

# General Strategies for Isolating the Genes Encoding Type I Collagen and for Characterizing Mutations Which Produce Osteogenesis Imperfecta<sup>a</sup>

DANIEL H. COHN,<sup>c</sup> RICHARD J. WENSTRUP,<sup>f</sup>  
MARCIA C. WILLING,<sup>b,e</sup> JEFFREY F. BONADIO,<sup>g</sup>  
AND PETER H. BYERS<sup>b,c,d,e</sup>

*Departments of<sup>c</sup> Pathology and<sup>d</sup> Medicine  
and*

*<sup>e</sup> Center for Inherited Disease  
University of Washington  
Seattle, Washington 98195*

*<sup>f</sup> Department of Pediatrics  
Division of Cardiology  
and*

*Department of Medicine  
Division of Dermatology  
Duke University  
Durham, North Carolina 27710*

*<sup>g</sup> Howard Hughes Medical Institute  
University of Michigan  
Ann Arbor, Michigan 48109*

To understand the nature of structural defects that produce osteogenesis imperfecta (OI), the modes of inheritance of the several forms of the disease, and the effects of specific mutations on the behavior of the type I collagen molecule, we have developed a general strategy for analyzing the basic molecular defects which produce the phenotype. The genes encoding the two chains of the type I collagen molecule are structurally complex,<sup>1,2</sup> and the proteins are large and internally repetitive,<sup>3,4</sup> making analysis of mutations at either the DNA or protein level difficult. The  $\alpha 1(I)$  and  $\alpha 2(I)$  chains which make up the type I collagen heterotrimer are encoded by the 18 kilobase (kb)

<sup>a</sup>Original investigations supported in part by grants from the National Institutes of Health (AR 21557, GM 15253, AR 07713, AR 37484 and AR 38473) and a Clinical Research Grant (6-298) from the March of Dimes Birth Defects Foundation.

<sup>b</sup>Address for correspondence: Peter H. Byers, MD, Department of Pathology SM-30, University of Washington, Seattle, WA 98195.

COL1A1<sup>2,5</sup> and 38 kb COL1A2<sup>6</sup> genes, respectively. Many methods have been used in the analysis of mutations which produce osteogenesis imperfecta, including a variety of genomic cloning strategies,<sup>7-12</sup> cDNA cloning,<sup>13,14</sup> protein sequence determination,<sup>15,39</sup> mRNA sequence determination,<sup>7</sup> S1 nuclease cleavage,<sup>16</sup> RNase cleavage,<sup>17</sup> and R-loop analysis.<sup>18</sup> In general, each study is tailored to the biochemical phenotype of the type I collagen molecules produced by the cell strain. Our goal has been to develop general strategies for analyzing mutations so that the same methods can be applied to the determination of the structure of mutations in many cell strains.

We have developed strategies by which most or all of the genes of interest can be isolated. DNA sequence analysis by localization of a structural alteration in the type I collagen molecule determined at the protein level is used to characterize the mutation. There are several advantages to this approach. Because the DNA fragments isolated are large and carry most of the gene of interest, the basic strategy will be useful in analyzing a wide variety of mutations. The fragments contain polymorphic restriction sites, so alleles from heterozygotes can be distinguished after isolation. When the proper haplotypes are present, the parental origin of mutation can also be determined.

## THE COL1A2 GENE

Most of the exons of the 38 kb COL1A2 gene are contained within a 27 kb BamHI restriction fragment. Patient DNA is digested with BamHI and size-fractionated. The DNA is ligated into the unique BamHI site of cosmid pRK311<sup>19</sup> and packaged into lambda phage heads *in vitro*.<sup>20</sup> Recombinants carrying the COL1A2 gene are identified by colony hybridization.<sup>21</sup> Typically one positive clone per  $3-15 \times 10^3$  colonies is seen.

