shift/footprinting assays were required to delineate DNA elements interacting with proteins in nuclear extracts, because of the low occupancy of these sites. We used a previously described copper phenanthroline footprinting technique, with minor modifications (40). Specialized reagents were purchased from Aldrich, and all steps were carried out at 4 °C. Scaled-up gel shift assays were carried out using 100,000 cpm of labeled probe, 40 μg of nuclear extract

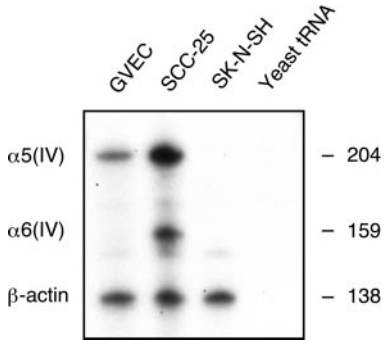

FIG. 1. **Expression profiles of model cell lines.** RNase protection assays for simultaneous detection of α5(IV), α6(IV), and β-actin mRNA expression in the indicated cell lines were carried out as described under "Experimental Procedures." Numbers refer to molecular size in bases.

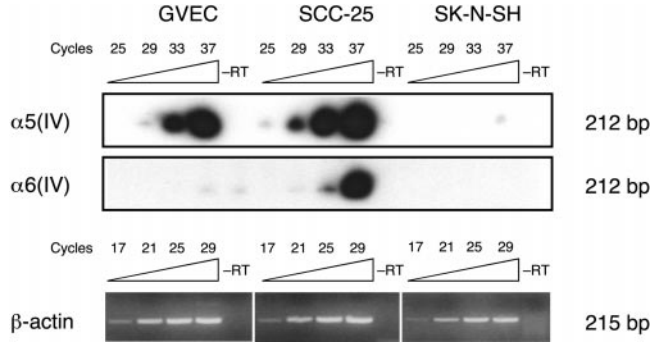

FIG. 2. **Identification of nascent transcripts.** Transcriptional activity of *COL4A5* and *COL4A6* was assessed by RT-PCR of nascent transcripts, as described under "Experimental Procedures." Products are shown for the indicated PCR cycles, with negative RT controls obtained at the highest cycle. RT-PCR amplifications of β-actin mRNA served as controls. Results are representative of three experiments. Molecular sizes of the products are indicated.

from SCC-25 cells, 15 μg of poly(dI·dC), and 15 μg of bovine serum albumin in 100 μl of gel shift buffer. After electrophoresis, gels were immersed in 400 ml of 10 mM Tris-HCl, pH 8.0, followed by addition of 40 ml of DNA cleavage reagent (36 ml of distilled water, added to 2 ml of 9 mM $CuSO_4$, and 2 ml of 40 mM 1,10-phenanthroline monohydrate in ethanol). The cleavage reaction was initiated by addition of 40 ml of 58 mM 3-mercaptopropionic acid, and terminated at 3–90 min by addition of 40 ml of 30 mM 2,9-dimethyl-1,10-phenanthroline. Bands corresponding to unbound DNA and the major bandshift S_1 (see "Results") were visualized by autoradiography, then excised. DNA was eluted overnight at 37 °C in 0.5 M ammonium acetate, 10 mM Mg acetate, 1 mM EDTA, 0.1% SDS and isolated by sequential phenol/chloroform extraction, chloroform extraction, and ethanol precipitation, with 10 μg of glycogen as a carrier. Pellets were rinsed twice in 85% ethanol and electrophoresed on 10% sequencing gels. Digestion patterns were visualized by autoradiography and analyzed by scanning densitometry.

RESULTS

Characterization of Model Systems—COL4A5 and *COL4A6* expression was analyzed in several cell lines by RNase protection assay. We report results for GVE cells, in which *COL4A5* alone is expressed; SCC-25 keratinocytes, in which both genes are expressed; and SK-N-SH neuroblastoma cells, in which neither gene is expressed (Fig. 1). To ascertain whether transcription accounts for differences among these mRNA expression patterns, we amplified nascent *COL4A5* and *COL4A6* transcripts, using intron-directed primers. In this surrogate for the nuclear runoff assay (37, 41), detection of nascent transcripts corresponded to steady-state mRNA levels (Fig. 2), indicating that transcription, rather than mRNA processing or degradation, is the major determinant of steady-state mRNA levels. The mRNA expression profiles in GVE and SCC-25 cells provide evidence that the type IV collagen composition of corresponding renal glomerular and epidermal basement membranes arises at the transcriptional level.

Structure of the Proximal Promoter Region—Transcription start sites were mapped by RNase protection assays, using RNA from SCC-25 and GVE cells (Fig. 3). Start sites for transcripts of *COL4A5* and 5′ splice variants A and B of *COL4A6* are upstream relative to earlier determinations by cDNA sequencing (4, 42). The intergenic region is 292 bp. By inspection, the 5′-flanking regions of exons E1 of *COL4A5* and E1a of *COL4A6* lack classical TATA boxes but lie downstream from AT-rich regions, which could confer specificity to sites of transcriptional initiation (Fig. 4). Additional features of the intergenic region include a tandem array of CCAAT boxes and a central CTC box, a functional motif, which is conserved in the intergenic region separating the *COL4A1* and *COL4A2* genes (28, 31). Transcript B of *COL4A6* is associated with at least two closely situated start sites. The 5′-flanking region of exon E1b also lacks a classical TATA box, but is associated at the upstream start site with a sequence that is identical to a transcriptional initiator (Inr) at 7 of 8 nucleotides (43).

