Am. J. Hum. Genet. 42:309–314, 1988

The Genes Coding for Human Pro α1(IV) Collagen and Pro α2(IV) Collagen Are Both Located at the End of the Long Arm of Chromosome 13

Charles D. Boyd,* SuEllen Toth-Fejel,‡ Inder K. Gadi,* Michael Litt,‡ Michael R. Condon,† Margret Kolbe,* I. Kathleen Hagen,* Markku Kurkinen,† James W. Mackenzie,* and Ellen Magenis‡

Departments of *Surgery and †Medicine, UMDNJ–Robert Wood Johnson Medical School, New Brunswick, NJ; and ‡Department of Medical Genetics, Crippled Children's Division, The Oregon Health Sciences University, Portland

Summary

We have isolated and characterized a cDNA clone containing DNA sequences coding for the noncollagenous carboxy-terminal domain of human pro α2(IV) collagen. Using this cDNA clone in both Southern blot analysis of DNA isolated from human-mouse somatic-cell hybrids and in situ hybridization of normal human metaphase chromosomes, we have demonstrated that the gene coding for human pro α2(IV) collagen is located at 13q33→34, in the same position on chromosome 13 as the pro α1(IV) collagen gene.

Introduction

The genes coding for fibrillar collagens have previously been shown to be widely dispersed in the human genome (Huerre et al. 1982; Solomon et al. 1983; Strom et al. 1984; Emanuel et al. 1985; Huerre-Jeanpierre et al. 1986*b*). Recently we and others have shown that a nonfibrillar collagen gene, coding for a basement membrane component, pro α1(IV) collagen, is located at the distal end of the long arm of chromosome 13 (Pihlajaniemi et al. 1985; Solomon et al. 1985; Boyd et al. 1986; Emanuel et al. 1986). This chromosomal location is different from any previously reported for collagenous coding sequences. To investigate the possibility that other nonfibrillar collagen genes are also dispersed throughout the genome, we undertook the chromosomal location of the pro α2(IV) collagen gene, using a cDNA clone containing sequences coding for human pro α2(IV) collagen. The present paper reports results demonstrating that the gene coding for human pro α2(IV) collagen is located at the distal end of the long arm of chromosome 13, in the same position on 13q as the pro α1(IV) collagen gene (Boyd et al. 1986).

Received March 4, 1987; revision received July 16, 1987.

Address for correspondence and reprints: Dr. Charles D. Boyd, Department of Surgery, UMDNJ–Robert Wood Johnson Medical School, One Robert Wood Johnson Place, New Brunswick CN-19, NJ 08903.

© 1988 by The American Society of Human Genetics. All rights reserved.

0002-9297/88/4202-0013$02.00

Material and Methods

A. Recombinant DNA Clones

A previously constructed library of recombinant cDNA clones (Pihlajaniemi et al. 1985), made by using RNA isolated from a human tumor-cell line (HT-1080), was screened with a mouse pro α2(IV) collagen cDNA clone (Kurkinen et al. 1987). The 1.9-kbp insert in the mouse cDNA recombinant contained DNA sequences coding for the entire NCI domain of mouse pro α2(IV) collagen, 96 amino acids of carboxy-terminal Gly-X-Y sequences of pro α2(IV) collagen, and 950 nucleotides corresponding to the 3′-untranslated region of the mouse pro α2(IV) collagen mRNA. The HT-1080 cDNA library was screened as described elsewhere (Pihlajaniemi et al. 1985). Autoradiographically positive colonies were characterized by means of Southern blot and DNA sequence analysis. Southern blot analysis was performed according to a method described elsewhere

(Vandenplas et al. 1984). DNA sequencing was performed using either the chemical modification procedure of Maxam and Gilbert (1977) or the method of dideoxy-chain termination using single-stranded DNA derived from M13 subclones of insert DNA sequences (Sanger et al. 1977; Messing 1983).

B. Parental and Hybrid Cell Lines

A panel of mouse-human somatic-cell hybrids was provided by Mohandas et al. (1986). DNA from hybrid cell lines and from both human and mouse parental cell lines was analyzed by means of Southern blot analysis using ^{32}P-labeled human pro α2(IV) collagen cDNA sequences.