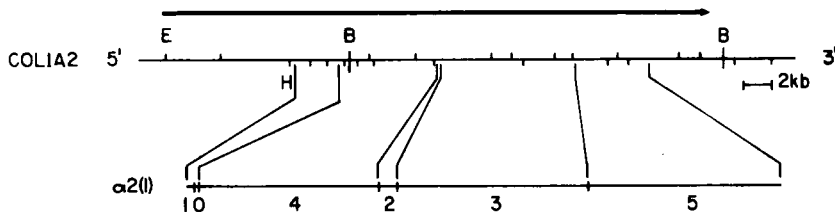
The isolated BamHI fragment contains 4 common polymorphic restriction sites<sup>22-25</sup> that have been used to distinguish alleles from individuals heterozygous at one of the sites. The frequencies of the 4 polymorphisms are such that heterozygosity is likely in most individuals. In families with dominantly inherited OI, linkage of a particular allele to the phenotype can also be determined.

From cDNA sequences,<sup>26,27</sup> genomic DNA sequences,<sup>6,28</sup> and genomic restriction maps,<sup>6</sup> we have developed a map that relates the linear arrangement of the protein (represented as cyanogen bromide (CB) cleavage fragments) to that of the exons in the gene (FIG. 1). The 27 kb BamHI fragment carries exons that encode the carboxyl-terminal 85% of  $\alpha 2(I)CB4$ , all of  $\alpha 2(I)CB3-5$ , and all of the C-propeptide.

We used the strategy to isolate alleles from an infant with OI type II carrying a 4.5 kb deletion within one COL1A2 allele.<sup>10</sup> The alleles were isolated, distinguished by restriction analysis, and heterozygosity at the polymorphic EcoRI site was used to establish that the deletion occurred in the paternally-derived allele. DNA sequence determination indicated that the deletion removed the 7 exons that encode amino acids 586-765 of the triple-helical domain. There was no homology between the endpoints of the deletion, suggesting a nonhomologous recombination mechanism. The endpoints of the deletion were located in introns, indicating that the protein product of the deletion allele should have an intact Gly-X-Y repeat. Even so, type I collagen molecules containing the short chain had excessive posttranslational modification (overmodification) amino-terminal to the deletion junction, were thermally unstable, and were poorly secreted, suggesting that elements of type I collagen primary structure beyond the Gly-X-Y repeat are important for triple-helix stability and mo-

lecular function. The characterization of this mutation also indicates that lethal mutations are not restricted to the COL1A1 gene.

We have also used this strategy to characterize a mutation which produces the mild to moderate OI type IV phenotype.<sup>11</sup> Linkage analysis using restriction fragment length polymorphic markers has demonstrated linkage of the phenotype to the COL1A2 gene in several families.<sup>23,29</sup> Biochemical defects including small deletions in the  $\alpha 2(I)$  chains of affected individuals<sup>30,31</sup> have supported the linkage data. We studied a small family in which the phenotype was inherited in an autosomal dominant fashion. Skin fibroblasts from affected individuals produced a population of normal type I collagen molecules and a population of molecules in which overmodification extended along the entire triple-helical region. Isoelectric focusing of peptides of type I collagen prepared by digestion with mammalian collagenase and cleavage with cyanogen bromide demonstrated a basic charge shift in a population of the  $\alpha 2(I)$ CB5B peptide fragments in affected family members, linking the phenotype to the COL1A2 gene. Analysis of COL1A2 restriction fragment length polymorphisms indicated that the phenotype associated with the RsaI(-) allele. Alleles were isolated from an affected



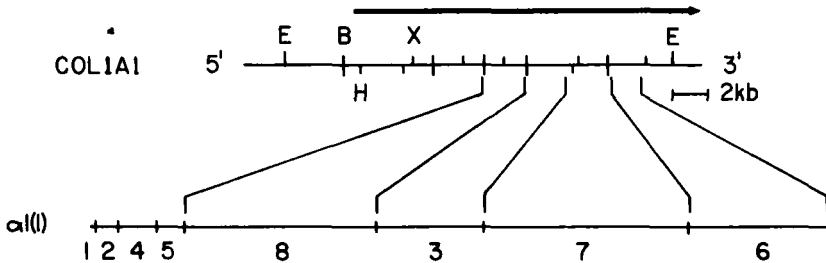
**FIGURE 1.** COL1A2 restriction map- $\alpha 2(I)$  cyanogen bromide cleavage map correlation. The *bold arrow* indicates the limits of the COL1A2 gene. Sites shown on the restriction map (*top*) are BamHI (B), EcoRI (E), and HindIII (H). Shown *below* the restriction map is a linear representation of the triple helix of the  $\alpha 2(I)$  chain. The locations of the methionine residues separating the cyanogen bromide peptides are indicated by *vertical lines* and the numerical designations of the peptides are shown. *Lines* between the restriction map and the protein map indicate the region of the gene encoding each peptide fragment.

