Am. J. Hum. Genet. 63:1329–1340, 1998

Determination of the Genomic Structure of the COL4A4 Gene and of Novel Mutations Causing Autosomal Recessive Alport Syndrome

Eileen Boye,[1] Géraldine Mollet,[1] Lionel Forestier,[1] Lola Cohen-Solal,[1] Laurence Heidet,[1] Pierre Cochat,[3] Jean-Pierre Grünfeld,[2] Jean-Bernard Palcoux,[4] Marie-Claire Gubler,[1] and Corinne Antignac[1]

[1]INSERM U423 and [2]Service de Néphrologie, Hôpital Necker-Enfants Malades, Université René Descartes, Paris; [3]Département de Pédiatrie, Hôpital Edouard Herriot, Lyon, France; and [4]Service de Pédiatrie B, Hôtel-Dieu, Clermont-Ferrand, France

Summary

Autosomal recessive Alport syndrome is a progressive hematuric glomerulonephritis characterized by glomerular basement membrane abnormalities and associated with mutations in either the COL4A3 or the COL4A4 gene, which encode the α3 and α4 type IV collagen chains, respectively. To date, mutation screening in the two genes has been hampered by the lack of genomic structure information. We report here the complete characterization of the 48 exons of the COL4A4 gene, a comprehensive gene screen, and the subsequent detection of 10 novel mutations in eight patients diagnosed with autosomal recessive Alport syndrome. Furthermore, we identified a glycine to alanine substitution in the collagenous domain that is apparently silent in the heterozygous carriers, in 11.5% of all control individuals, and in one control individual homozygous for this glycine substitution. There has been no previous finding of a glycine substitution that is not associated with any obvious phenotype in homozygous individuals.

Introduction

Type IV collagen forms the basic structural component of basement membranes. There are six different type IV collagen chains, named α1–α6, which are encoded by distinct but highly homologous genes (Hudson et al. 1993). The genes are unique in that they have characteristic pairwise, head-to-head chromosomal arrangements. These arrangements are the result of evolution by the intrachromosomal and subsequent interchromosomal duplication of a common ancestral gene, and they can allow coordinated regulation of a particular gene pair. The genes of each pair can be divided, by their homology, into two classes, the COL4A1- or COL4A2-like genes, thus defining their evolutionary origin (Zhou et al. 1994). The corresponding α chains may assemble into various different heterotrimers to produce different functional isoforms. The α1 and α2 chains are known to form (α1)2α2 trimers and are ubiquitously coexpressed in all basement membranes (Timpl 1989), whereas the other chains have more restricted tissue-specific expression patterns (Kleppel et al. 1989; Ninomiya et al. 1995; Peissel et al. 1995). The α3, α4, and α5 chains are strongly expressed in the glomerular basement membrane (GBM) (Peissel et al. 1995; Ninomiya et al. 1995).

The hereditary nephritis Alport syndrome (AS) (MIM 104200), a progressive, hematuric glomerulonephritis often associated with sensorineural hearing loss and specific eye lesions, is a predominantly X-linked condition characterized by GBM abnormalities (Atkin et al. 1988), and many cases have been shown to be due to mutations in the X chromosome–located COL4A5 gene (Barker et al. 1990; Knebelmann et al. 1996; Lemmink et al. 1997). However, up to 15% of AS cases are autosomally inherited and are thought to be due to mutations in the COL4A3 and COL4A4 genes, which are located on chromosome 2 (Mariyama et al. 1992). In addition, autosomal dominant forms of AS and dominantly inherited familial benign hematuria (FBH) (MIM 141200) have been shown to be linked to the COL4A3-COL4A4 locus, which identifies these as a spectrum of disorders with similar underlying pathology (Lemmink et al. 1996; Jefferson et al. 1997). Mutation screening in these two genes to evaluate their involvement in the molecular etiology of such disorders has been hampered by the availability of only partial intron-exon structure of the genes. To date, only five COL4A3 mutations (Mochizuki et al. 1994; Lemmink et al. 1994; Ding et al. 1995; Knebelmann et al. 1995) and two COLA4A4 mutations (Moch-

Received May 14, 1998; accepted for publication September 10, 1998; electronically published October 26, 1998.

Address for correspondence and reprints: Dr. Corinne Antignac, INSERM U 423, Tour Lavoisier 6° étage, Hôpital Necker-Enfants Malades, 149, rue de Sèvres, 75015 Paris, France. E-mail: antignac@necker.fr

0002-9297/98/6305-0010$02.00

izuki et al. 1994) have been reported in autosomal recessive AS, and only one COL4A4 mutation (Lemmink et al. 1996) in familial benign hematuria.

The human COL4A4 cDNA is translated into a 1,690–amino acid polypeptide, which consists of a 38-residue signal peptide, a 1,421-residue collagenous domain, and a 231-residue noncollagenous (NC1) domain (Leinonen et al. 1994). Only 6 exons at the 3′ end have been characterized (Kamagata et al. 1992; Sugimoto et al. 1993; Mochizuki et al. 1994). More recently, the organization of the 5′ end has been determined (Momota et al. 1998). We report here the complete characterization of the COL4A4 gene structure, and we demonstrate the usefulness of this information by the detection of 10 novel mutations in eight patients diagnosed with autosomal recessive AS.

Patients and Methods

Patients

Thirty-one unrelated autosomal recessive AS patients of European (21), North African (6), Middle Eastern (2), and West Indian (2) origin were investigated. All fulfilled the diagnostic criteria for AS (Antignac et al. 1994), and autosomal recessive inheritance was assessed on the existence of one or more of the following: (1) severity of the disease in young females, (2) consanguinity in the family, (3) microscopic hematuria in the father of a male patient, (4) immunohistochemical changes typical of autosomal recessive AS (Gubler et al. 1995), and (5) exclusion of linkage to the COL4A5 locus. Informed consent was obtained from all individuals or their parents. A control group was composed of 48 individuals with no obvious phenotype. The group comprised 34 Europeans, 8 North Africans, 2 Africans, and 4 individuals from the Middle East.

