Am. J. Hum. Genet. 46:293–307, 1990

Consistent Linkage of Dominantly Inherited Osteogenesis Imperfecta to the Type I Collagen Loci: COL1A1 and COL1A2

Bryan Sykes,* Donald Ogilvie,* Paul Wordsworth,* Gillian Wallis,† Chris Mathew,‖‖ Peter Beighton,† Alan Nicholls,‡ F. Michael Pope,‡ Elizabeth Thompson,§ Petros Tsipouras,‖ Robin Schwartz,‖ Olafur Jensson,# Alfred Arnason,# Anne-Lise Børresen,** Arvid Heiberg,†† Daniel Frey,‡‡ and Beat Steinmann§§

*Collagen Genetics Group, University of Oxford, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford; ‡Medical Research Council Clinical Research Centre, Harrow, Middlesex, England; §Institute of Child Health, London; ‖Department of Pediatrics, University of Connecticut Health Center, School of Medicine, Farmington, CT; #Genetics Division, The Blood Bank, National Hospital, Reykjavík; **Genetics Department, Institute of Cancer Research, Norwegian Radium Hospital, Oslo; ††Frambu Health Centre, Siggerud, Norway; ‡‡Institut für Medizinische Genetik and §§Department of Paediatrics, University of Zurich, Zurich; and ‖‖MRC Centre for Molecular and Cellular Biology, University of Stellenbosch Medical School, Tygerberg, South Africa

Summary

The segregation of COL1A1 and COL1A2, the two genes which encode the chains of type I collagen, was analyzed in 38 dominant osteogenesis imperfecta (OI) pedigrees by using polymorphic markers within or close to the genes. This was done in order to estimate the consistency of linkage of OI genes to these two loci. None of the 38 pedigrees showed evidence of recombination between the OI gene and *both* collagen loci, suggesting that the frequency of unlinked loci in the population must be low. From these results, approximate 95% confidence limits for the proportion of families linked to the type I collagen genes can be set between .91 and 1.00. This is high enough to base prenatal diagnosis of dominantly inherited OI on linkage to these genes even in families which are too small for the linkage to be independently confirmed to high levels of significance. When phenotypic features were compared with the concordant collagen locus, all eight pedigrees with Sillence OI type IV segregated with COL1A2. On the other hand, Sillence OI type I segregated with both COL1A1 (17 pedigrees) and COL1A2 (7 pedigrees). The concordant locus was uncertain in the remaining six OI type I pedigrees. Of several other features, the presence or absence of presenile hearing loss was the best predictor of the mutant locus in OI type I families, with 13 of the 17 COL1A1 segregants and none of the 7 COL1A2 segregants showing this feature.

Introduction

Osteogenesis imperfecta (OI) is one of the heritable disorders affecting the skeleton and other connective tissues. The diagnostic hallmark is bone fragility, although this definition embraces a very wide range of phenotype, from, at one end, intrauterine or perinatally lethal forms through an almost continuous spectrum to, at the other end, a barely noticeable increase in fracture tendency. When the phenotype allows it, OI is almost always inherited as an autosomal dominant trait.

Bone and the other tissues most affected in OI depend on type I collagen for mechanical strength, so abnormalities in this protein, or in the genes that control it, have been suspected in the disorder for some time (McKusick 1972; Penttinen et al. 1975; Sykes et al. 1977). Collagen 1 is a trimer of two α1 chains and one α2 chain. These subunits are encoded, in man, at the nonsyntenic loci COL1A1 (17q 21.31–22.15) and COL1A2 (7q 21.3–22.1), respectively. Biochemical analysis of isolated cases has shown new dominant muta-

Received January 25, 1989; final revision received September 22, 1989.

Address for correspondence and reprints outside the United States: Dr. Bryan Sykes, University of Oxford, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU; England. Address for reprints within the United States: Dr. Petros Tsipouras, Department of Pediatrics, University of Connecticut Health Center, Farmington, CT 06032.

0002-9297/90/4602-0010$02.00

tions in these genes to be a frequent cause of perinatal lethal OI (Byers et al. 1988), but the association between the inherited forms of the disease and the collagen loci has come from the genetic analysis of extensive pedigrees.