individual, distinguished by restriction analysis, and the RsaI(-) allele was selected for DNA sequence analysis. The  $\alpha 2(I)$ CB5B peptide fragment is encoded by a 3.0 kb XhoI-XbaI restriction fragment which was subcloned into M13mp19. DNA sequence determination showed that the basic charge shift resulted from a single nucleotide change which substituted an arginine for the last glycine (residue 1012) of the triple helix. This result indicated that substitutions for glycine in the  $\alpha 2(I)$  chain can have milder clinical consequences than analogous defects in the  $\alpha 1(I)$  chain.

## THE COL1A1 GENE

The entire 18 kb COL1A1 gene except for the last 12 codons and the 3' untranslated region is contained within a ca. 23 kb EcoRI fragment. Using a similar strategy to that developed for the COL1A2 gene, we have isolated this fragment using cosmid

vector pLAFR1.<sup>32</sup> The fragment contains the only known intragenic polymorphic restriction site,<sup>33</sup> which allows alleles from heterozygotes to be distinguished. We used cDNA sequences,<sup>34</sup> genomic DNA sequences<sup>7-9</sup> (D. H. Cohn, unpublished; R. E. Gelinas, unpublished), and genomic restriction maps<sup>2,5,7</sup> to relate the COL1A1 gene structure to the  $\alpha 1(I)$  protein structure (FIG. 2). The genomic DNA sequence information known for the COL1A1 gene has facilitated efficient strategies for DNA sequence analysis. We used the genomic sequence information to design and synthesize exon-specific primers for some regions of the gene.<sup>40</sup> The primers are a series of 20mers which anneal to each intron near the intron-exon junctions. The primers are used in chain termination DNA sequence analysis to rapidly assay the exons encoding a particular region of the protein. As an example, for the 2 kb region of the gene encoding the 192 residues of  $\alpha 1(I)CB6$ , 9 primers have been used to determine the sequence of each exon.



**FIGURE 2.** COL1A1 restriction map- $\alpha 1(I)$  cyanogen bromide cleavage map correlation. The *bold arrow* indicates the limits of the COL1A1 gene. Sites shown on the restriction map (*top*) are BamHI (B), EcoRI (E), HindIII (H), and XhoI (X). Shown *below* the restriction map is a linear representation of the triple helix of the  $\alpha 1(I)$  chain. The locations of the methionine residues separating the cyanogen bromide peptides are indicated by *vertical lines* and the numerical designations of the peptides are shown. *Lines* between the restriction map and the protein map indicate the region of the gene encoding each peptide fragment.

## PERSPECTIVES

We have developed general strategies to isolate the genes encoding type I collagen for use in determining the structures of mutations which produce osteogenesis imperfecta. Characterization of mutations has allowed us to test hypotheses relating to the determinants of triple-helix stability, the distinct functions of the  $\alpha 1(I)$  and  $\alpha 2(I)$  chains, and the relationship between mutational structure and clinical phenotype.

While exceedingly useful, the methods described here cannot be applied to cell strains from all OI patients. Some cell strains do not show an obvious biochemical phenotype. In those that do, where mutations fall outside the identified fragment, alternative cloning strategies will be necessary. In sporadic cases of OI, assigning the mutation to COL1A1 or COL1A2 may not be easy and there may be cases where

the mutation does not reside in either gene. Although all of the point mutations that result in OI type II have been found in the COL1A1 gene, too few have been characterized so far to know whether this is a general finding. Similarly, linkage analysis and biochemical data indicate that mutations in the COL1A2 gene account for many cases of OI type IV, but there are individuals with the OI type IV phenotype with COL1A1 lesions (B. J. Starman and P. H. Byers, unpublished). There are examples of COL1A1 and COL1A2 lesions in individuals with OI type III.<sup>16,39</sup>

Several methods are available to make gene assignments for cells harboring apparent point mutations. At the protein level, isoelectric focusing<sup>35,36</sup> of cyanogen bromide peptide fragments of type I collagen for those mutations which produce a charge change in a type I collagen peptide fragment<sup>37,41</sup> is the best method. At the RNA level, cleavage with RNase at single base mismatches in mRNA-cDNA duplexes<sup>38</sup> should also be useful. The RNase cleavage method will also detect deletions too small to be seen at the protein level. Where there is a family history of the disease, linkage analysis combined with analysis of type I collagens produced by cells from affected family members can be used. While none of these methods alone can be used to determine the exact nature of the mutations, in combination with the gene isolation and DNA sequence determination methods outlined above, they can provide a powerful and rapid method for the systematic characterization of a number of interesting mutations.