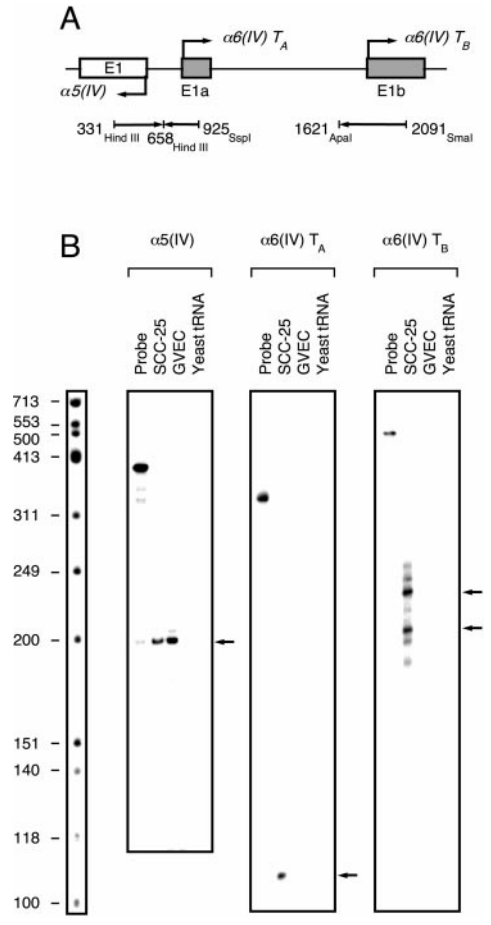

FIG. 3. **Determination of transcription start sites.** *A*, map of the promoter region of *COL4A5* and *COL4A6*. Numbering is according to GenBank™/EBI accession D28116. Restriction fragments used as templates are indicated, with *arrows* denoting transcription in the antisense orientation. T_A and T_B refer to Transcripts A and B of *COL4A6*. *B*, analysis of protected bands. The major protected species are indicated by *arrows*. Sizing was by comparison against an end-labeled *Hin*fI-digested ØX174 DNA ladder (Promega), giving 200 bp for *COL4A5*, 105 bp for Transcript A of *COL4A6*, and 209 and 234 bp for predominant bands associated with Transcript B of *COL4A6*. The data are from two different gels, which were run under similar conditions and gave overlapping molecular size markers.

We undertook a detailed analysis of the proximal promoter region, as this is where cell-specific regulatory elements are expected to exert their effects. In transient transfections with

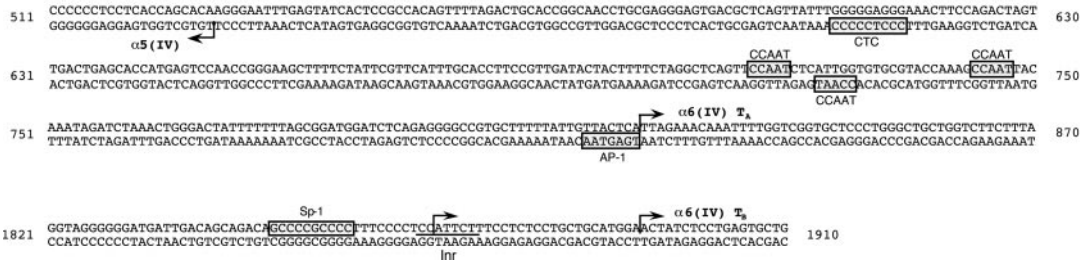

FIG. 4. **Structure of the *COL4A5* and *COL4A6* promoter region.** Transcription start sites, sequence bearing identity to a transcriptional initiator (*Inr*) at 7 of 8 nucleotides, and CCAAT, AP-1, Sp-1, and CTC consensus regulatory elements are indicated (see text).

luciferase reporter gene constructs, we sought to characterize the minimal *COL4A5* and *COL4A6* promoters and to identify important proximal regulatory elements.

A parent construct pα5E1 reporting in the *COL4A5* orientation gave luciferase activity 17- and 47-fold higher than the promoterless vector pGL3-Basic, in SCC-25 and SK-N-SH cells, respectively (Fig. 5). This finding localizes the minimal *COL4A5* promoter to the intergenic region. Inasmuch as the two cell types differ in *COL4A5* expression (Figs. 1 and 2), overlying mechanisms are implicated in regulation of the endogenous promoter. Transfections with construct pα6E1a gave luciferase activity 17- and 40-fold higher than the promoterless control, in SCC-25 and SK-N-SH cells (Fig. 5). These activities place the minimal *COL4A6* promoter in the intergenic region as well and implicate additional mechanisms in repressing *COL4A6*, in SK-N-SH cells. In GVE cells, luciferase activity was only 5- to 6-fold over promoterless controls, in both the *COL4A5* and *COL4A6* orientations (Fig. 5). The lack of orientation specificity for these constructs suggests that distal elements are involved in the selective activation of *COL4A5* (see "Discussion").