Concordant segregation of the dominantly inherited OI phenotype and a collagen structural gene was first reported in 1983 (Tsipouras et al. 1983) when no recombinants between the disease and COL1A2 were found in eight informative meioses within a single family when an *Eco*RI RFLP was used at the locus. As additional markers at COL1A2 were discovered and more families were tested, discordant as well as concordant families emerged (Tsipouras et al. 1984; Grobler-Rabie et al. 1985*b*; Wallis et al. 1986). When markers at COL1A1 were tested in addition to those at COL1A2, there were no examples of discordance at both loci in 11 families (Sykes et al. 1986). This indicated that the extent of genetic locus heterogeneity might be limited to these two collagen genes. This in turn suggested that the widespread prenatal diagnosis of OI on the basis of identification of the mutant type I collagen allele by using genetic markers might be feasible in families which were too small in themselves to provide sufficiently good independent evidence of linkage.

The study which the present article reports was organized to address two questions of considerable practical importance to prenatal diagnosis. First, what is the frequency of OI loci linked to type I collagen genes? The 11-family study was clearly too limited to provide an accurate estimate of this frequency, the derived 95% confidence intervals extending from .71 to 1.00. Second, is there an association between the clinical phenotype and the mutant locus? A correspondence between COL1A2 and Sillence OI type IV had been suggested by earlier studies (Falk et al. 1986; Sykes et al. 1986), and any method of predicting the mutant locus in a small family in which segregation analysis might not provide a clear answer deserved investigation. Plainly, the only sensible way to answer these important questions was to coordinate the analyses of several centers with access to extensive OI families.

Methods

The seven groups taking part in this study analyzed a total of 508 individuals from 38 pedigrees. In all of these OI was segregating as a dominant trait, and any individuals in which the diagnosis was in doubt were excluded. Parameters of the phenotype were recorded and summarized as the most severe expression of the gene within each pedigree. In addition, collaborators were asked to assign the disease in each pedigree to one of the categories suggested by Sillence for dominantly inherited OI (Sillence et al. 1979).

DNA preparation, restriction-enzyme digestion, electrophoresis, blotting, and hybridization schedules were not standardized, and each laboratory used its own version of conventional techniques. The filters were hybridized to [^{32}P]-labeled probes which revealed fragment-length variation at the two type I collagen loci. The map positions of the markers are shown in figure 1, and information on probes and fragment sizes is given in table 1. Haplotypes were constructed according to the notation in table 2.

Incorporated into the protocol of the study was the stipulation that the maternity/paternity of apparently recombinant meioses should be checked. This was done by hybridizing *Hin*fI digests with the minisatellite probe 15.1.11.4, which detects a highly polymorphic series of bands in an individual which are all derived from the natural parents (Jeffreys et al. 1985). In the end only one group analyzed all recombinant meioses in this way, the others restricting their checks to crucial recombinants only. It was also agreed that if, after confirming paternity/maternity, any pedigrees were discordant at *both* COL1A1 and COL1A2, the appropriate branches should be resampled. This condition, unlike the first, was rigidly enforced. These two conditions reflected the anticipated significance to the overall study of doubly discordant pedigrees, which would, of course, indicate the presence in the population of a third, unlinked, OI locus.

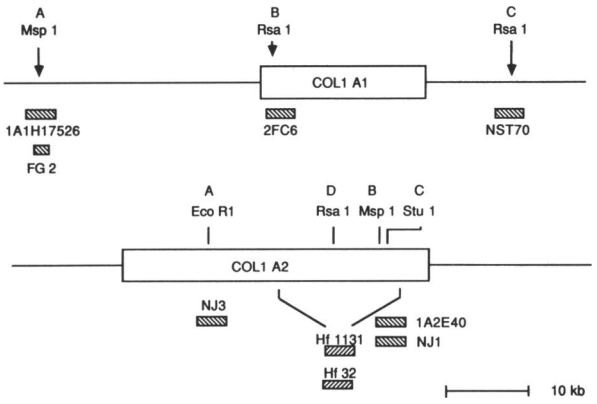


Figure 1 Physical map of dimorphic restriction sites at COL1A1 and COL1A2, showing the probes used to reveal them.