As additional mutations are characterized, a relationship between the nature of the mutation and the clinical outcome is expected to emerge. The results should begin to dissect functional domains within the type I collagen molecule and allow us to derive new hypotheses regarding the roles this protein assumes in the extracellular matrix.

## ACKNOWLEDGMENTS

We thank Francesco Ramirez and Michael Bernard for providing DNA sequence information prior to publication.

## REFERENCES

1. YAMADA, Y., V. E. AVVIDEMENTO, M. MUDRY, H. OHKUBO, G. VOGELI, M. IRANI, I. PASTAN & B. DECROMBRUGGHE. 1980. The collagen gene: evidence for its evolutionary assembly by amplification of a DNA segment containing an exon of 54 bp. *Cell* **22**: 887-892.
2. CHU, M.-L., W. DE WET, M. BERNARD, J.-F. DING, M. MORABITO, J. MYERS, C. WILLIAMS & F. RAMIREZ. 1984. Human pro- $\alpha$ 1(I) collagen gene structure reveals evolutionary conservation of a pattern of introns and exons. *Nature* **310**: 337-340.
3. BORNSTEIN, P. & W. TRAUB. 1975. The chemistry and biology of collagen. *In* The Proteins. H. Neurath and R. Hill, Eds. Vol. 4: 411-632. Academic Press. New York, NY.
4. BORNSTEIN, P. & H. SAGE. 1980. Structurally distinct collagen types. *Annu. Rev. Biochem.* **49**: 957-1003.
5. BARSH, G. S., C. L. ROUSH & R. E. GELINAS. 1984. DNA and chromatin structure of the human  $\alpha$ 1(I) collagen gene. *J. Biol. Chem.* **259**: 14906-14913.
6. MYERS, J. C., L. A. DICKSON, W. J. DEWET, M. P. BERNARD, M.-L. CHU, M. DILIBERTO,