Long-Range PCR Amplification

Twenty 18–20-base primer pairs located ~250 nucleotides apart were chosen at random to cover the entire 5.2-kb COL4A4 cDNA sequence (Leinonen et al. 1994), by means of the OLIGO 5.0 program (NBI). The primers were used to amplify COL4A4 genomic DNA from the YAC 929_G_1 from the CEPH library. We amplified ~50 ng of total yeast DNA from melted YAC plugs, using the Expand Long Template PCR System (Boehringer Mannheim) according to the manufacturer's protocol. The 10 × buffer containing 22.5 mM $MgCl_2$ and optimized for amplifying the largest templates was used. The PCR reaction was performed with a GeneAmp PCR System 9600 (Perkin-Elmer) under the following conditions (Rozet et al. 1996): an initial denaturation step at 94°C for 4 min; 10 cycles of denaturation at 93°C for 10 s and annealing/extension at 59°C for 15 s; 20 cycles of denaturation at 93°C for 10 s and annealing/extension at 59°C for 15 min, with a 20 s/cycle stepwise augmentation of the annealing/extension time; and a final extension step at 68°C for 10 min. Failure to obtain an amplification product under these conditions, followed by less stringent annealing temperature conditions, was thought to indicate either the presence of one or more large introns (>12 kb) or the presence of an exon-exon boundary within one or both of the primer sequences. To rectify this problem, we took advantage of the high homology between type IV collagen chains and redesigned primers outside putative exon-exon boundaries, modeled on the mouse col4a2 genomic structure (Buttice et al. 1990). The amplification products were purified with either the Geneclean II kit (Bio 101) or the Wizard PCR Preps DNA purification system (Promega). Purified fragments were sequenced directly, with specific primers for the cDNA or the genomic sequence and the PRISM Ready Reaction Sequencing kit (Perkin-Elmer) on an automatic fluorometric DNA sequencer (Perkin-Elmer).

Ligation-Mediated PCR

To characterize the structures of the exons, which were inaccessible by long-range PCR amplification because of the presence of very large introns, we designed a series of external and internal primers in specific regions of interest (i.e., within an exon with an undefined exon-intron boundary and directed toward the potential boundary). We thereby characterized the intervening genomic sequences by ligation-mediated PCR, using the technique described by Kere et al. (1992), replacing YAC-arm primers with region-specific primers. Amplification products were purified and sequenced as described above.

Mutation Detection by SSCP Analysis and Direct Sequencing

We amplified all coding exons by PCR using flanking intronic primers selected with the OLIGO 5.0 program (NBI) and screened by SSCP analysis, as described elsewhere (Saunier et al. 1997). Where mobility shifts were detected, automated direct sequencing was used to characterize the sequence variation. SSCP analysis and direct sequencing were used to detect and to evaluate the presence of the nucleotide variants in the control population.

Results

Intron-Exon Structure Determination

Three YAC clones from the human CEPH library (Cohen et al. 1993) containing at least part of the COL4A4 gene—clones 870_B_6 (400 kb), 929_G_1 (1220 kb), and 933_C_7 (1690 kb)—were found by screening Unigene, a gene-oriented view of European Molecular Bi-

ology Laboratory/GenBank sequence entries at the National Center for Biotechnology Information, and the STS-based map of the human genome at the Whitehead Institute for Biomedical Research/MIT Center for Genome Research. 929_G_1 was shown, by PCR and Southern blot analysis, to contain the 3′ ends of both COL4A3 and COL4A4 and thus was chosen for use in the study. Long-range PCR amplification of YAC genomic DNA, using randomly chosen cDNA PCR primers in combination with ligation-mediated PCR, allowed identification of a total of 48 individual exons covering the COL4A4 cDNA sequence. All of these had recognizable 3′ and 5′ splice sites conforming to the GT-AG rule for intron-exon boundaries, except for the 5′ splice site of exon 21 (see table 1). The sizes of the coding exons range from 9 bp to 287 bp. The initiation methionine codon is in exon 2, and the signal peptide is encoded by exons 2 and 3. Exons 4–45 and the 5′ part of exon 46 code for the collagenous domain. The carboxyl NC1 domain is encoded by the 3′ end of exon 46, exon 47, and the coding part of exon 48. PCR amplification of the YAC DNA as well as various human control DNAs, using primers located in the 3′ untranslated region (5452F: 5′-ATGCTGCACAGATGGATTTG-3′ and 5569R: 5′-TGGCCTTTATCATCTACTTG-3′), consistently give a 137-bp fragment, which proves that this region is entirely encoded by a single exon (48). The exons, their sizes, the immediately flanking intronic sequences, and the approximate intron sizes are indicated in table 1. All sequences derived from YAC genomic DNA were confirmed on human genomic DNA, and no exonic deletions were found in the YAC.

As expected, comparison of the deduced COL4A4 gene structure, with respect to exon number and size, with the other previously characterized type IV collagen genes revealed a high degree of similarity to the X-linked COL4A6 gene, as well as to mouse col4a2, as is shown in figure 1. Twenty-five COL4A4 exons are the same size as the corresponding COL4A6 exons, and 19 of them are also the same size as the corresponding mouse col4a2 exons. As with COL4A5 and COL4A6, the first several exons (5–18) of the collagenous domain and three other exons (23, 25, and 34) begin with an intact glycine codon, whereas the remaining 25 exons encoding the collagenous domain start with the second nucleotide of a glycine codon.

SSCP Mutation Analysis

PCR primers designed for SSCP screening are indicated in table 2. SSCP screening of all the coding exons allowed the detection of multiple band shifts. In eight cases (exons 17, 21, 29, 33, 39, 42, 44, and 47), highly variable patterns were observed among patients and controls because of the presence of numerous polymorphisms. Novel primers were chosen, to exclude the polymorphic sequences within the PCR products of two exons (exons 42 and 44). Although the amended amplification products do not completely cover the relevant exons and their genomic flanking regions, they allowed the detection of two novel band shifts.

A total of 16 sequence variants were identified (table 3). Five of these result in premature termination codons or frameshifts and are potentially null mutations. Another is an 18-bp deletion in exon 20 and would be expected to result in the deletion of two Gly-X-Y repeats. Two other mutations affect splicing. One is a 46-bp deletion, which includes 17 bp of exon 15 and 29 bp from the 5′ part of intron 15; the second is an A→G substitution at position −23 in intron 45, which creates a potential acceptor splice site. The presence of a potential branch point at position −37 (GGTTTAAC) may allow this cryptic splice site to be used, which leads to a frameshifted transcript. This mutation was not found in any of the 48 unrelated controls screened. Thus, all of these mutations are expected to be pathogenic. Seven further nucleotide differences cause presumptive amino acid substitutions. One is a leucine for proline substitution in the carboxyterminal NC1 domain, P1572L. This change was not found in 48 control DNAs, involves a proline conserved among all of the human type IV α chains, and thus is probably pathogenic. All the other amino acid substitutions occur in the collagenous domain. Four are substitutions of nonglycine residue. Two of these, L1004P and P1402S, are obviously amino acid–sequence polymorphisms, since their frequencies are .46 and .68, respectively, in 90 normal chromosomes. The other two, E570Q and A931T, do not involve conserved amino acids. One, E570Q, is present in 1 of the 48 controls screened and therefore is likely not to be pathogenic. A931T was not found in the control population, but, as this group does not adequately represent the West Indian population from which the patient originates, its significance remains unclear. The last two variants are glycine substitutions, G545A and G1030V, within continuous stretches of Gly-X-Y repeats in the collagenous domain. The glycine-to-valine substitution was found in only one patient and in no control individuals. The alanine for glycine substitution, on the other hand, was found to be homozygous in two unrelated patients, both from consanguinous families, and was also found to be heterozygous in one patient (family AR12), in whom two nonsense mutations had already been found, as well as in numerous control individuals. In fact, 30 (11.5%) of 261 control individuals of various geographic origins display the G545A substitution in the heterozygous state. Interestingly, this number includes 15 (of 60) individuals originating from North Africa. In addition, in one North African individual without any obvious phenotype, G545A was present in the homo-