Construct pα6E1b containing the 5′-flanking region of E1b was used to assess whether transcription of *COL4A6* is directed by two promoters. Reporter gene activity was 7- to 9-fold higher than the promotorless control in the three cell lines (Fig. 5), indicating the presence of a functional promoter, which is distinct from that upstream of E1a and thus a site of differential regulation for Transcript B. Given comparable promoter activities among the cell lines, additional regulatory mechanisms must apply, to explain differences in steady-state levels of Transcript B.

We refined localization of the minimal promoters for *COL4A5* and *COL4A6* by deletion analysis. In SCC-25 and SK-N-SH cells, deletion of nt 755–876 (GenBank™/EBI accession D28116), in construct pα5E1, caused 82 and 87% reductions in activity, respectively (Fig. 5), indicating the presence of a positive regulatory element in this region. Further deletion to nt 625 did not alter reporter gene activity appreciably. Construct pα5E1(625), which contains 99 base pairs upstream from the transcription start site, gave luciferase activity significantly higher than promoterless controls, suggesting that the minimal promoter resides within this fragment. In GVE cells, serial deletions of pα5E1 to nt 625 had no significant effects, with activity for pα5E1(625) remaining significantly higher than control. These results consign the minimal promoter to the immediate 5′-flanking region, as in the other cell lines, but provide evidence against the function of a positive regulatory element. Deletion of the tandemly arranged CCAAT boxes or of the CTC box (see Fig. 4) had no significant effects on reporter gene activity.

By comparison to serial deletions of pα5E1, those of pα6E1a were associated with only modest (<55%) reductions in luciferase activity (Fig. 5). Plasmid pα6E1a (755), containing 67 bp upstream of the transcription start site, gave luciferase activity significantly above promoterless controls. These data indicate the presence of a minimal promoter immediately upstream of exon E1a, which is thus functionally separable from the *COL4A5* promoter, as it is from a second downstream *COL4A6* promoter.

Taken together, our findings support a configuration of nonoverlapping minimal promoters for *COL4A5* and *COL4A6*. In SCC-25 and SK-N-SH cells, the positive regulatory effects of an intergenic, potentially bidirectional, element are evident. This configuration has valuable attributes for regulation of the paired genes, although it is by itself insufficient to explain cell-specific expression patterns (see "Discussion").

A Positive Regulatory Element in the Intergenic Region—To investigate further the positive regulatory element between nt 755 and 876, we utilized constructs with directed substitutions in this region, in SCC-25 cells (Fig. 6). Four sites MT1 through MT4 were chosen to span the region, and a fifth, containing an AP-1 consensus site (MT5), was targeted directly. The MT2 mutation had a clear bidirectional effect, reducing luciferase activity 82% in the *COL4A5* orientation, and 68% in the *COL4A6* orientation, relative to the parent constructs. Among the remaining constructs, we observed appreciable changes relative to the parent constructs with MT4 only, which caused a 2-fold increase in luciferase activity, in the *COL4A6* orientation. This mutation was associated with the recruitment of additional shifted bands in gel shift assays (data not shown), suggesting the introduction of binding sites for spurious transcriptional activation. Substitution of the AP-1 site had no significant effects, suggesting that this element is not required for basal transcription.

To investigate the interactions of nuclear factors with the bidirectional element, gel shift assays were carried out, using labeled probes containing the region of the MT2 mutation (nt 728–800). Several band shifts were observed following incubation of wild-type probes with nuclear extracts from SCC-25 or GVE cells but not following incubation of MT2 mutant probes (Fig. 7). The banding patterns were complex and differed between the two cell types, suggesting multiple interactions between the DNA fragments and cell-specific nuclear factors. Most of the band shifts were abolished by competition with excess unlabeled wild-type fragment, but not with MT2 mutant fragments, supporting the specificity of the interactions, and the importance of the region affected by the MT2 mutation.

To delineate the boundaries of the bidirectional enhancer, a coupled gel shift/chemical footprinting assay was carried out, using the predominant shifted band S_1 (Fig. 7). The shifted probe was protected from cleavage at the site 5′-TAGC-3′ (Fig. 8), a site of preferential cleavage for copper phenanthroline (44). Analysis by scanning densitometry indicated strong protection of the region 5′-TAGCGGATGGATCTCA-3′, with weaker protection of the adjacent region 5′-GAGGGG-3′. This site lacks known regulatory elements, by database search (TRANSFAC, release 3.5; Ref. 45). The MT2 mutation likely disrupts binding at the site secondarily.