- G. PEPE, F. O. SANGIORGI & F. RAMIREZ. 1983. Analysis of the 3' end of the human pro- $\alpha$ 2(I) collagen gene: Utilization of multiple polyadenylation sites in cultured fibroblasts. *J. Biol. Chem.* **258**: 10128-10135.
7. COHN, D. H., P. H. BYERS, B. STEINMANN & R. E. GELINAS. 1986. Lethal osteogenesis imperfecta resulting from a single nucleotide change in one human pro- $\alpha$ 1(I) collagen allele. *Proc. Natl. Acad. Sci. USA* **83**: 6045-6047.
  8. BARSH, G. S., C. L. ROUSH, J. BONADIO, P. H. BYERS & R. E. GELINAS. 1985. Intron-mediated recombination may cause a deletion in an  $\alpha$ 1(I) type I collagen chain in a lethal form of osteogenesis imperfecta. *Proc. Natl. Acad. Sci.* **82**: 2870-2874.
  9. CHU, M.-L., V. GARGIULO, C. WILLIAMS & F. RAMIREZ. 1985. Multixon deletion in an osteogenesis imperfecta variant with increased type III collagen mRNA. *J. Biol. Chem.* **260**: 691-694.
  10. WILLING, M. C., D. H. COHN, B. J. STARMAN, K. A. HOLBROOK, C. R. GREENBERG & P. H. BYERS. 1988. Heterozygosity for a large deletion in the  $\alpha$ 2(I) collagen gene (COL1A2) has a dramatic effect on type I collagen secretion and produces perinatal lethal osteogenesis imperfecta. *J. Biol. Chem.* **263**: 8398-8404.
  11. WENSTRUP, R. J., D. H. COHN, T. COHEN & P. H. BYERS. 1988. Arginine for glycine substitution in the triple helical domain of the products of one  $\alpha$ 2(I) collagen allele (COL1A2) produces the osteogenesis imperfecta type IV phenotype. *J. Biol. Chem.* **263**: 7734-7740.
  12. PIHLAJANIEMI, T., L. A. DICKSON, F. M. POPE, V. R. KORHONEN, A. NICHOLLS, D. J. PROCKOP & J. C. MYERS. 1984. Osteogenesis imperfecta: cloning of a pro- $\alpha$ 2(I) collagen gene with a frameshift mutation. *J. Biol. Chem.* **259**: 12941-12944.
  13. VOGEL, B. E., R. R. MINOR, M. FREUND & D. J. PROCKOP. 1987. A point mutation in a type I procollagen gene converts glycine 748 of the  $\alpha$ 1 chain to cysteine and destabilizes the triple helix in a lethal variant of osteogenesis imperfecta. *J. Biol. Chem.* **262**: 14737-14744.
  14. CONSTANTINOU, C. D., K. B. NIELSEN & D. J. PROCKOP. 1988. The molecular defect in a lethal variant of osteogenesis imperfecta is a single base mutation that substitutes cysteine for glycine 904 of the  $\alpha$ 1(I) chain of type I procollagen. Submitted.
  15. BATEMAN, J. F., D. CHAN, I. D. WALKER, J. G. ROGERS & W. G. COLE. 1987. Lethal perinatal osteogenesis imperfecta due to substitution of arginine for glycine at residue 391 of the  $\alpha$ 1(I) chain of type I collagen. *J. Biol. Chem.* **262**: 7021-7027.
  16. DICKSON, L. A., T. PIHLAJANIEMI, S. DEAK, F. M. POPE, A. NICHOLLS, D. J. PROCKOP & J. C. MYERS. 1984. Nuclease S<sub>1</sub> mapping of a homozygous mutation in the carboxy-propeptide coding region of the pro- $\alpha$ 2(I) collagen gene in a patient with osteogenesis imperfecta. *Proc. Natl. Acad. Sci. USA* **81**: 4524-4528.
  17. GENOVESE, C. & D. ROWE. 1986. Analysis of cytoplasmic and nuclear messenger RNA in fibroblasts from patients with type I osteogenesis imperfecta. *Methods Enzymol.* **145**: 223-235.
  18. DEWET, W., M. SIPPOLA, M. BERNARD, D. PROCKOP, M.-L. CHU & F. RAMIREZ. 1985. Electronmicroscopic localization of deletions in the human pro- $\alpha$ 2(I) collagen gene. *Ann. N.Y. Acad. Sci.* **460**: 415-418.
  19. DITTA, G., T. SCHMIDHAUSER, E. YAKOBSON, P. LU, X.-W. LIANG, D. R. FINLAY, D. GUINEY & D. R. HELINSKI. 1985. Plasmids related to the broad host range vector, pRK290, useful for gene cloning and for monitoring gene expression. *Plasmid* **13**: 149-153.
  20. ENQUIST, C. & N. STERNBERG. 1979. *In vitro* packaging of Dam vectors and their use in cloning DNA fragments. *Methods Enzymol.* **68**: 281-298.
  21. GRUNSTEIN, M. & D. HOGNESS. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* **72**: 3961-3965.
  22. FALK, C. T., R. C. SCHWARTZ, F. RAMIREZ & P. TSIPOURAS. 1986. Use of molecular haplotypes specific for the human pro- $\alpha$ 2(I) collagen gene in linkage analysis of the mild autosomal dominant form of osteogenesis imperfecta. *Am. J. Hum. Genet.* **38**: 269-279.
  23. GROBLER-RABIE, A. F., D. K. BREBNER, S. VANDENPLAS, G. WALLIS, R. DALGLEISH, R. E. KAUFMAN, A. J. BESTER, C. G. MATHEW & C. D. BOYD. 1985. Polymorphism of DNA sequence in the human pro- $\alpha$ 2(I) collagen gene. *J. Med. Genetics* **22**: 182-186.