Table 1

COL4A4 Intron-Exon Boundaries

Exon Number	Exon Size (bp)	3′ Splice Site		5′ Splice Site		Intron Number	Approximate Intron Size[a] (kb)
1	178[b]			CGTCCCCACAG	gtaaggcgcccataggtggggcaaccgctg	1	5.4
2	172	tgttttaagtgatatatttatatctaacag	GAACAGACTTT	GGT CCC TG.	gtgagtatcctcagtgtgattgcctgcttt	2	2.6
3	43	aaattaatctgcgtttttgtttttctgtag	..G TCA CTT	GTA TAT GGG	gtaagtacattcactagctctggttgcgtt	3	4.2
4	78	gtatactaaataagatgcttcactttccag	AGT GGA AAG	GGG TCT CGG	gtaagtgatgtaaagaccaacatgggcagc	4	>12
5	135	cccccatttctttttaatcttccttctcag	GGT CCA CCA	GGA GAT AAG	gtaagccacatcttatgtggagattcacat	5	1
6	45	atctctttgttttatttttctgggattttag	GGT CCA ACT	GGC ATA CCT	gtaagtatccctaaattttctcccatttttt	6	1.2
7	117	atgcttcacttataattttgttttctgcag	GGG CAC CCA	GGC CCC CTA	gtaagatacccacagttcatcttaacagga	7	3.4
8	69	ctgttaagtacatacatattttatttcag	GGC CAT CCT	GGT ATT CAG	gtaagtattatgatcatgatgggttaaaaa	8	2.8
9	36	acttaaagcgatgcattcttatttttgcag	GGA GAC AGA	GGC TTA CCA	gtaagtctcctgagcaggctctatgcttct	9	2.5
10	63	aacagatgcactgatttttctcatgtttag	GGA TCT TGG	GGG TTA GTG	gtaagtacagcactgctgatgtgatccctt	10	.355[c]
11	36	tgatttgtgtttttttctcccttgatttag	GGA CCT CCG	GGT TTG AAG	gtaggcagtcatgagcaagatacatcaata	11	.210[c]
12	42	caattagattctttttcatttctttctcag	GGA AAT CCC	GGA GAC CCG	gtaagaattccctcttgctctgaggagcaa	12	4.7
13	81	aattttcattatcttgctttttttttttaag	GGT GAG GTT	GGA GAA AAG	gtaattcccaaatccatatgtcaccttgga	13	.7
14	54	agaccaaattatattcattgtaatttttag	GGT ATA AAA	GGA CGC AAG	gtagtttatttaagtacctttttttttttca	14	.3
15	60	aacccctctaaatgttgtcatctcttaaag	GGA GAA TCT	GGG CCT CGG	gtaagcatcatttgctgttaacatcagtga	15	.8
16	45	tgtttatcctgatttttgtgaatttattag	GGG GAT CCT	GGA TTA AAG	gtaaatgtacctcattatagagataataaa	16	.308[c]
17	54	tttttaaggcatgtttgtgcttgtttgtag	GGA GAA CTA	GGC CCA AAG	gtaaaaagtgagcctattttaatatatcct	17	1.8
18	70	tttaaaaagtgaatgaatgaatgatttcag	GGG GAT CCT	CTT AAA G..	gtttgtgttcctattcagttcctccataaa	18	.8
19	105	tttattctcatgtaccattatctttggcag	.GC CCA CCA	TGT GCA G..	gtacggatgccctgatgtgccctggctttt	19	4.3
20	165	tgtaattttcattcttgatttatcttgcag	.GC ATG ATA	TCA AGT G..	gtgaagtgcgccttattcatatccattagc	20	3.8
21	90	tctctgcccacccccttacctccttgacag	.TG ATA TAC	GAA AAA G..	gcaagtaggcaccaaaggtagaatttctag	21	.5
22	164	tgacaaatctgccattgatcctctgtctag	.GA AAT GAA	GGG CTA CCA	gtgagtacctcttaccagtcatctttccag	22	6.8
23	73	cctaatttgtttctcttatttttcaccaag	GGA AAG CAT	GTT AAA G..	gtgagcataatcaatcacttaactccctgc	23	1.7
24	107	gcttgaatatcttttgacctctatttccag	.GG CAC AAA	GGA CCT CCA	gtatgtagaattcttctcttgctatagaat	24	2.4
25	184	ggtaattgactcatgacataacattcccag	GGG GAT CAT	CAG AAA G..	gtaattttaaaaaaattttatttatttcaa	25	>12
26	69	gtgaatgccgcttttcttttctcattgcag	.GT GAC ACA	CCT CCA G..	gtattatcaatgagaagttactgtcttata	26	2.7

(*continued*)

Table 1 (continued)