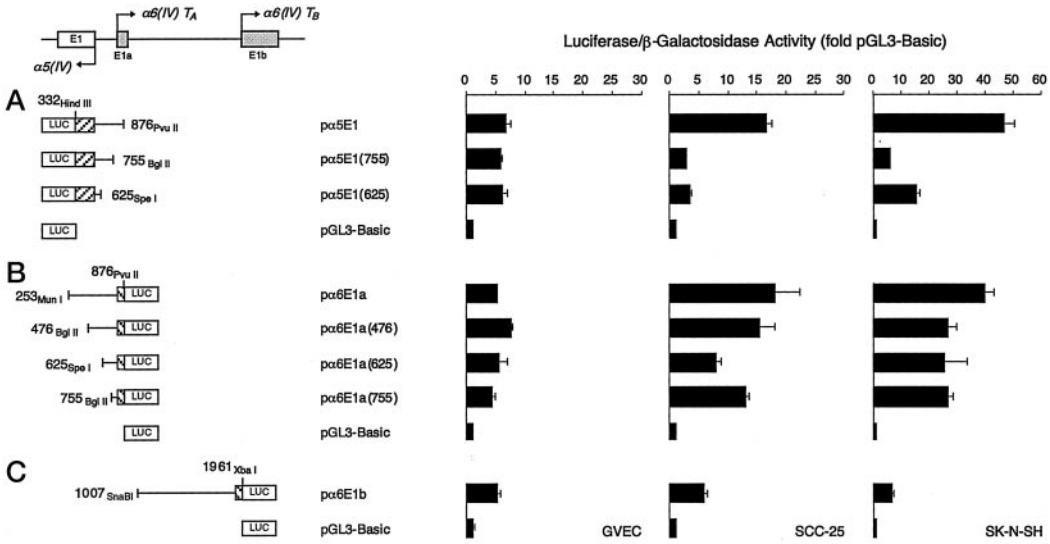

FIG. 5. **Promoter activity in model cell lines.** Parent constructs containing the 5′-flanking regions of the *COL4A5* and *COL4A6* genes were transiently transfected into the indicated cell lines. A map of the region and restriction sites used in generating the constructs are shown. Numbering is according to GenBank™/EBI accession D28116. *Hatched areas* denote 5′-untranslated regions. *A*, *COL4A5* constructs; *B*, *COL4A6* E1a constructs; *C*, *COL4A6* E1b constructs. Luciferase activities represent mean ± S.E. (n = 3).

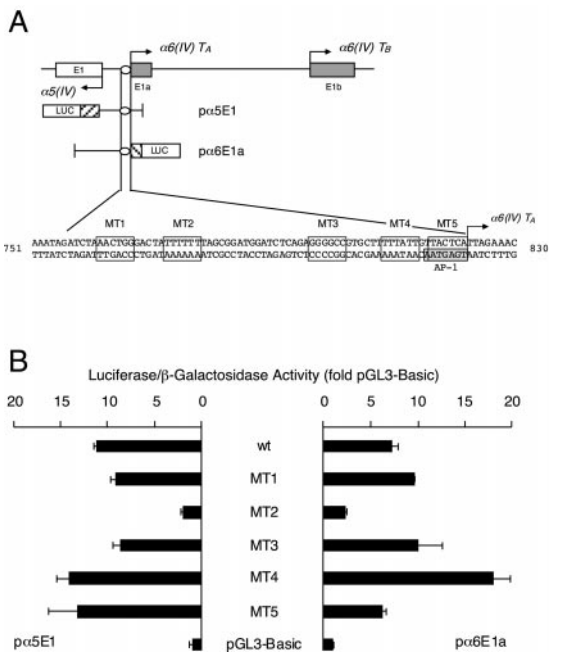

FIG. 6. **Mutational analysis of a putative bidirectional element.** *A*, substitution mutations. The region containing a positive regulatory element is indicated by an *oval*. MT1 through MT5 were introduced into pα5E1 and pα6E1a, as described under "Experimental Procedures." Numbering is according to GenBank™/EBI accession D28116. *B*, luciferase activity for wild-type and mutant constructs. Transient transfections were carried out in SCC-25 cells, and activities are shown for constructs reporting in the *COL4A5* or *COL4A6* orientation.

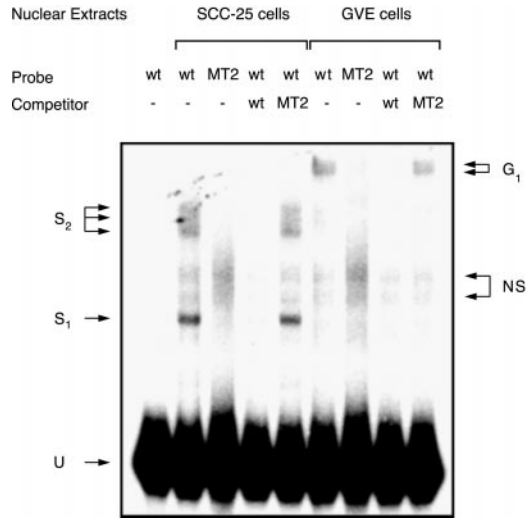

FIG. 7. **Binding of the bidirectional element.** Nuclear extracts from SCC-25 or GVE cells were incubated in gel shift assays with wild-type or MT2-mutant probes, in the absence or in the presence of excess ~150-fold unlabeled competitor. Prominent band shifts are designated S_1 and S_2 for SCC-25 cells, the latter representing a triplet, and G_1 for GVE cells, representing a doublet. Nonspecific band shifts (*NS*) remained in the presence of excess competitor. *U* indicates unbound probe.