24. GROBLER-RABIE, A. F., G. WALLIS, D. K. BREBNER, P. BEIGHTON, A. J. BESTER & C. G. MATHEW. 1985. Detection of a high frequency RsaI polymorphism in the human pro- $\alpha$ 2(I) collagen gene which is linked to an autosomal dominant form of osteogenesis imperfecta. *EMBO J.* **4**: 1745-1748.
25. TSIPOURAS, P., J. C. MYERS & F. RAMIREZ. 1983. Restriction fragment length polymorphism associated with the pro- $\alpha$ 2(I) gene of human type I procollagen. *J. Clin. Invest.* **72**: 1262-1267.
26. BERNARD, M. P., J. C. MYERS, M.-L. CHU, F. RAMIREZ, E. EIKENBERRY & D. J. PROCKOP. 1983. Structure of a cDNA for the pro- $\alpha$ 2 chain of human type I procollagen. Comparison with chick cDNA for pro- $\alpha$ 2(I) identifies structurally conserved features of the protein and the gene. *Biochemistry* **22**: 1139-1145.
27. DEWET, W., M. BERNARD, V. BENSON-CHANDA, M.-L. CHU, L. DICKSON, D. WEIL & F. RAMIREZ. 1987. Organization of the human pro- $\alpha$ 2(I) collagen gene. *J. Biol. Chem.* **262**: 16032-16036.
28. DICKSON, L. A., W. J. DEWET, M. DILIBERTO, D. WEIL & F. RAMIREZ. 1985. Analysis of the promoter region and the N-propeptide domain of the human pro- $\alpha$ 2(I) collagen gene. *Nucl. Acids Res.* **13**: 3427-3438.
29. TSIPOURAS, P., J. C. MYERS, F. RAMIREZ & D. J. PROCKOP. 1983. Restriction fragment length polymorphism associated with the pro- $\alpha$ 2(I) gene of human type I procollagen. Application to a family with an autosomal dominant form of osteogenesis imperfecta. *J. Clin. Invest.* **72**: 1261-1267.
30. BYERS, P. H., J. R. SHAPIRO, D. W. ROWE, K. E. DAVID & K. A. HOLBROOK. 1983. Abnormal  $\alpha$ 2-chain in type I collagen from a patient with a form of osteogenesis imperfecta. *J. Clin. Invest.* **71**: 689-697.
31. WENSTRUP, R. J., P. TSIPOURAS & P. H. BYERS. 1986. Osteogenesis imperfecta type IV. Biochemical confirmation of genetic linkage to the pro- $\alpha$ 2(I) gene of type I collagen. *J. Clin. Invest.* **78**: 1449-1455.
32. FRIEDMAN, A. M., S. R. LONG, S. E. BROWN, W. J. BUIKEMA & F. M. AUSUBEL. 1982. Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. *Gene* **18**: 289-296.
33. SYKES, B., D. OGILVIE, P. WORDSWORTH, J. ANDERSON & N. JONES. 1986. Osteogenesis imperfecta is linked to both type I collagen structural genes. *Lancet* **1**: 69-72.
34. BERNARD, M. P., M.-L. CHU, J. C. MYERS, F. RAMIREZ, E. EIKENBERRY & D. J. PROCKOP. 1983. Nucleotide sequences of complementary deoxyribonucleic acids for the pro- $\alpha$ 1 chain of human type I procollagen. Statistical evaluation of structures that are conserved during evolution. *Biochemistry* **22**: 5213-5223.
35. BENYA, P. 1981. Two-dimensional CNBr peptide patterns of collagen types I, II and III. *Coll. Rel. Res.* **1**: 17-26.
36. COLE, W. & D. CHAN. 1981. Analysis of the heterogeneity of human collagens by two-dimensional polyacrylamide-gel electrophoresis. *Biochem. J.* **197**: 377-393.
37. BATEMAN, J. F., T. MASCARA, D. CHAN & W. G. COLE. 1987. A structural mutation of the collagen  $\alpha$ 1(I)CB7 peptide in lethal perinatal osteogenesis imperfecta. *J. Biol. Chem.* **262**: 4445-4451.
38. MYERS, R. M., Z. LARIN & T. MANIATIS. 1985. Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**: 1242-1246.
39. STARMAN, B. J., D. H. COHN, L. WEISS & P. H. BYERS. In preparation.
40. COHN, D. H. & P. H. BYERS. In preparation.
41. COHN, D. H., B. J. STARMAN, BLUMBERG & P. H. BYERS. Detection of germinal mosaicism for a mutation in a type I collagen gene (COL1A1) indicates that a small number of cells is irrevocably allocated to the germline early in human development. Submitted.