Exon Number	Exon Size (bp)	3′ Splice Site		5′ Splice Site		Intron Number	Approximate Intron Size[a] (kb)
27	108	agtctgtgttttgtttttattattgtttag	.GT CCG AAG	CCA CCT G..	gtcagtgaggttttttttttttttttttt	27	.4
28	219	cattaaaaatttttttttgttgttaaatag	.GT TTT CGT	GCT GAA G..	gtatgagggctttgctgtccttttggtgca	28	1.6
29	162	gaatgtcatggatgtattgtatcatctcag	.GG CCA GCT	AGC CCA G..	gtaagtgtcaagaactgtgaagggtctggg	29	1.3
30	171	acacactgtggtcactgtggtttgtcttag	.GT GCT CCA	CCA AAG G..	gtaggtagtgctttttccatgatgtcctcc	30	.5
31	144	tcctaaaactttatgctctcatttcaccag	.GA CCC CGG	GCC AAA G..	gtaaataaaatacagccatgcaacatgcat	31	2
32	108	attttgttcttgcataattaatatccatag	.GA GCC ATA	GAA AGA G..	gtaaacacttaagagtttttaaccatatca	32	1.1
33	182	agagacctgtaacacttccaacttcatcag	.GT GAT AAA	GGT GAC CAG	gtaaggggctactgctttgtgggggaagag	33	.9
34	64	ttaaggccttttgtttcattaaactgtcag	GGT GAG CCA	CCT AAA G..	gtatggtttggagatacagaagccatctgt	34	2.4
35	75	gtaatgtcacatgcattttatccctttcag	.GA AAC AAA	TGT CCA G..	gtatagctgcatatttttcttactaaaaca	35	4.7
36	108	aactctgatgtttgatctttccttctttag	.GG CAT TTT	TGC CCA G..	gtaccttgaaaggaaatctggagacttcct	36	.8
37	108	ccatctgcaaaacttccctgtcatcctcag	.GT GAT CAC	ATA AAA G..	gtaagtcaaaatgaagcccatcccacaaga	37	10.8
38	72	tttgggtcacaggcctgtggtgtttttcag	.GT CCC TCC	GCT TCA G..	gtaagcacctaatcgattcgtaacagactg	38	1.2
39	129	tatagtattaattctgttttccccattaag	.GT TTG CAT	CCC CCA G..	gtaagctgtagcatgctttgccccttctttc	39	.092[c]
40	111	aggataacatgtgaacatgtctcttcttag	.GG AGT TCA	CCA AGA G..	gtatgagagaatcatgacaaaggatggcag	40	1.3
41	156	agaatatttttgtctcttctctgtggtcag	.GA GCA CCT	CAG AAA G..	gtatgaatgttgttccgttaagtaatatgt	41	3
42	108	cctcatttttatgttttgtttttttttaag	.GA CCA ATG	CTT CCA G..	gtgagcttccaacctacatacagcattttg	42	2
43	9	gcctcgtttgttgtctttggatccctctag	.GT CCC	AGA G..	gtaagctttgctataaaattggtaatttcc	43	3.7
44	126	ccattacacaagcggtgattccaatttcag	.GT GAA CCG	GGA CCA G..	gtatgagccgcatgctgattgttcattatt	44	9
45	117	tcaccaccagcatcataaacttttgtgcag	.GG TGC AAA	GGA CCA G..	gtagagtttcagtcactcctcttttctctg	45	1.5
46	189	gccttcaatttttttccttttgttcctccag	.GG CCT CCT	GAC CTT G..	gtatggatacaggatcgcccatctccattg	46	1.7
47	287	acaccagctgtctcttcttccccttcccaag	.GT CTG GCA	TTC CTG ATG	gtgagtcccgtgcccttgaaaaccctggct	47	.4
48	>264	agccctaattctctcttgttttctgggtag	CAC ACA GGA				

NOTE.—Sixty–base pair intronic sequences for each boundary have been deposited in European Molecular Biology Laboratory/GenBank.

[a] Intron sizes were determined from YAC DNA and may have been underestimated if internal deletions of the YAC have occurred.

[b] According to Momota et al. (1998).

[c] Exact intron size (obtained by sequencing).

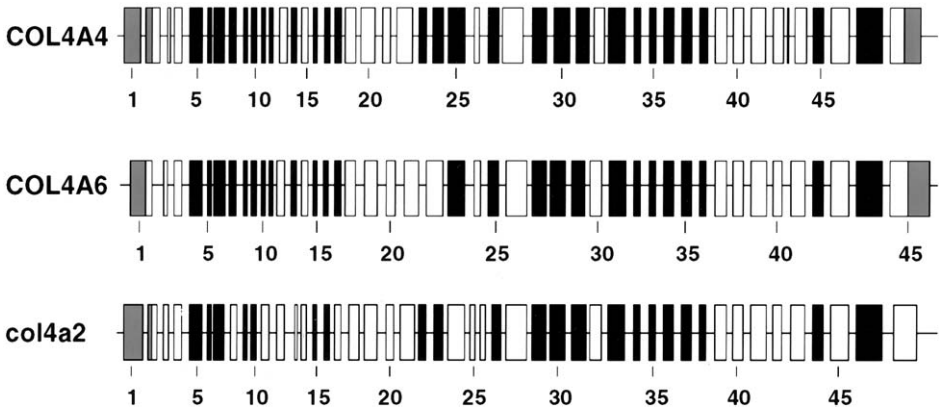

Figure 1 Comparison of exon sizes in the COL4A4 gene with the COL4A6 (Oohashi et al. 1995) and col4a2 (Buttice et al. 1989) genes. Exon numbers are indicated by the numbers below each gene. Exons of conserved size between COL4A4 and either col4a2 or COL4A6 are indicated in black, coding exons of differing sizes are indicated by open boxes. Noncoding regions are represented in gray.

zygous state. These data suggest that the G545A variant has little or no functional significance and that substitution of alanine for glycine at this position does not cause any severe pathogenesis.

In total, 10 mutations that are likely to be pathogenic were found in eight patients, with one nonsense mutation being found in two unrelated patients. In three patients originating from consanguineous marriages, the mutation, as expected, was homozygous. In three others, both heterozygous mutations were found, and in two cases, only one mutation was detected.

Segregation of the Mutations in the Families

In three families, one consanguineous (AR2), and in two families in which both mutations were found in the probands (AR18, AR23), the segregation of the mutation could be studied (fig. 2). In family AR2, six individuals are homozygous for the mutation (fig. 2*a*), and two patients each from families AR18 and AR23 (fig. 2*b* and *c,* respectively) share two different COL4A4 mutations, one inherited from each parent. All of these patients have features of severe AS, including deafness. Five of them, aged 22–34 years, already have end-stage renal disease or chronic renal failure, and the five others, all aged <28 years, present hematuria and proteinuria. In these families, various members have permanent microscopic hematuria. The individuals with hematuria all have normal renal function (even at age 90 years), except one (patient II.2 of family AR23), aged 54 years, whose chronic renal failure might be the result of lithium therapy. One individual (III.2 of family AR18) has both microscopic and episodes of macroscopic hematuria. A renal biopsy performed in individual III.7 of family AR23 at age 27 years displayed thin basement membranes. All hematuric individuals also display a COL4A4 mutation in the heterozygous state, but not all individuals carrying the same mutation seem to have microscopic hematuria; however, the latter group tends to be younger than the hematuric individuals.

Discussion

We report the complete intron-exon structure of the human COL4A4 gene, which encodes the α4 chain of type IV collagen. This gene had previously been mapped, head-to-head with the COL4A3 gene, to chromosome 2q34-37 (Mariyama et al. 1992), and cDNAs had been isolated covering the entire coding sequence (Leinonen et al. 1994), but its genomic structure had not been determined, with the exception of its 3′ end (Kamagata et al. 1992; Sugimoto et al. 1993; Mochizuki et al. 1994). Using a combination of long-range PCR and ligation-mediated PCR, we amplified and sequenced YACs covering the chromosomal region and searched for deviations from the published cDNA sequence that would indicate the presence of introns. By virtue of the sizes of overlapping long range PCR products, using YAC DNA as template, we can put the size of the COL4A4 gene at a minimum of 113 kb. This estimate does not take into account the sizes of introns 4 and 25, which must be >12 kb, since we could not amplify them. This large size is typical of the human type IV collagen genes. With reference to the published cDNA sequence (Leinonen et al. 1994), the COL4A4 gene was found to have 48 exons, with a 5′ UTR covering exon 1 and part of exon 2. Recently, Momota et al. (1998) further characterized the transcription start site, increasing the size of exon 1 to 178 bp, and described an alternative first exon (exon 1′) with rather ubiquitous and low expression and located upstream of exon 1. Contrary to a