DISCUSSION

Several divergently transcribed gene pairs have been described in higher eukaryotes. Within the type IV collagen gene family itself, pairing has been conserved through an evolutionary pattern of duplications, across mammalian species (46). This implies biological advantage, deriving perhaps from stoichiometric requirements for related gene products. Pairing has been conserved within other gene families as well, including the histone genes, which are arranged within duplicated clusters as oppositely transcribed pairs, which share common promoter elements (47, 48). Additional examples such as that afforded by the functionally related genes *GPAT-AIRC* (49), whose products catalyze steps in the purine nucleotide biosynthetic pathway, and *TAP1-LMP2* (50), whose products are central in antigen presentation, reinforce the notion that paired gene arrangements have been conserved in response to evolutionary selection pressures. For the type IV collagen genes specifically, regulatory schemes operating within this framework account for a highly specific subunit distribution, among tissues that vary considerably in their basement membrane composition and turnover requirements.

We have demarcated the promoter region of the genes *COL4A5* and *COL4A6* and defined a novel positive regulatory element situated in the intergenic region. Our results are consistent with a scheme in which *COL4A5* and *COL4A6* are associated with functionally distinct promoters, which can be regulated separately or coordinately. The existence of separate

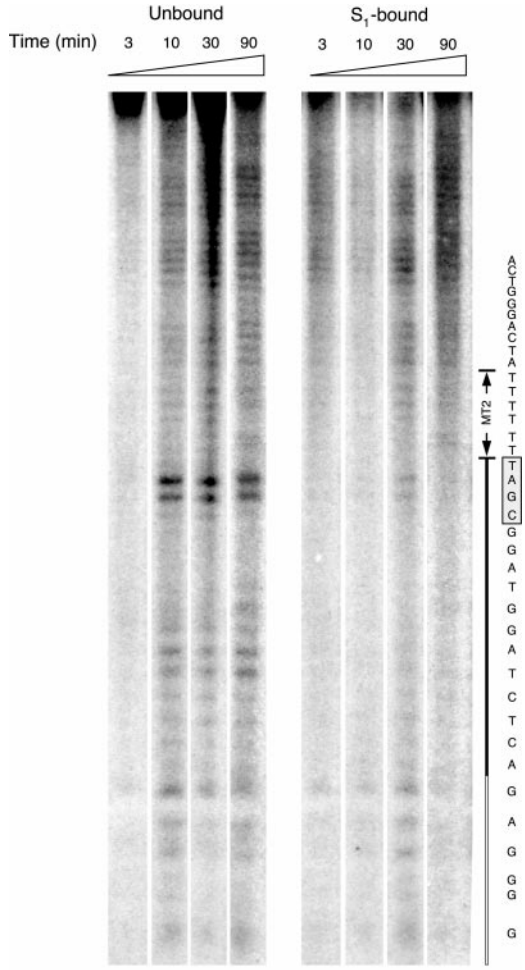

FIG. 8. **Footprinting of the bidirectional element.** S_1-bound and unbound probes were resolved by gel shift assays, subjected to chemical cleavage *in situ*, for the indicated periods, and isolated as described under "Experimental Procedures." Approximately equal amounts of the recovered products were analyzed on 10% sequencing gels. A strongly protected region is indicated by the *filled box*, with a weakly protected region indicated by the *open box*. The *shaded sequence* 5′-TAGC-3′ is cleaved preferentially by copper phenanthroline (44). The site of the MT2 mutation is indicated.

promoters for Transcripts A and B of *COL4A6* in particular affords a basis for their differential expression in several tissues, including kidney, lung, skin, and placenta (42). If the chief benefit of type IV collagen gene pairing is to facilitate coordinated transcription, then the overlapping *COL4A1-COL4A2* promoter, and the distinct *COL4A5* and *COL4A6* promoters, represent minor but important variations permitting finely regulated synthesis of their respective gene products. The positive element defined here may confer the capacity for co-expression of *COL4A5* and *COL4A6* genes in keratinocytes and other cell types (12, 13).

The complexity of DNA binding in the region of the bidirectional element is evidenced by the appearance of multiple band shifts in our gel shift assays, each reflecting one of multiple DNA-protein interactions. Abolition of these band shifts by the MT2 mutation, with corresponding functional effects in SCC-25 cells, suggests that binding of this *cis*-acting element is a condition for bidirectional transcription. If the site is viewed as a nidus for the binding of transcription factor complexes, then differential effects in SCC-25 and GVE cells can reflect the activity of a transcriptional activator in SCC-25 cells, and/or the activity of a transcriptional repressor in GVE cells. Inasmuch as the footprints in our assays do not correspond to