C. In Situ Hybridization

pBR322 recombinants, containing DNA sequences coding for human pro α2(IV) collagen, were radiolabeled, by means of either nick-translation or primer extension (Feinberg and Vogelstein 1983) using ^{3}Hd-CTP, to specific activities of 1.2–2.7 × 10^8 dpm/μg DNA. In situ hybridization to metaphase chromosomes was performed according to a method described elsewhere (Boyd et al. 1986).

Results

A. cDNA Analysis

Two cDNA recombinants were isolated from the HT-1080 cDNA library after screening 8 × 10^4 transformants with a mouse pro α2(IV) collagen cDNA clone. The insert DNA fragments within the *Pst*I site of the pBR322 vector were overlapping DNA sequences of 1.5 kbp and 1.7 kbp in length (HT-68 and HT-39, respectively). DNA sequence analysis of the 5′ end of the 1.5-kbp DNA insert in HT-68 revealed nucleotide sequences coding for polypeptide sequences homologous to a portion of the carboxy-terminal noncollagenous domain of human pro α2(IV) collagen and mouse pro α2(IV) collagen (fig. 1). Polypeptide sequences derived from HT-68 are identical to peptide sequences obtained from isolates of human pro α2(IV) collagen (R. Glanville, personal communication). Protein sequences derived from HT-68 were also identical to polypeptide sequences obtained from a similar coding region within a mouse pro α2(IV) collagen cDNA (Kurkinen et al. 1987). Further, comparison of derived amino acid sequences from the NCI domains of human pro α1(IV) collagen and pro α2(IV) collagen revealed extensive sequence homology, particularly within conserved sequence elements previously shown to be part of a tandem repeat within the noncollagenous carboxy-terminal domain of human pro α1(IV) collagen. These conserved sequence elements are shown in figure 1.

CCA	GGC	CGC	AGC	↓ GTC	AGC	ATC	GGC	TAC	CTC
Pro	Gly	Arg	Ser	Val	Ser	Ile	Gly	Tyr	Leu
CTG	GTG	AAG	CAC	AGC	CAG	ACG	GAC	CAG	GAG
Leu	Val	Lys	His	Ser	Gln	Thr	Asp	Gln	Glu
			—	—	—	—	Ile		
CCC	ATG	TGC	CCG	GTG	GGC	ATG	AAC	AAA	CTC
Pro	Met	Cys	Pro	Val	Gly	Met	Asn	Lys	Leu
—		—	—	Ser	—				—
TGG	AGT	GGA	TAC	AGC	CTG	CTG	TAC	TTC	GAG
Trp	Ser	Gly	Tyr	Ser	Leu	Leu	Tyr	Phe	Glu
		—	—	—					
GGC	CAG	GAG	AAG	GCG	CAC	AAC	CAG	GAC	CTG
Gly	Gln	Glu	Lys	Ala	His	Asn	Gln	Asp	Leu
						Gly	—		—
GGG	CTG	GCC	GGC	TCC	TGC	CTG	GCG	CGG	TTC
Gly	Leu	Ala	Gly	Ser	Cys	Leu	Ala	Arg	Phe
			—	—	—	—			—
AGC	ACC	ATG	CCC	TTC	CTG	TAC	TGC	AAC	CCT
Ser	Thr	Met	Pro	Phe	Leu	Tyr	Cys	Asn	Pro
			—	—			—		
GGT	GAT								
Gly	Asp.								

Figure 1 Nucleotide sequences and derived amino acid sequences from a human pro α2(IV) collagen cDNA. *First line,* Nucleotide sequences from a human pro α2(IV) collagen cDNA, HT-68. *Second line,* Derived amino acid sequences from nucleotide sequences of HT-68. Underlined amino acids correspond to identical amino acids found in cyanogen bromide fragments of placental human pro α2(IV) collagen. *Third lines,* Amino acid sequences derived from conserved regions within human pro α1(IV) collagen (Pihlajaniemi et al. 1985). Identical amino acids are indicated with a dash (—). ↓ Indicates the beginning of the NCI domain.