Table 2

Oligonucleotide Primers and PCR Conditions Used for SSCP Analysis

Exon Number	Sense (No. of Intron Bases between Primer and Exon) (5′→3′)	Antisense (No. of Intron Bases between Primer and Exon) (5′→3′)	PCR Annealing Temperature (°C)	PCR Product Size (bp)
2	TCTGGAAGAGAAGACTGGCA[a] (32)[b]	(10) AAGCAGGCAATCACACTGA	54	153
3	TGTTTAAATTAATCTGCGTT (15)	(11) GCAACCAGAGCTAGTG	45	105
4	CGATGAGTACTGGTATACTA (22)	(12) ATGCTGCCCATGTTGGTCTT	50	152
5	ACCCCCATTTCTTTTTAATC (11)	(22) GGTGAGTCTTTCATGTGAAT	54	208
6	TCTCTTTGTTTTATTTTCTG (9)	(33) GATGAGTACTTCTGCCTTTT	47	127
7	TTTCGCAAAAATGCTTCACT (20)	(34) CCACAGGGCCTGTTCACTTA	50	211
8	TACTGAAATGGTAATACGCT (36)	(39) CATGGGCTTACCTATTTGGA	48	184
9	TGTGTGGACTTAAAGCGATG (17)	(3) TAGAGCCTGCTCAGGAGACT	53	96
10	TTGGGTAACAGATGCACTGA (16)	(10) AAGGGATCACATCAGCAGTG	55	129
11	TTGTGTTTTTTTCTCCCTTG (5)	(28) TTTCATTGTTCAGGGCTCTA	50	109
12	AGCCAGAAGTCTTAATTGCT (57)	(17) TCACCATTTGCTCCTCAGAG	54	156
13	GGGTGGAAACCTTCAAAACA (43)	(15) TACTTTCCAAGGTGACATAT	50	179
14	GGAGATGGAATTCAGTATGT (44)	(59) AAAGACCATGAGAAATAACA	53	197
15	CCCCTCTAAATGTTGTCATC (8)	(72) TTTGAGCTTGTGGGACTACT	54	180
16	AATGATGCACTGAGCTGGTT (49)	(66) GCACGCAACAGTACAACTTC	53	200
17	ATTTGTCACCCCGTCACTTT (34)	(73) GAATGATTCCTGGCAATACT	50	201
18	CCAGGCAACATGAGTAAAAT (39)	(6) TGGAGGAACTGAATAGGAAC	50	155
19	TGCACATACCATTTGTTTAT (25)	(5) CCAGGGCACATCAGGGCATC	50	175
20	TTCTTCTACAGAGACGTTT (44)	(11) TGCTAATGGATATGAATAAG	52	259
21	TATAGAAGACAGTCAGAAAA (42)	(9) TAGAAATTCTACCTTTGGTG	50	181
22	AAATATGACAAATCTGCCAT (15)	(8) GGAAAGATGACTGGTAAGAG	50	227
23	TGATCCATCACAATTAACCT (29)	(9) CAGGGAGTTAAGTGATTGAT	53	149
24	ACTTTACCCTCTGCTGATAA (32)	(48) GGGAAATAGTTGTTTGTATG	50	223
25	GACATTCAGTGGTTGGTAAT (24)	(33) TAAACACTTGTACCCCAAAG	57	280
26	TCAGTTATGTGAATGCCGCT (19)	(19) TGGGAAGTATATAAGACAGT	50	147
27	TGAGTCTGTGTTTTGTTTTT (12)	(5) AAAAAAAAAAACCTCAC	52	210
28	ATTGGTTCTATACTTGCACA (35)	(14) TCTATGCACCAAAAGGACAG	54	309
29	TGGGCCATCTGTATAGTTTT (37)	(30) TAATAGTAAGTAGGGTAAGC	57	269
30	GCCTTCACACACTGTGGTCA (16)	(13) ATGGGAGGACATCATGGAAA	55	240
31	TCCTAAAACTTTATGCTCTC (10)	(27) TCAAATACCAGAAACAAATG	53	221
32	CCTGTTCATTTTGTTCTTGC (17)	(22) TGTCAACTTATTTGATATGG	57	187
33	TTTCAGCAGAGACCTGTAAC (17)	(32) AAGAACAGAAAGGTTTTATT	52	271
34	GTTGTGCATGTGCCATTTGT (43)	(9) GATGGCTTCTGTATCTCC	50	154
35	TGAGACCAAATTAAATTGTC (34)	(80) TCATTGCCAGCTAGAAGTAA	52	210
36	CAAACGGCAACTCTGATGTT (18)	(17) AGTGCTCAGGAAGTCTCCAG	55	183
37	TATCTGGCCATCTGCAAAAC (17)	(8) TTGTGGGATGGGCTTCATTT	55	173
38	GCGTTTGTGGCTAGAGTGAG (49)	(27) GAACCATGGACTGAAGCTCAG	57	190
39	AGGCACTATAACAGGGACAAGA (40)	(45) GGAGTAACGTAAACCTTCCA	60	256
40	ACCTTCCAAATGCAATGAGG (27)	(4) CATCCTTTGTCATGATTCTCTC	53	184
41	TTTTTGTCTCTTCTCTGTGG (4)	(18) AGTTATTCACATATTACTTA	48	218
42	GCCCTCATTTTTATGTTTTG (12)	(29) ATTTTAGTAAATGTTTAGAC	50	190
		GTTGGAAGCTCACCTGGAAG[c]	54	153
43	GACTGGCCTCGTTTG (20)	(114) TTAATATCCTTACAGCACCC	50	180
44	ATTACACAAGCGGTGATTCC (6)	(39) AGAATTTCATTTCAGCAATA	50	214
		TGGCTCCTTCTGGTCCTCTC[c]	56	118
45	CACCAGCATCATAAACTT (8)	(23) AGGTTTACAGTGTCAGAGAA	53	186
46	AGTGCCAGAACAGAGGTGCT (61)	(7) GGAGATGGGCGATCCTGTA	57	297
47	ACACCAGCTGTCTCTTCTTC (10)	(19) TGAATGAGCCAGGGTTT	57	353
48	GTGTGTGTCTGAGCCCTAAT (20)	TGGTGAATTTCGCATTCT	50	322

[a] Exonic sequences are underlined.
[b] No. of exon bases between primer and beginning of translation.
[c] Novel primer designed to exclude polymorphic sequences within the PCR product.

report by Sugimoto et al. (1993), our data show that the entire 3′ untranslated region of COL4A4 lies in a single exon (48), as do those of the other type IV collagen genes. All of the exons have conventional GT-AG donor and acceptor splice sites, except exon 21, which has a donor site with GC instead of GT, confirmed by sequencing several human control DNAs. Although extremely rare, this 5′ splice site sequence has been de-