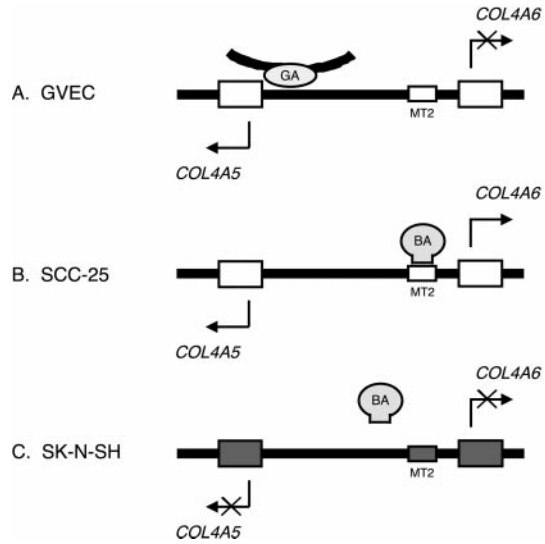

FIG. 9. **Model of the *COL4A5-COL4A6* proximal promoter region.** The model contains non-overlapping sites for recruitment of transcriptional initiation complexes (*large boxes*), and a positive regulatory element near the MT2 site (*small box*). *A*, uncoupled activation of *COL4A5* transcription, as in GVE cells, depends on inactivity of the bidirectional element. A proposed glomerular activator of *COL4A5* (GA) acts through a distal enhancer. *B*, coupled transcriptional activation, as in SCC-25 cells, occurs through the MT2 site, bound by a bidirectional activator (BA). *C*, in SK-N-SH cells the genes are repressed (*darkened symbols*) and activator-unresponsive.

consensus regulatory elements, we cannot posit roles for specific transcription factors in binding the bidirectional element. Moreover, the activity of this element in nonexpressing SK-N-SH cells implicates higher-order regulation by sequence-specific transcriptional repressors operating through distal elements, or by general, *i.e.* chromatin-mediated, mechanisms of transcriptional repression (51).

We have also demonstrated selective transcription of *COL4A5* in GVE cells, corresponding to the discordant expression pattern of the α5(IV) and α6(IV) chains in the renal GBM. On the basis of our promoter studies, distal elements are required to account for selective activation of the *COL4A5* promoter. Nonetheless, features of the proximal promoter region may afford a partial explanation. In GVE cells, inactivity of the bidirectional element is evidenced functionally in transient transfections, and is corroborated by differential patterning in gel shift assays. This inactivity could be permissive for selective *COL4A5* activation, ensuring absence of the α6(IV) chain in GBM, by uncoupling transcription of the two genes. In principle, selective transcription could also be accomplished through asymmetric effects of *cis*-acting elements on a shared promoter; however, this scheme, described for other co-expressed gene pairs including *COL4A1-COL4A2* (31, 32), may not afford sufficient selectivity for *COL4A5* and *COL4A6*.

A working model for the proximal promoter region is presented in Fig. 9. Its major features are separability of the minimal promoters for the two genes, and provision for the bidirectional element, which can act as a shared near-upstream positive regulator. In this model, as applied to GVE cells, inactivity of the bidirectional element is an important feature of selective *COL4A5* transcription, which is mediated by distal enhancers (*panel A*). In other cell types, activity of the bidirectional element is determined by the activation state of the locus as a whole, and the presence of specific MT2-binding transcription factors (*panels B*, *C*). Distinctions between SCC-25 and SK-N-SH cells reside in the mechanisms by which the genes are "pre-activated," *i.e.* rendered activator-responsive, by favorable local chromatin structures and DNA methylation patterns

(51, 52). These mechanisms could involve keratinocyte-specific enhancers, and/or somatic silencing mechanisms in cells like those from the SK-N-SH line, derived from the central nervous system. Predictions from the working model can now be addressed in further studies, including comparisons among DNase I-hypersensitive sites in chromatin from these model cell lines, and analysis of promoter constructs *in vivo*, in transgenic mouse lines.

An interesting issue raised by this and earlier work on other paired genes is the mechanism by which a proximate bidirectional element coordinates transcription. One possibility is that elements bound at the MT2 site alternate among minimal promoters, contributing to assembly and/or activation of basal transcription factors at each of these sites discretely. Another possibility is that promoter modulation occurs simultaneously, even to the point of interlocking divergent transcriptional initiation events. Simultaneous interactions may provide a more compelling case for conservation of the paired gene arrangement, as has been suggested for other gene pairs (53).

The bidirectional element is the most potent *cis*-acting element in the intergenic region, in SCC-25 and SK-N-SH cells. By deletion analysis, we could not demonstrate roles for a tandem array of CCAAT boxes, or for a central CTC box. These elements may play important roles in directing *COL4A5* and *COL4A6* transcription under different experimental conditions, or in cellular response to external mediators, such as have been shown to affect the *COL4A1-COL4A2* promoter (54). A putative regulatory mutation causing ectopic α6(IV) expression in the GBM is in close proximity to tandem CCAAT boxes (14), raising the possibility that the mutation facilitates transcriptional activation through these sites. It remains to be seen whether any of these promoter elements are conserved in the intergenic region separating the *COL4A3* and *COL4A4* genes, therein providing a possible means of coordinating the expression of the four minor chains. Such elements could be targets of altered regulation in cases of Alport syndrome in which decreased mRNA levels for all of the minor chains accompany mutation in a single type IV collagen gene (23). Whereas it is possible that this effect arises at the level of mRNA stability, as has been proposed (23), it is also possible that the presence of a mutation generates a signal feeding back on the transcription of related genes.