B. Somatic-Cell Hybrid Analysis

Table 1 illustrates the panel of mouse-human hybrid cell lines used in the present study. Southern blot analysis of DNA obtained from hybrid cell lines and from both mouse and human parental cell lines is presented in figure 2. A human-specific 18-kbp pro α2(IV) collagen genomic DNA fragment is evident in *Eco*RI digests of DNA isolated from either the human parental cell lines (1MR91) or circulating lymphocytes (fig. 2, lane 1). Mouse-specific pro α2(IV) collagen genomic DNA fragments are also evident,

Table 1

Chromosome Content of Hybrid Cell Lines

Cell Line	Chromosome Complement[a] (%)																							Scoring[b]
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	
CF84-2/8	82	73	...	82	64	100	91	91	...	...	9	73	...	100	100	73	100	91	73	100	...	100	...	...	−
CF84-4/8	83	83	92	92	...	100	83	83	...	75	...	...	100	100	92	...	92	92	92	100	...	...	...	...	+
CF84-5/8	...	90	80	...	40	20	...	100	...	...	...	...	20	90	100	...	100	...	50	...	90	80	60	90	+
CF84-7/8	...	...	80	67	...	67	7	20	...	...	53	27	...	87	...	...	87	...	...	47	...	...	...	...	−
CF84-20/8	...	50	64	57	14	...	93	43	...	...	36	50	93	93	7	43	100	93	...	36	93	43	57	...	+
CF84-21/5	...	...	...	...	...	...	82	...	...	85	...	82	...	97	...	...	100	...	...	97	100	...	...	18	−
CF84-25/8	...	...	...	62	92	85	...	62	...	54	85	77	...	...	...	...	100	...	15	...	100	...	...	...	−
CF84-27/8	...	83	...	...	83	100	...	75	...	...	83	92	...	...	100	...	100	100	...	75	92	...	...	...	−
CF84-34/9	...	...	...	33	...	100	100	58	...	83	92	...	...	100	...	...	92	...	25	58	...	42	...	...	−
CF84-35/8	...	...	60	80	...	90	90	100	...	...	...	70	...	80	80	...	100	100	...	...	...	...	...	...	−
Discordancy fraction	.27	.27	.27	.54	.40	.70	.60	.60	.27	.45	.60	.72	.0	.45	.50	.27	.72	.36	.30	.45	.36	.27	.09	.20	

[a] Percent of cells in each hybrid cell line that contain human chromosomes. Hybrid lines, in which human chromosomes were present in <20% of cells, were excluded from analysis; this is reflected in the calculation of the discordancy fraction. The "Scoring" column indicates the presence (+) or absence (−) of human-specific *Eco* DNA restriction fragments following Southern blot analysis with HT-68.

following *Eco*RI digestion of DNA isolated from mouse parental cell lines (B82; fig. 2, lane 2). These DNA fragments of 5.8, 3.2, 1.4, and 1.3 kbp are also present in *Eco*RI digests of DNA isolated from mouse-human hybrids (fig. 2, lanes 3–12). Only DNA isolated from hybrid cell lines that contained human chromosome 13 demonstrated the presence of an 18-kbp human pro α2(IV) collagen genomic DNA fragment (fig. 2, lanes 4, 5, and 7, and table 1).

C. In Situ Hybridization

Hybridization of radiolabeled HT-39 to normal metaphase chromosomes revealed a clear association of autoradiographic grains with the distal region of the long arm of chromosome 13 (13q33→34) (figs. 3, 4). The chromosomes from 100 metaphase cells were