Table 3

Characteristics of Nucleotide Variants Detected in the COL4A4 Gene

Mutation or Variant	Nucleotide	Effect on Coding Sequence	Exon	Family	Mutation Status[a]	Presence in Control Individuals[a]
Pathogenic mutations						
Nonsense						
R1377X	C→T at 4337	Arg→Stop at 1377	44	AR12	h	ND[b]
				AR18	h	
C1641X	C→A at 5131	Cys→Stop at 1641	48	AR12	h	ND
Frameshift						
1013/6insA	Ins of A at 1013/6	Frameshift at Gly 270 Stop at 366	13	AR18	h	ND
2009delC	Del of C at 2009	Frameshift at Pro 601 Stop at 652	24	AR30	H	ND
3854/6delG	Del of G at 3854/6	Frameshift at Ser 1217 Stop at 1287	39	AR23	h	ND
In-frame deletion						
1527del18bp	Del of 18 bp at 1527	Del of Pro-Gly-Lys-Pro-Gly-Ala from Pro at 441	20	AR23	h	ND
Splicing						
1122del46bp	Del of 46 bp at 1122	5′ splice site	15[c]	AR2	H	ND
4542-23A→G	A→G at 4542-23	Novel splice site	46	AR4	H	0/48
Missense						
Mutations in the NC1 domain						
P1572L	C→T at 4923	Pro→Leu at 1572	47	AR25	h	0/48
Glycine substitution in the collagenous domain						
G1030V	G→T at 3297	Gly→Val at 1030	33	AR22	h	0/48
Variants of unknown significance						
Nonglycine substitutions in the collagenous domain						
E570Q	A→G at 1917	Glu→Gln at 570	24	AR19	h	1h/48
A931T	G→A at 2999	Ala→Thr at 931	31	AR7	H	0/48
Intronic variant						
2925-5A→T	A→T at 2925-5	3′ splice site	31	AR31	H	2h/48
Coding region polymorphisms						
Glycine substitution in the collagenous domain						
G545A	G→C at 1842	Gly→Ala at 545	23	AR6	H	30h/261
				AR9	H	1H/261
				AR12	h	
Nonglycine substitutions in the collagenous domain						
L1004P	T→C at 3219	Leu→Pro at 1004	33			11H/45
						19h/45
P1402S	C→T at 4414	Pro→Ser at 1402	44			19H/42
						19h/42

[a] H = homozygous change; h = heterozygous change.
[b] ND = not done.
[c] Deletes 17 bp of exon 14 and 29 bp in the 5′ part of intron 15, deleting the 5′ splice site.

scribed, elsewhere, for single exons in a few genes (Senapathy et al. 1990) and, in particular, for COL4A1 exon 34 (Soininen et al. 1989).

Current knowledge of the type IV collagen gene family indicates that these genes derive from a common ancestral gene (Zhou et al. 1994). By comparison of their peptide sequences, as predicted from their cDNA sequences, the α chains can be divided into two classes by virtue of their homology to either α1- or α2-like chains. The deduced α4 chain was shown to belong to the α2-like group (Leinonen et al. 1994) by virtue of homology at the amino acid level; this is further supported here by a high proportion of identical exon sizes and positions of split glycine residues at exon-intron boundaries, compared with human COL4A6 (Oohashi et al. 1995) and mouse col4a2 (Buttice et al. 1990).

Having characterized the COL4A4 gene structure, we performed mutation screening by SSCP in all COL4A4 coding exons of 31 patients with recessive AS, and we detected 10 novel mutations that are likely to be caus-