In summary, we have shown that the transcription of the *COL4A5* and *COL4A6* genes is directed by distinct promoters, which permit highly selective expression of their gene products. In keratinocytes, co-expression of α5(IV) and α6(IV) correlates with the activity of a bidirectionally active proximal promoter element, which is bound at a previously unrecognized site. This element is likely to be important in establishing tissue-specific patterns of α5(IV) and α6(IV) co-expression. Nonetheless, it is clear from the complexity of binding, as well as from the largely unexplained phenomenon of discordant expression in the GBM, that the proximal promoter region functions within a transcriptional hierarchy, which is largely responsible for the specialized distribution of type IV collagen.

Acknowledgments—We thank Christine Herzog, Ursula Kaiser, and Stephen Reeders for critical review of the manuscript.

REFERENCES

1. Timpl, R. (1989) *Eur. J. Biochem.* **180,** 487–502
2. Hudson, B. G., Reeders, S. T., and Tryggvason, K. (1993) *J. Biol. Chem.* **268,** 26033–26036
3. Barker, D. F., Hostikka, S. L., Zhou, J., Chow, L. T., Oliphant, A. R., Gerken, S. C., Gregory, M. C., Skolnick, M. H., Atkin, C. L., and Tryggvason, K. (1990) *Science* **248,** 1224–1227
4. Zhou, J., Mochizuki, T., Smeets, H., Antignac, C., Laurila, P., de Paepe, A., Tryggvason, K., and Reeders, S. T. (1993) *Science* **261,** 1167–1169
5. Mochizuki, T., Lemmink, H. H., Mariyama, M., Antignac, C., Gubler, M.-C., Pirson, Y., Verellen-Dumoulin, C., Chan, B., Schröder, C. H., Smeets, H. J., and Reeders, S. T. (1994) *Nat. Genet.* **8,** 77–82
6. Lemmink, H. H., Nillesen, W. N., Mochizuki, T., Schröder, C. H., Brunner, H. G., van Oost, B. A., Monnens, L. A. H., and Smeets, H. J. M. (1996) *J. Clin. Invest.* **98,** 1114–1118
7. Butkowski, R. J., Wieslander, J., Kleppel, M., Michael, A. F., and Fish, A. J. (1989) *Kidney Int.* **35,** 1195–1202
8. Kleppel, M. M., Santi, P. A., Cameron, J. D., Wieslander, J., and Michael, A. F. (1989) *Am. J. Path.* **134,** 813–825
9. Sanes, J. R., Engvall, E., Butkowski, R., and Hunter, D. D. (1990) *J. Cell Biol.* **111,** 1685–1699
10. Miner, J. H., and Sanes, J. R. (1994) *J. Cell Biol.* **127,** 879–891
11. Yoshioka, K., Hino, S., Takemura, T., Maki, S., Wieslander, J., Takekoshi, Y., Makino, H., Kagawa, M., Sado, Y., and Kashtan, C. E. (1994) *Am. J. Path.* **144,** 986–996
12. Peissel, B., Geng, L., Kalluri, R., Kashtan, C., Rennke, H. G., Gallo, G. R., Sun, M. J., Hudson, B. G., Yoshioka, K., Reeders, S. T., Neilson, E., and Zhou J. (1995) *J. Clin. Invest.* **96,** 1948–1957
13. Ninomiya, Y., Kagawa, M., Iyama, K., Naito, I., Kishiro, Y., Seyer, J. M., Sugimoto, M., Oohashi, T., and Sado, Y. (1995) *J. Cell Biol.* **130,** 1219–1229
14. Kawai, S., Nomura, S., Naito, I., Sado, Y., Fukushima, T., and Osawa, G. (1996) *J. Am. Soc. Nephrol.* **7,** 1615
15. Laurie, G. W., Horikoshi, S., Killen, P. D., Segui-Real, B., and Yamada, Y. (1989) *J. Cell Biol.* **109,** 1351–1362
16. Desjardins, M., and Bendayan, M. (1991) *J. Cell Biol.* **113,** 689–700
17. Kleppel, M. M., Fan, W. W., Cheong, H. I., and Michael, A. F. (1992) *J. Biol. Chem.* **267,** 4137–4142
18. Lee, L. K., Pollock, A. S., and Lovett, D. H. (1993) *J. Cell. Physiol.* **157,** 169–177
19. Heidet, L., Cai, Y., Guicharnaud, L., Antignac, C., and Gubler, M.-C. (2000) *Am. J. Pathol.* **156,** 1901–1910
20. Simoneau, A., Herring-Gillam, F. E., Vachon, P. H., Perreault, N., Basora, N., Bouatrouss, Y., Pageot, L.-P., Zhou, J., and Beaulieu, J.-F. (1998) *Dev. Dyn.* **212,** 437–447
21. Enders, G. C., Kahsai, T. Z., Lian, G., Funabiki, K., Killen, P. D., and Hudson, B. G. (1995) *Biol. Reprod.* **53,** 1489–1499
22. Gubler, M. C., Knebelman, B., Beziau, A., Broyer, M., Pirson, Y., Haddoum, F., Kleppel, M. M., and Antignac, C. (1995) *Kidney Int.* **47,** 1142–1147