Table 1

Information on Restriction-Fragment Sizes Revealed by Each Probe

Collagen Locus, Enzyme Notation, and Enzyme	Probe	Fragment Size (kb)		Reference
		Allele 1	Allele 2	
COL1A1:				
A, *Msp*I	1A1H17526	2.7	3.0	Sykes et al. 1986
	FG2[a]	2.7	3.0	Gilchrist 1987
B, *Rsa*I	2FC6	2.6	3.6	Sykes et al. 1986
C, *Rsa*I	NST70	2.1	2.4	Ogilvie et al. 1987
COL1A2:				
A, *Eco*RI	NJ3	10.5	14.0	Tsipouras et al. 1983
B, *Msp*I	Hf 32	1.6	2.1	Tsipouras et al. 1984
	1A2E40	1.6	2.1	Grobler-Rabie et al. 1985*a*
C, *Stu*I	NJ1	5.9+4.3	10.2	Børressen et al. 1985
D, *Rsa*I	Hf 32, Hf1131	2.1	2.9	Grobler-Rabie et al. 1985*b*
	1A2E40	2.1	2.9	Sykes et al. 1986

Note.—See references cited give further details.

[a] Subclone FG2 is a derivative of 1A1H17526 containing reduced amounts of repetitive sequence.

Results

Figure 2 displays the detail of the pedigrees and type I collagen genotypes. Table 3 is a summary of the lod scores at $\theta = .00$ that were calculated using MLINK or LIPED programs or, in simple cases, by arithmetic. Other values of θ were not considered, since for candidate genes using internal or very close markers, linkage analysis becomes segregation analysis with only two discrete conclusions. If any true recombinants are observed, segregation