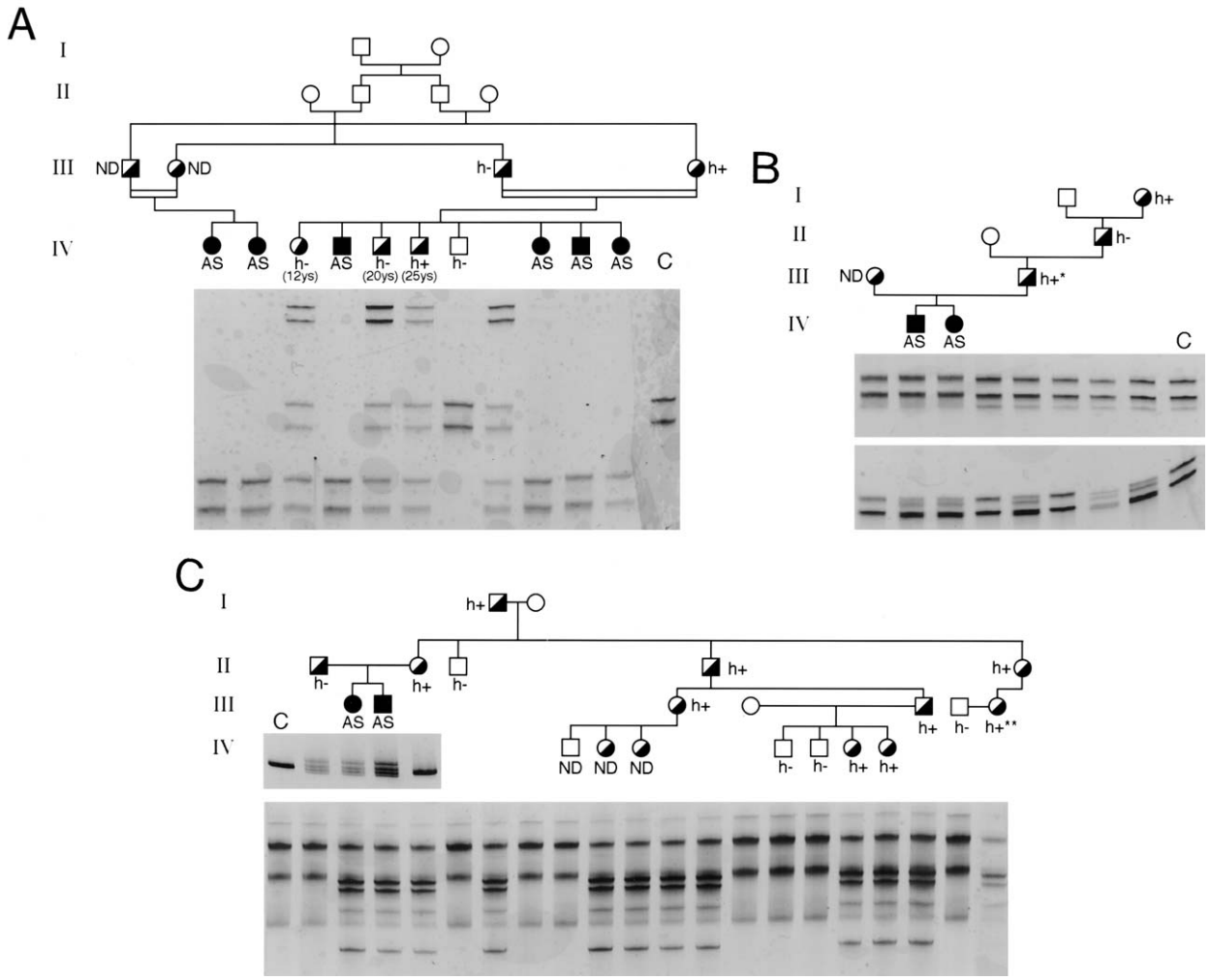

Figure 2 Mobility shifts observed in SSCP/heteroduplex analysis of DNA from members of three families. *a,* Family AR2, mutation 1122del46bp in exon 15. No clinical information or DNA samples were available for generations I and II. The uppermost bands, found only in heterozygotes, represent heteroduplex products. *b,* Family AR18, upper SSCP panel, mutation 1013/6insA in exon 13, present in the two siblings affected with autosomal recessive AS (IV.1 and IV.2) as well as in their mother; lower SSCP panel, mutation R1377X in exon 44, present in the two affected children and in the paternal branch. *c,* Family AR23, upper SSCP panel, mutation 3854delG in exon 39, present in the two children affected with autosomal recessive AS (III.1 and III.2) and in their father; lower SSCP panel, mutation 1527del18bp in exon 20, present in the two affected sibs as well as in their mother and several individuals from the maternal branch of the family. c indicates control DNA; h+, microscopic hematuria; h−, no hematuria; h+*, microscopic and episodes of macroscopic hematuria; h+**, microscopic hematuria, and thin basement membranes on renal biopsy; and ND, microscopic hematuria not determined.

ative and six different probably nonpathogenic amino acid substitutions. Only two mutations in COL4A4 have been described, to date, in autosomal recessive AS (Mochizuki et al. 1994). The pathogenic mutations were detected in eight unrelated families, including two nonconsanguinous families, in which only one heterozygous mutation was detected. This represents a relatively low mutation-detection rate compared with the 85% detection rate obtained in this laboratory using the identical method in the screening of the cystinosis gene (G. Jean, personal communication). It is possible that some COL4A4 mutations have been missed because of the presence of numerous polymorphisms that render the detection of additional bandshifts very difficult. Of course, mutations of the COL4A3 gene may also be responsible for the disease in a significant number of these patients. A further consideration is the possibility that autosomal recessive AS can exist as a "digenic" disease (i.e., with a COL4A4 mutation at one allele and a COL4A3 mutation at the other). This could account for the missing mutations in the two patients in whom only one mutation was found. This question cannot be addressed properly until the COL4A3 gene has been fully characterized and comprehensively screened in this and other patient cohorts.

Two mutations detected were glycine substitutions,

G545A and G1030V, which would be predicted to interrupt the Gly-X-Y repeat stretches in the collagenous domain. The G1030V mutation was not found in any controls and thus was considered likely to be pathogenic. In contrast, and rather suprisingly, we found the G545A variant in a patient with two nonsense mutations, as well as in a large number of control individuals, which excluded the possibility that it is responsible for the phenotype. Moreover, we found a control individual with this substitution in the homozygous state. This is the first description of a homozygous glycine substitution in a collagen chain that does not produce any obvious phenotype. The small size of the glycine residue, due to the absence of a side chain, is critical for the formation of a stable collagen triple helix. Although alanine is a bulkier and more hydrophobic amino acid, structurally, it is the most similar to glycine, and its substitution would thus be predicted to produce the most minimal disruption of the helix. This could explain why glycine-to-alanine substitutions in the collagenous domain are less frequent than other glycine substitutions, as is observed in severe forms of osteogenesis imperfecta (Byers and al. 1991) and in X-linked AS (3 of 69 glycine substitutions reported in the Human Gene Mutation Database Cardiff; Krawczak and Cooper 1997). Furthermore, type IV collagens contain numerous interruptions in the collagenous regions and thus might be likely to tolerate glycine substitutions at particular sites in the gene. The glycine at position 545 is located in a stretch of two Gly-X-Y triplets that lack prolines (in the X and Y positions) thought to be essential for thermostability (Westerhausen et al. 1990). This could suggest that local stability of this region is not critical to overall stability of the helix, so that disruption here does not significantly affect protein structure, folding, or function.

Given the frequency of this mutation in the population, it should be present in ~1/300 individuals in the homozygous state. One could therefore speculate that the latter group represents sporadic cases of microscopic hematuria, which seems to be relatively frequent in the general population (Cameron 1998) and would generally escape detailed molecular investigation.

In one family, the mutation 1527del18bp segregates with the presence of microscopic hematuria and thin basement membranes, typical features of familial benign hematuria. Although one COL4A4 mutation has already been described in a family with familial benign hematuria (Lemmink et al. 1996), this is the first report of a type IV collagen mutation in a family showing the complete clinical and histological features of FBH. Our results confirm that carriers of autosomal recessive AS may have FBH, as suggested by Lemmink et al. (1996).

In this study, the elucidation of the COL4A4 genomic structure enabled us to perform a comprehensive mutation screen of the gene in 31 patients with autosomal recessive AS and to detect 10 mutations in eight patients. All of these mutations are likely to be causative, as they can be predicted to alter the putative protein structure significantly. These data, which document potentially pathogenic COL4A4 mutations in AS and FBH and the presence of glycine substitutions in the general population, will contribute to genotype-phenotype correlation studies and, along with the possibility of the existence of compound heterozygotes with mutations in COL4A3 as well, will enhance arguments for including those syndromes in a more generalized framework of type IV collagen disorders.

Acknowledgments

We thank the patients, the families, and the physicians who have contributed to this project, and we gratefully acknowledge G. Albouze. This study was supported by the Fondation pour la Recherche Médicale, the Assistance Publique-Hôpitaux de Paris, the Mutuelle Générale de l'Education Nationale, the Association Claude Bernard, and the Association pour l'Utilisation du Rein Artificiel.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

European Molecular Biology Laboratory, http://www.ebi.ac.uk/ebi_docs/embl_db/ebi/topembl.html (for COL4A4 exon and flanking intronic sequences, accession numbers Y17396–Y17443)

GenBank, http://www.ncbi.nim.nih.gov/Web/Genbank (for COL4A4 exon and flanking intronic sequences, accession numbers Y17396–Y17443)

Human Gene Mutation Database, http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html (for glycine substitutions in collagen disorders)

National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/Unigene/Hs.Home.html (for YAC clones containing at least part of the COL4A4 gene)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for autosomal recessive AS [MIM 203780], dominantly inherited familial benign hematuria (FBH) (MIM 141200, and hereditary nephritis AS [MIM 104200])

Whitehead Institute for Biomedical Research/MIT Center for Genome Research, http://www-genome.wi.mit.edu (for YAC clones containing at least part of the COL4A4 gene)

References

Antignac C, Knebelmann B, Drouot L, Gros F, Deschenes G, Hors-Cayla MC, et al (1994) Deletions in the COL4A5 collagen gene in X-linked Alport syndrome: characterization of the pathological transcripts in nonrenal cells and correlation with disease expression. J Clin Invest 93:1195–1207

Atkin CL, Gregory MC, Border WA (1988) Alport syndrome.