23. Thorner, P. S., Zheng, K., Kalluri, R., Jacobs, R., and Hudson, B. G. (1996) *J. Biol. Chem.* **271,** 13821–13828
24. Miner, J. H., and Sanes, J. R. (1996) *J. Cell Biol.* **135,** 1403–1413
25. Burbelo, P. D., Martin, G. R., and Yamada, Y. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 9679–9682
26. Pöschl, E., Pollner, R., and Kühn, K. (1988) *EMBO J.* **7,** 2687–2695
27. Burbelo, P. D., Bruggeman, L. A., Gabriel, G. C., Klotman, P. E., and Yamada, Y. (1991) *J. Biol. Chem.* **266,** 22297–22302
28. Bruggeman, L. A., Burbelo, P. D., Yamada, Y., and Klotman, P. E. (1992) *Oncogene* **7,** 1497–1502
29. Tanaka, S., Kaytes, P., and Kurkinen, M. (1993) *J. Biol. Chem.* **268,** 8862–8870
30. Heikkilä, P., Soininen, R., and Tryggvasson, K. (1993) *J. Biol. Chem.* **268,** 24677–24682
31. Fischer, G., Schmidt, C., Opitz, J., Cully, Z., Kühn, K., and Pöschl, E. (1993) *Biochem. J.* **292,** 687–695
32. Schmidt, C., Fischer, G., Kadner, H., Genersch, E., Kühn, K., and Pöschl, E. (1993) *Biochim. Biophys. Acta* **1174,** 1–10
33. Haniel, A., Welge-Lüβen, U., Kühn, K., and Pöschl, E. (1995) *J. Biol. Chem.* **270,** 11209–11215
34. Takeoka, H., Iehara, N., Uematsu-Yanagita, M., Abe, H., Sunamoto, M., Yamada, Y., Kita, T., and Doi, T. (1998) *Biochem. Biophys. Res. Commun.* **252,** 290–295
35. Delarue, F., Virone, A., Hagege, J., Lacave, R., Peraldi, M.-N., Adida, C., Rondeau, E., Feunteun, J., and Sraer, J.-D. (1991) *Kidney Int.* **40,** 906–912
36. Renieri, A., Bassi, M. T., Galli, L., Zhou, J., Giani, M., de Marchi, M., and Ballabio, A. (1994) *Hum. Mutat.* **4,** 195–198
37. Elferink, C. J., and Reiners, J. J. (1996) *BioTechniques* **20,** 470–477
38. Segal, Y., Peissel, B., Renieri, A., de Marchi, M., Ballabio, A., Pei, Y., and Zhou, J. (1999) *Am. J. Hum. Genet.* **64,** 62–69
39. Dignam, J. D., Martin, P. L., Shastry, B. S., and Roeder, R. G. (1983) *Methods Enzymol.* **101,** 582–598
40. Papavassiliou, A. G. (1994) in *Methods in Molecular Biology* (Kneale, G. G., ed), Vol. 30, pp. 43–78, Humana Press, Totowa, NJ
41. Gabbitas, B., Pash, J. M., Delany, A. M., and Canalis, E. (1996) *J. Biol. Chem.* **271,** 9033–9038
42. Sugimoto, M., Oohashi, and Ninomiya, Y. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11679–11683
43. Kollmar, R., and Farnham, P. J. (1993) *Proc. Soc. Exp. Biol. Med.* **203,** 127–139
44. Veal, J. M., and Rill, R. L. (1988) *Biochemistry* **27,** 1822–1827
45. Wingender, E., Kel, A. E., Kel, O. V., Karas, H., Teinemeyer, T., Dietze, P., Knüppel, R., Romaschenko, A. G., and Kolchanov, N. A. (1997) *Nucleic Acids Res.* **25,** 265–268
46. Zhou, J., Ding, M., Zhao, Z., and Reeders, S. T. (1994) *J. Biol. Chem.* **269,** 13193–13199
47. Sturm, R. A., Dalton, S., and Wells, J. R. E. (1988) *Nucleic Acids Res.* **16,** 8571–8587
48. Huh, N.-E., Hwang, I., Lim, K., You, K.-K., and Chae, C.-B. (1991) *Nucleic Acids Res.* **19,** 93–98
49. Gavalas, A., and Zalkin, H. (1995) *J. Biol. Chem.* **270,** 2403–2410
50. Wright, K. L., White, L. C., Kelly, A., Beck, S., Trowsdale, J., and Ting, J. P. (1995) *J. Exp. Med.* **181,** 1459–1471
51. Ashraf, S. I., and Ip, Y. T. (1998) *Curr. Biol.* **8,** R683–R686
52. Blackwood, E. M., and Kadonaga, J. T. (1998) *Science* **281,** 60–63
53. Yu, W., Misulovin, Z., Suh, H., Hardy, R. R., Jankovic, M., Yannoutsos, N., and Nussenzweig, M. C. (1999) *Science* **285,** 1080–1084
54. Kuncio, G. S., Alvarez, R., Li, S., Killen, P. D., and Neilson, E. G. (1996) *Am. J. Physiol.* **259,** F120–F125

Regulation of the Paired Type IV Collagen Genes*COL4A5* and *COL4A6* : ROLE OF THE PROXIMAL PROMOTER REGION

Yoav Segal, Liyan Zhuang, Eric Rondeau, Jean-Daniel Sraer and Jing Zhou

J. Biol. Chem. 2001, 276:11791-11797.
doi: 10.1074/jbc.M007477200 originally published online November 28, 2000