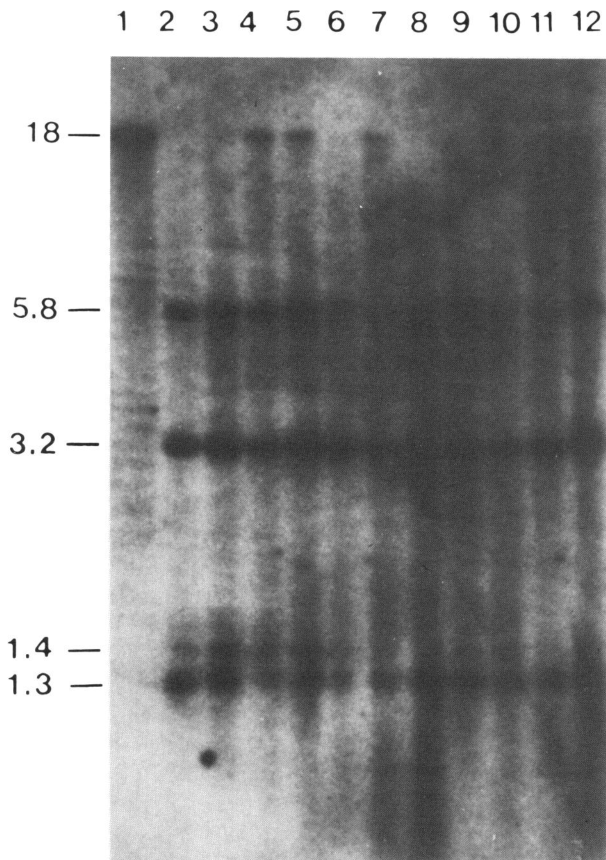

Figure 2 Southern blot analysis of DNA isolated from somatic-cell hybrid lines. DNA was isolated from human and mouse parental cell lines (1MR91 and B82 cell lines, respectively) in addition to the human-mouse hybrid cell lines indicated in table 1. DNA preparations were digested with *Eco*RI and analyzed by means of Southern blot hybridization using a radiolabeled human pro α2(IV) collagen cDNA, HT-68. Lane 1, 1MR91 human DNA. Lane 2, B82 mouse DNA. Lanes 3–12, DNA isolated from cell lines CF84 - 2/8 (lane 3), 4/8 (lane 4), 5/8 (lane 5), 7/8 (lane 6), 20/8 (lane 7), 21/5 (lane 8), 25/8 (lane 9), 27/8 (lane 10), 34/9 (lane 11), and 35/8 (lane 12). The sizes (in kbp) of the indicated restriction fragments were calculated on the basis of the comigration of radiolabeled *Hin*dIII fragments of lambda DNA.

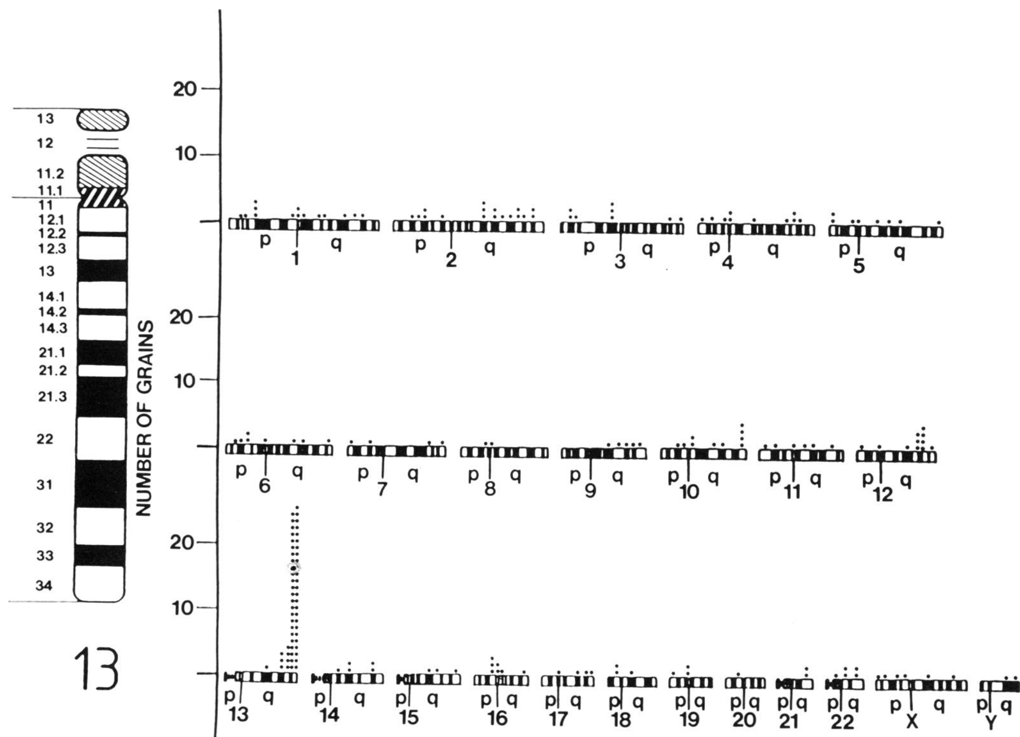

Figure 3 *Left,* Idiogram of chromosome 13 (ISCN 1985). *Right,* Grain distribution in 100 metaphase cells from a normal individual examined following in situ hybridization with a human pro α2(IV) collagen cDNA, HT-39. Note the significant number of grams over distal 13 (right). The number system and location of bands are indicated by the ideogram of chromosome 13 (left).

analyzed. Forty-one percent of these cells revealed hybridization to gene sequences within the q33→q34 banding region of chromosome 13. This is equivalent to 28% (57/203) of the total number of autoradiographic grains.