is discordant and the candidate is excluded as the mutant locus. If segregation is concordant, the likelihood that the candidate gene is at or very close to the mutant locus is conveniently expressed as the positive lod score.

If we assume that OI families may belong to one of two discrete types—a fraction (α) linked to one or other of the collagen loci at $\theta = .00$ or a fraction ($1-\alpha$) not linked to either of the collagen loci—then the data can be used to estimate the relative proportion of each family type. Since no family was discordant at both loci, we have no direct evidence of unlinked families; but their existence in the population can never be excluded, and this study reports results from only a sample of the population. Further, one or more unlinked families may have been included in the study but not demonstrated because the observed concordant segregation was due to chance, not to linkage.

The study can be used to derive the probability of its detecting one or more unlinked families at different values of α. The probability (p) at $\theta = .00$ that in a family with m offspring none is a recombinant, is given by

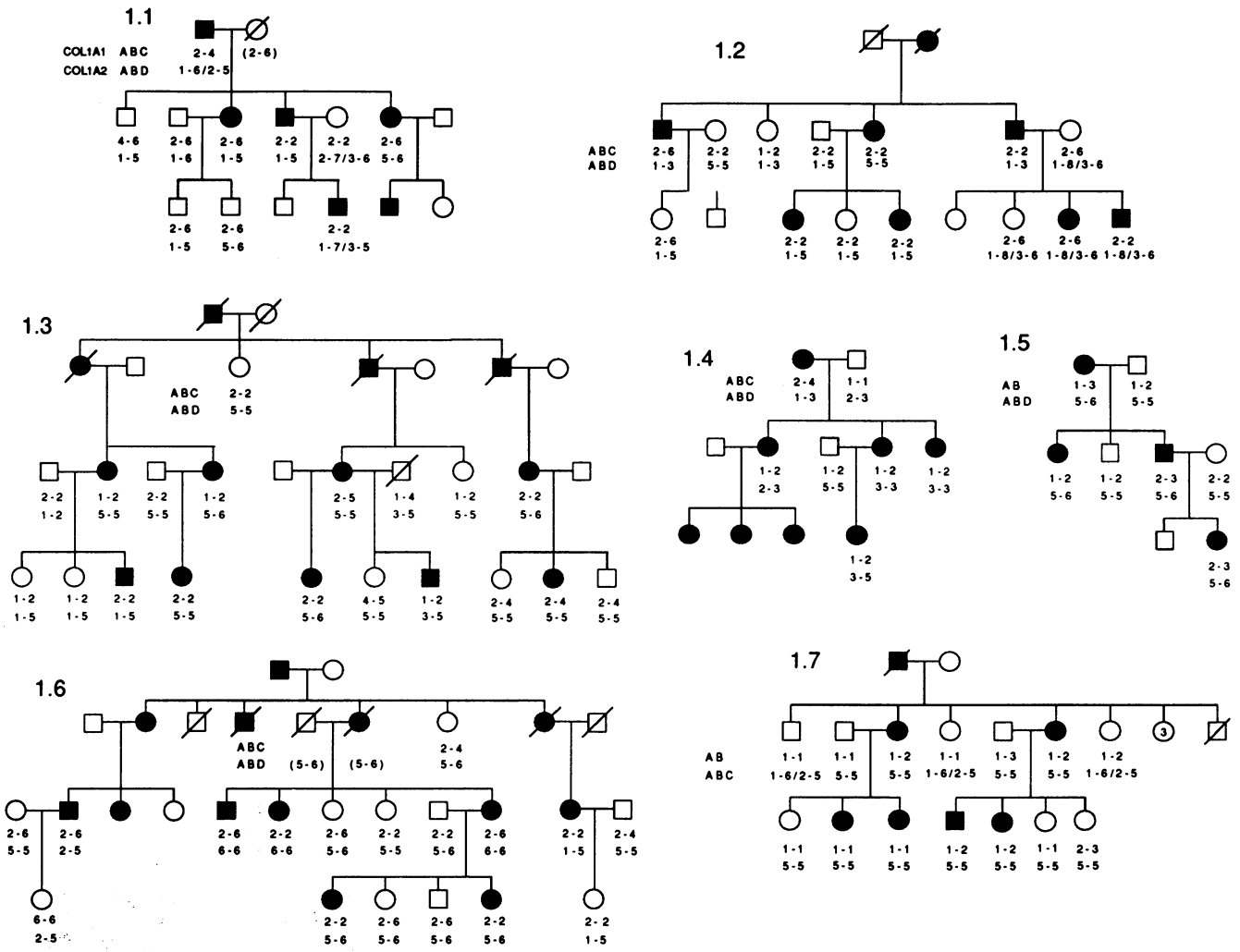
$$p = \alpha + (1-\alpha)(1/2)^m , \quad (1)$$