In: Schrier RW, Gottschalk CW (eds) Diseases of the kidney. Little Brown, Boston, pp. 617–641

Barker DF, Hostikka SL, Zhou J, Chow LT, Oliphant AR, Gerken SC, et al (1990) Identification of mutations in the COL4A5 collagen gene in Alport syndrome. Science 248: 1224–1227

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

Byers PH, Wallis GA, Willing MC (1991) Osteogenesis imperfecta: translation of mutation to phenotype. J Med Genet 28:433–442

Cameron JS (1998) The patient with proteinuria and/or hematuria. In: Davison AM, Cameron JS, Grünfeld J-P, Kerr DNS, Ritz E, Winearls CG (eds) Oxford textbook of clinical nephrology. Oxford University Press, Oxford, 441–459

Cohen D, Choumakov I, Weissenbach J (1993) A first generation physical map of the human genome. Nature 366: 698–701

Ding J, Stitzel J, Berry P, Hawkins E, Kashtan CE (1995) Autosomal recessive Alport syndrome: mutation in the COL4A3 gene in a woman with Alport syndrome and post-transplant antiglomerular basement membrane nephritis. J Am Soc Nephrol 5:1714–1717

Gubler MC, Knebelmann B, Beziau A, Broyer M, Pirson Y, Haddoum F, Kleppel MM, et al (1995) Autosomal recessive Alport syndrome: immunohistochemical study of type IV collagen chain distribution. Kidney Int 47:1142–1147

Hudson BG, Reeders ST, 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

Jefferson JA, Lemmink HH, Hughes AE, Hill CM, Smeets HJM, Doherty CC, Maxwell AP (1997) Autosomal dominant Alport syndrome linked to the type IV collagen α3 and α4 genes (COL4A3 and COL4A4). Nephrol Dial Transplant 12:1595–1599

Kamagata Y, Mattei M-G, Ninomiya Y (1992) Isolation and sequencing of cDNAs and genomic DNAs encoding the α4 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

Kere J, Nagaraja R, Mumm S, Ciccodicola A, D'Urso M, Schlessinger D (1992) Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts. Genomics 14:241–248

Kleppel MM, Santi PA, Cameron JD, Wieslander J, Michael AF (1989) Human tissue distribution of novel basement membrane collagen. Am J Pathol 134:813–825

Knebelmann B, Forestier L, Drouot L, Quinones S, Chuet C, Benessy F, Saus J, et al (1995) Splice-mediated insertion of an Alu sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. Hum Mol Genet 4:675–679

Knebelmann B, Breillat C, Forestier L, Arrondel C, Jacassier D, Giatras I, Drouot L, et al (1996) Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport syndrome. Am J Hum Genet 59:1221–1232

Krawczak M, Cooper DN (1997) The Human Gene Mutation Database. Trends Genet 13:121–122

Leinonen A, Mariyama M, Mochizuki T, Tryggvason K, Reeders ST (1994) Complete primary structure of the human type IV collagen α4(IV) chain–comparison with structure and expression of the other α(IV) chains. J Biol Chem 269: 26172–26177

Lemmink HH, Mochizuki T, Vandenheuvel LPWJ, Schroder CH, Barrientos A, Monnens LAH, Van Oost BA, et al (1994) Mutations in the type IV collagen α3 (COL4A3) gene in autosomal recessive Alport syndrome. Hum Mol Genet 3: 1269–1273

Lemmink HH, Nillesen WN, Mochizuki T, Schroder CH, Brunner HG, Van Oost BA, Monnens LA, et al (1996) Benign familial hematuria due to mutation of the type IV collagen α4 gene. J Clin Invest 98:1114–1118

Lemmink HH, Schroder CH, Monnens LAH, Smeets HJM (1997) The clinical spectrum of type IV collagen mutations. Hum Mutat 9:477–499

Mariyama M, Zheng KG, Yangfeng TL, Reeders ST (1992) Colocalization of the genes for the α3(IV) and α4(IV) chains of type IV collagen to chromosome 2 bands-q35-q37. Genomics 13:809–813

Mochizuki T, Lemmink HH, Mariyama M, Antignac C, Gubler MC, Pirson Y, Verellen-Dumoulin C, et al (1994) Identification of mutations in the α3(IV) and α4(IV) collagen genes in autosomal recessive Alport syndrome. Nat Genet 8:77–82

Momota R, Sugimoto M, Oohashi T, Kigasawa K, Yoshioka H, Ninomiya Y (1998) Two genes, COL4A3 and COL4A4 coding for the human α3(IV) and α4(IV) collagen chains, are arranged head-to-head on chromosome 2q36. FEBS Lett 424:11–16

Ninomiya Y, Kagawa M, Iyama K, Naito I, Kishiro Y, Seyer JM, Sugimoto M, et al (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

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

Peissel B, Geng L, Kalluri R, Kashtan C, Rennke HG, Gallo GR, Yoshioka K, et al (1995) Comparative distribution of the α1(IV), α5(IV), and α6(IV) collagen chains in normal human adult and fetal tissues and in kidneys from X-linked Alport syndrome patients. J Clin Invest 96:1948–1957

Rozet JM, Gerber S, Perrault I, Camuzat A, Calvas P, Viegas-Pequignot E, Molina-Gomes D, et al (1996) Structure and physical mapping of DR1, a TATA-binding protein-associated phosphoprotein gene, to chromosome 1p22.1, and its exclusion in Stargadt disease (STGD). Genomics 36: 554–556

Saunier S, Calado J, Heilig R, Silbermann F, Benessy F, Morin G, Konrad M, et al (1997) A novel gene that encodes a protein with a putative src homology 3 domain is a candidate gene for familial juvenile nephronophthisis. Hum Mol Genet 6:2317–2323

Senapathy P, Shapiro MB, Harris NL (1990) Splice junctions,

branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol 183:252–278

Soininen R, Huotari M, Ganguly A, Prockop DJ, Tryggvason K (1989) Structural organization of the gene for the α1 chain of human type IV. J Biol Chem 264:13565–13571

Sugimoto M, Oohashi T, Yoshioka H, Matsuo N, Ninomiya Y (1993) cDNA isolation and partial gene structure of the human α4(IV) collagen chain. FEBS Lett 330:122–128

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

Westerhausen A, Kishi J, Prockop DJ (1990) Mutations that subsitute serine for glycine a1-598 and glycine α1-631 in type I procollagen. The effects on thermal unfolding of the triple helix are position-specific and demonstrate that the protein unfolds through a series of cooperative blocks. J Biol Chem 265:13995–14000

Zhou J, Ding M, Zhao Z, Reeders ST (1994) Complete primary structure of the sixth chain of human basement membrane collagen, α6(IV). Isolation of the cDNAs for α6(IV) and comparison with five other type IV collagen chains. J Biol Chem 269:13193–13199