Boyd et al. (1986) have shown that this region of 13q contains the gene coding for pro α1(IV) collagen. To exclude the possibility that the results presented in figures 3 and 4 represent cross-hybridization to pro α1(IV) collagen gene sequences, a series of Southern blots were undertaken using human pro α1(IV) and human pro α2(IV) collagen cDNA clones (fig. 5).

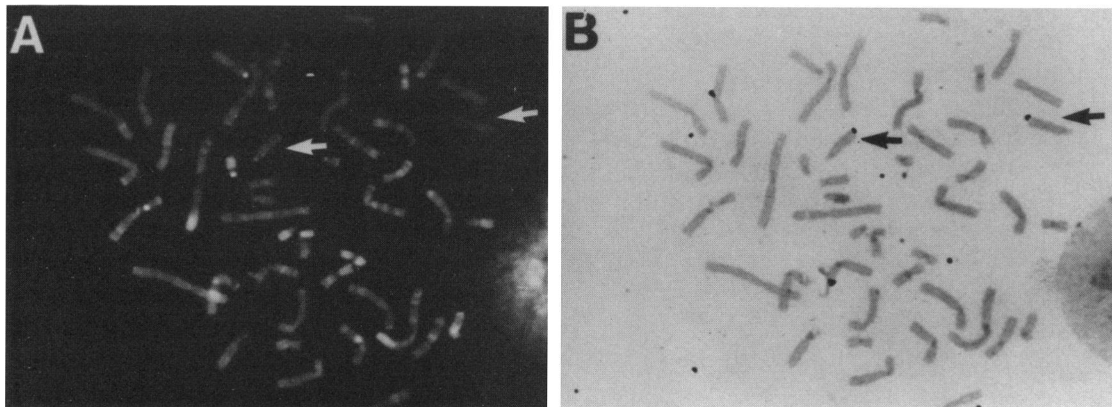

Figure 4 Metaphase chromosome spread of a cell from a normal individual, hybridized to a human pro α2(IV) collagen cDNA, HT-39. The spread is R-banded (*A*) for chromosome identification, destained, and restained with Wright's (*B*) to show location of grains. Arrows indicate grains associated with the end of 13q.

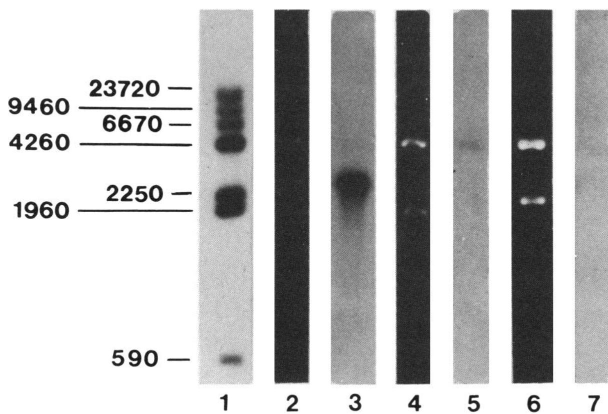

Figure 5 Southern blot analysis of restriction digests of human pro α2(IV) collagen cDNAs. Restriction digests of HT-68 and HT-39 were analyzed by means of Southern blot analysis using a radiolabeled *Pst*I, 2.6-kbp DNA fragment isolated from HT-21, a human pro α1(IV) collagen cDNA (Pihlajaniemi et al. 1985). Restriction digestion, radiolabeling, and Southern blot analysis were performed according to a method described elsewhere (Vandenplas et al. 1984). Lane 1, Radiolabeled *Hin*dIII digest of lambda DNA used as molecular-weight marker. The sizes (in bp) of individual DNA fragments are indicated. Lane 2, Ethidium bromide–stained DNA fragments of HT-21, following DNA digestion with *Pst*I. Lane 3, Autoradiography after hybridization of HT-21 DNA fragments, shown in lane 2, with a ^{32}P-labeled 2.6-kbp DNA fragment isolated from HT-21 after *Pst*I digestion. Previous characterization of this insert DNA fragment (Pihlajaniemi et al. 1985) has shown DNA sequences coding for human pro α1(IV) collagen, including the entire NCI domain. Lanes 4 and 6, Ethidium bromide–stained DNA fragments of HT-68 and HT-39, following digestion of DNA with *Pst*I. Lanes 5 and 7, Autoradiography after hybridization of HT-68 and HT-39 DNA fragments with the radiolabeled insert of HT-21.