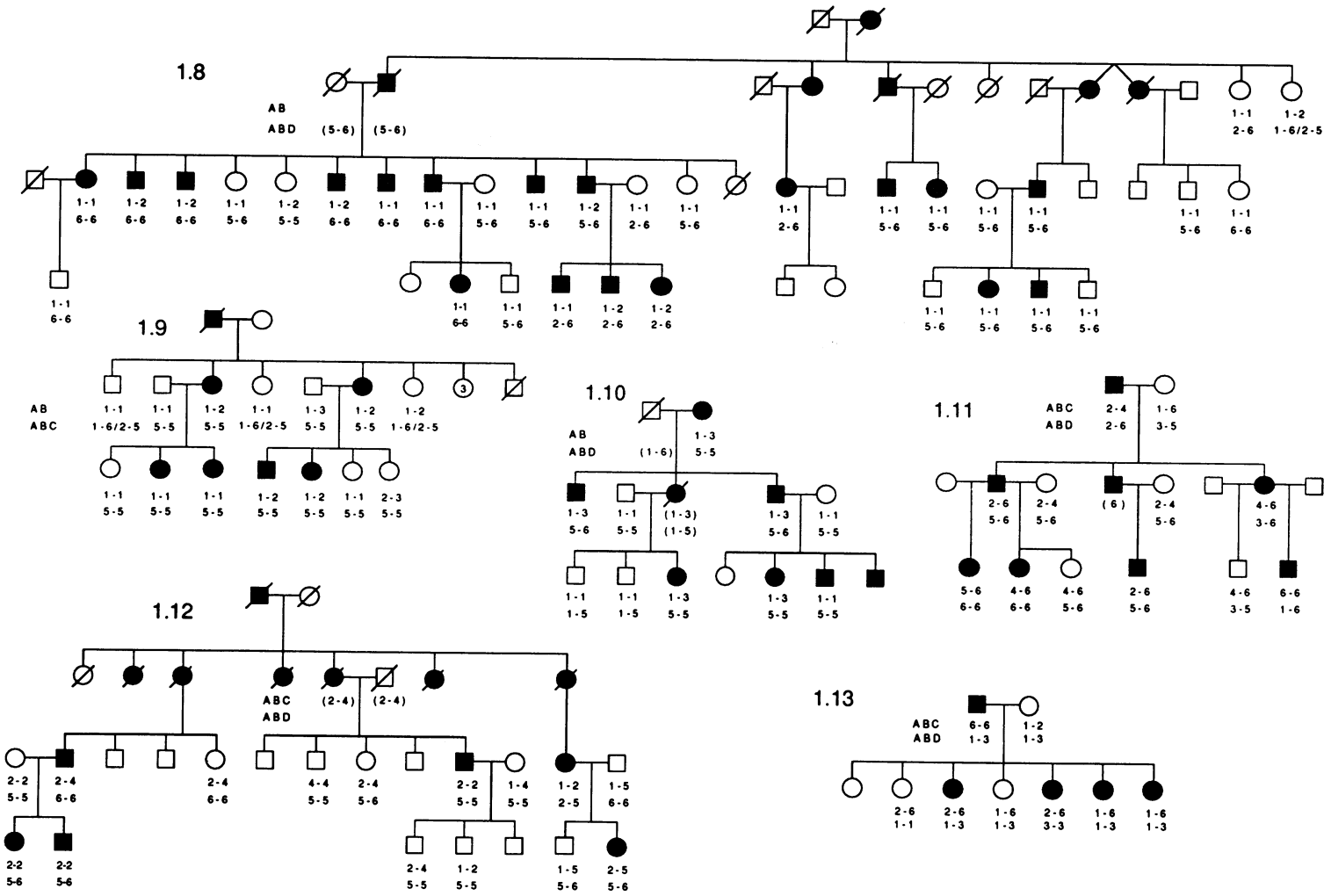
Table 2

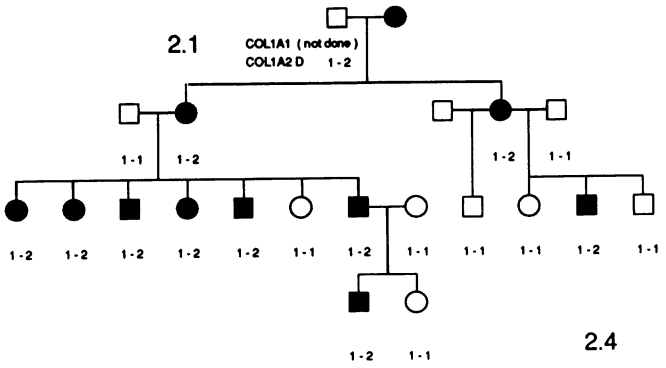
Haplotype Notation

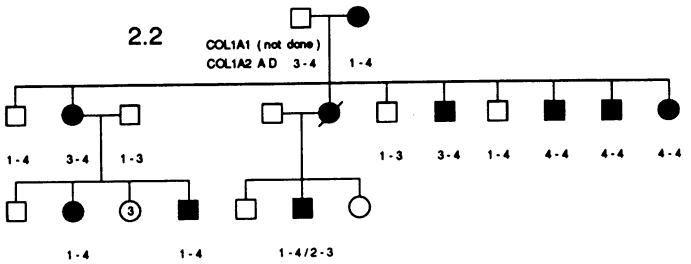
No. of Variants and Haplotype	First Site	Second Site	Third Site
1:			
1	+		
2	−		
2:			
1	+	+	
2	+	−	
3	−	+	
4	−	−	
3:			
1	+	+	+
2	+	+	−
3	+	−	+
4	+	−	−
5	−	+	+
6	−	+	−
7	−	−	+
8	−	−	−

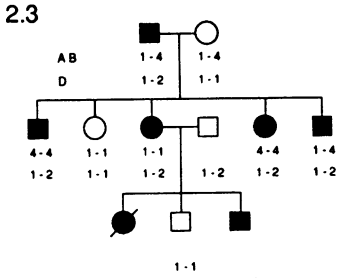
Note.—Combinations of variants used to assemble haplotypes are shown. Haplotype notation is that used to construct the genotypes shown in fig. 2.