HT-21 contains the DNA sequence coding for the entire NCI domain of human pro α1(IV) collagen; both the cDNA clones, HT-39 and HT-68, contain DNA sequences coding for the NCI domain of human pro α2(IV) collagen. In spite of considerable homology at the level of amino acid sequence within these domains of human pro α1(IV) and pro α2(IV) collagen, no significant nucleotide sequence homology is apparent following Southern blot analysis of the human pro α2(IV) collagen cDNA clones with HT-21 (fig. 5, lanes 3, 6, and 7). Therefore, hybridization of HT-39 to DNA sequences within the q33→q34 region of chromosome 13 specifically detects only homologous pro α2(IV) collagen gene sequences.

Discussion

The extensive homology between derived amino acid sequences from the cDNA clone HT-68 and mouse pro α2(IV) collagen cDNA clones, in addition to the homology of amino acid sequences obtained from human pro α2(IV) collagen peptide sequences, clearly demonstrate that HT-68 contains DNA sequences coding for human pro α2(IV) collagen. Further, Southern blot analysis of DNA isolated from a library of mouse-human somatic-cell hybrids and in situ hybridization using this human pro α2(IV) collagen cDNA clone have localized the gene coding for pro α2(IV) collagen to the distal end of the long arm of chromosome 13.

The syntenic nature of the genes coding for basement membrane collagens is significant for a number of reasons. The chromosomal localization of gene sequences coding for pro α1(IV) and pro α2(IV) collagen represents the first chromosomal assignments of nonfibrillar collagen genes. With the exception of the syntenic arrangement of the genes coding for pro α1(III) and pro α2(V) collagen (Emanuel et al. 1985; Huerre-Jeanpierre et al. 1986*a*) on chromosome 2, all collagenous coding sequences analyzed to date are widely dispersed in the genome. The location of basement membrane collagen genes to a chromosome different from those previously shown to contain Gly-X-Y coding sequences is consistent with the dispersed nature of the collagen multigene family. However, the syntenic arrangement of both pro α1(IV) and pro α2(IV) collagen genes contrasts with the overall chromosomal dispersion of Gly-X-Y coding sequences. It is significant to note that the synteny of the genes coding for the nonfibrillar heterotrimeric type IV collagen contrasts particularly with the dispersed chromosome location of the genes coding for the fibrillar heterotrimer, type I collagen (Huerre et al. 1982)

The functional significance of these findings remains speculative. We and others have suggested elsewhere that genomic dispersion of repetitive Gly-X-Y coding sequences is a necessary prerequisite to reducing the incidence of unequal crossover within these coding sequences during meiotic recombination. (Boedtker et al. 1985; Boyd et al. 1986). The presence of multiple introns within Gly-X-Y coding sequences and the dispersed chromosomal locations of these multiexon genes is consistent with this hypothesis. However, further detailed analysis of the structure of syntenic collagenous coding sequences will be necessary to determine whether this is a necessary prerequisite in the evolution of all Gly-X-Y coding sequences or is confined only to those fibrillar collagen genes in which alterations in the final size of

the procollagen protein affect triple-helix formation. It seems that aberrant triple-helix formation of nonfibrillar procollagens, a consequence of meiotic or somatic recombinations within the genes coding for these proteins, is less likely to produce the lethal phenotypes characteristic of changes within genes coding for fibrillar collagens (Prockop and Kivirikko 1984; Boedtker et al. 1985). Should this increased tolerance prove to be correct, the syntenic nature of the nonfibrillar basement membrane collagen genes may be of considerable significance both to considerations of the evolution of Gly-X-Y coding sequences and the role of basement membrane collagen in a variety of human disorders.

Acknowledgments

Since completion of this work, we have learned of similar results obtained by Drs. Jeanne Myers and Ellen Solomon (personal communications). This work was supported in part by National Institutes of Health grant AM22051 and the Gerlinger fund. We are grateful to Dr. Susan Deak for many constructive comments and to Camille Vanderzee for typing the manuscript.

References

Boedtker, H., M. Finer, and S. Aho. 1985. The structure of the chicken α2 collagen gene. Pp. 85–116 *in* R. Fleischmajer, B. R. Olsen, and K. Kuhn, eds. Biology, chemistry and pathology of collagen, Vol. **460**. New York Academy of Sciences, New York.

Boyd, C. D., K. Weliky, S. Toth-Fejel, S. B. Deak, A. M. Christiano, J. W. Mackenzie, L. J. Sandell, K. Tryggvason, and E. Magenis. 1986. The single copy gene coding for human α1(IV) procollagen is located at the terminal end of the long arm of chromosome 13. Hum. Genet. **74**:121–125.

Emanuel, B. S., L. A. Cannizzaro, J. M. Seyer, and J. C. Myers. 1985 Human α1(III) and α2(V) procollagen genes are located on the long arm of chromosome 2. Proc. Natl. Acad. Sci. USA **82**:3385–3389.

Emanuel, B. S., B. T. Sellinger, L. J. Gudas, and J. C. Myers. 1986. Localization of the human procollagen α1(IV) gene to chromosome 13q34 by in situ hybridization. Am. J. Hum. Genet. **38**:38–44.

Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.

Harnden, D. G., and H. P. Klinger, eds. 1985. ISCN (1985): an international system for human cytogenetic nomenclature. S. Karger, Basel.

Huerre, C., C. Junien, D. Weil, M.-L. Chu, M. Morabito, M. Van Cong, J. C. Myers, C. Foubert, M.-S. Gross, D. J. Prockop, A. Bove, J.-C. Kaplan, A. de la Chapette, and F. Ramirez. 1982. Human type I procollagen genes are located on different chromosomes. Proc. Natl. Acad. Sci. USA **79**:6627–6630.

Huerre-Jeanpierre, C., I. Henry, M. Bernard, P. Gallano, D. Weil, K.-H. Grzeschik, F. Ramirez, and C. Junien. 1986*a*. The pro α2(V) collagen (COL5A) maps to 2q14→2q32, syntenic to the pro α1(III) collagen locus (COL3A1) Hum. Genet. **73**:64–67.

Huerre-Jeanpierre, C., M.-G. Mattei, D. Weil, K. H. Grzeschik, M.-L. Chu, F. O. Sangiorgi, M. E. Sobel, F. Ramirez, and C. Junien. 1986*b*. Further evidence for the dispersion of the human fibrillar collagen genes. Am. J. Hum. Genet. **38**:26–37.

Kurkinen, M., M. R. Condon, B. Blumberg, D. Barlow, S. Quinones, J. Saus, and T. Pihlajaniemi. 1987. Extensive homology between the carboxyl-terminal peptides of mouse α1(IV) and α2(IV) collagen. J. Biol. Chem. **262**:8496–8499.

Maxam, A., and W. Gilbert. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.

Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.

Mohandas, T., C. Heinzmann, R. F. Sparks, J. Wasmuth, P. Edwards, and A. Lusis. 1986. Assignment of human 3-hydroxy-3 methyl glutaraldehyde co-enzyme A reductase gene to q13→q23 region of chromosome 5. Somatic Cell Mol. Genet. **12**:89–94.

Pihlajaniemi, T., K. Tryggvason, J. C. Myers, M. Kurkinen, R. Lebo, M.-C. Cheung, D. J. Prockop, and C. D. Boyd. 1985. cDNA clones coding for the pro α1(IV) chain of human type IV procollagen reveal an unusual homology of amino acid sequence in two halves of the carboxy-terminal domain. J. Biol. Chem. **260**:7681–7687.

Prockop, D. J., and K. I. Kivirikko. 1984. Heritable diseases of collagen. N. Engl. J. Med. **311**:376–386.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

Solomon, E., L. R. Hiorn, N. N. Spurr, M. Kurkinen, D. Barlow, B. L. M. Hogan, and R. Dalgleish. 1985. Chromosomal assignments of the genes coding for human type II, III and IV collagen: a dispersed gene family. Proc. Natl. Acad. Sci. USA **82**:3330–3334.

Solomon, E., L. Hiorne, R. Dalgleish, R. Tolstoshev, R. Crystal, and B. Sykes. 1983. Regional localization of the human α2(I) collagen gene on chromosome 7 by molecular hybridization. Cytogenet. Cell Genet. **35**:64–66.

Strom, C. M., R. L. Eddy, and T. B. Shows. 1984. Localization of human type II procollagen gene (COL2A1) to chromosome 12. Somatic Cell Mol. Genet. **10**:651–655.

Vandenplas, S., I. Wiid, A. Grobler-Rabie, K. Brebner, M. Ricketts, G. Wallis, A. Bester, C. Boyd, and C. Mathew. 1984. Blot hybridization analysis of genomic DNA. J. Med. Genet. **21**:164–172.