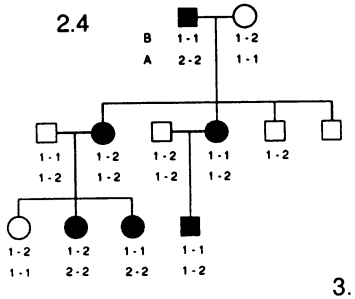

1.1
COL1A1 ABC
COL1A2 ABD
2-4
1-6/2-5
(2-6)
4-6
1-5
2-6
1-6
2-6
1-5
2-2
1-5
2-2
2-7/3-6
2-6
5-6
2-6
1-5
2-6
5-6
2-2
1-7/3-5
1.2
ABC
ABD
2-6
1-3
2-2
5-5
1-2
1-3
2-2
1-5
2-2
5-5
2-2
1-3
2-6
1-8/3-6
2-6
1-5
2-2
1-5
2-2
1-5
2-2
1-5
2-6
1-8/3-6
2-6
1-8/3-6
2-2
1-8/3-6
1.3
ABC
ABD
2-2
5-5
2-2
1-2
1-2
5-5
2-2
5-5
1-2
5-6
2-5
5-5
1-4
3-5
1-2
5-5
2-2
5-6
1-2
1-5
1-2
1-5
2-2
1-5
2-2
5-5
2-2
5-6
4-5
5-5
1-2
3-5
2-4
5-5
2-4
5-5
2-4
5-5
1.4
ABC
ABD
2-4
1-3
1-1
2-3
1-2
2-3
1-2
5-5
1-2
3-3
1-2
3-3
1-2
3-5
1.5
AB
ABD
1-3
5-6
1-2
5-5
1-2
5-6
1-2
5-5
2-3
5-6
2-2
5-5
2-3
5-6
1.6
ABC
ABD
(5-6)
(5-6)
2-4
5-6
2-6
5-5
2-6
2-5
6-6
2-5
2-6
6-6
2-2
6-6
2-6
5-6
2-2
5-5
2-2
5-6
2-6
6-6
2-2
5-6
2-6
5-6
2-6
5-6
2-2
5-6
2-2
1-5
2-4
5-5
1.7
AB
ABC
1-1
1-6/2-5
1-1
5-5
1-2
5-5
1-1
1-6/2-5
1-3
5-5
1-2
5-5
1-2
1-6/2-5
3
1-1
5-5
1-1
5-5
1-1
5-5
1-2
5-5
1-2
5-5
1-1
5-5
2-3
5-5

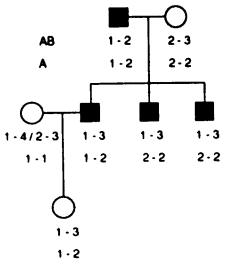

1.8
AB
ABD
(5-6)
(5-6)
1.9
AB
ABC
1.10
AB
ABD
1.11
ABC
ABD
1.12
ABC
ABD
1.13
ABC
ABD

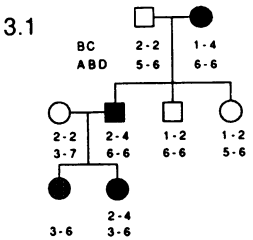

2.1
COL1A1 (not done)
COL1A2 D 1-2
1-1 1-2
1-2 1-1
1-2 1-2 1-2 1-2 1-2 1-1 1-2 1-1 1-1 1-1 1-2 1-1
1-2 1-1

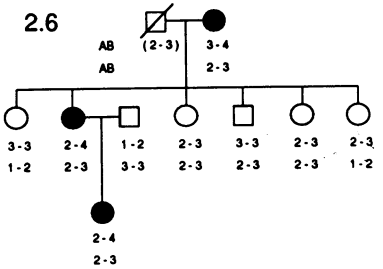

2.2
COL1A1 (not done)
COL1A2 A D 3-4 1-4
1-4 3-4 1-3
1-3 3-4 1-4 4-4 4-4 4-4
3
1-4 1-4 1-4/2-3

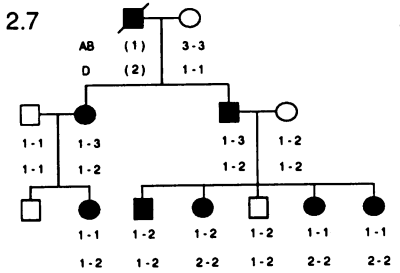

2.3
AB 1-4 1-4
D 1-2 1-1
4-4 1-1 1-1 4-4 1-4
1-2 1-1 1-2 1-2 1-2 1-2
1-1

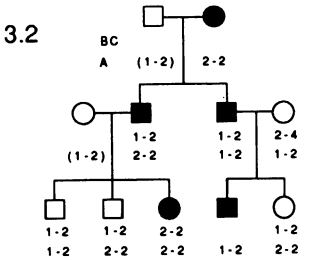

2.4
B 1-1 1-2
A 2-2 1-1
1-1 1-2 1-2 1-1 1-2
1-2 1-2 1-2 1-2
1-2 1-2 1-1 1-1
1-1 2-2 2-2 1-2

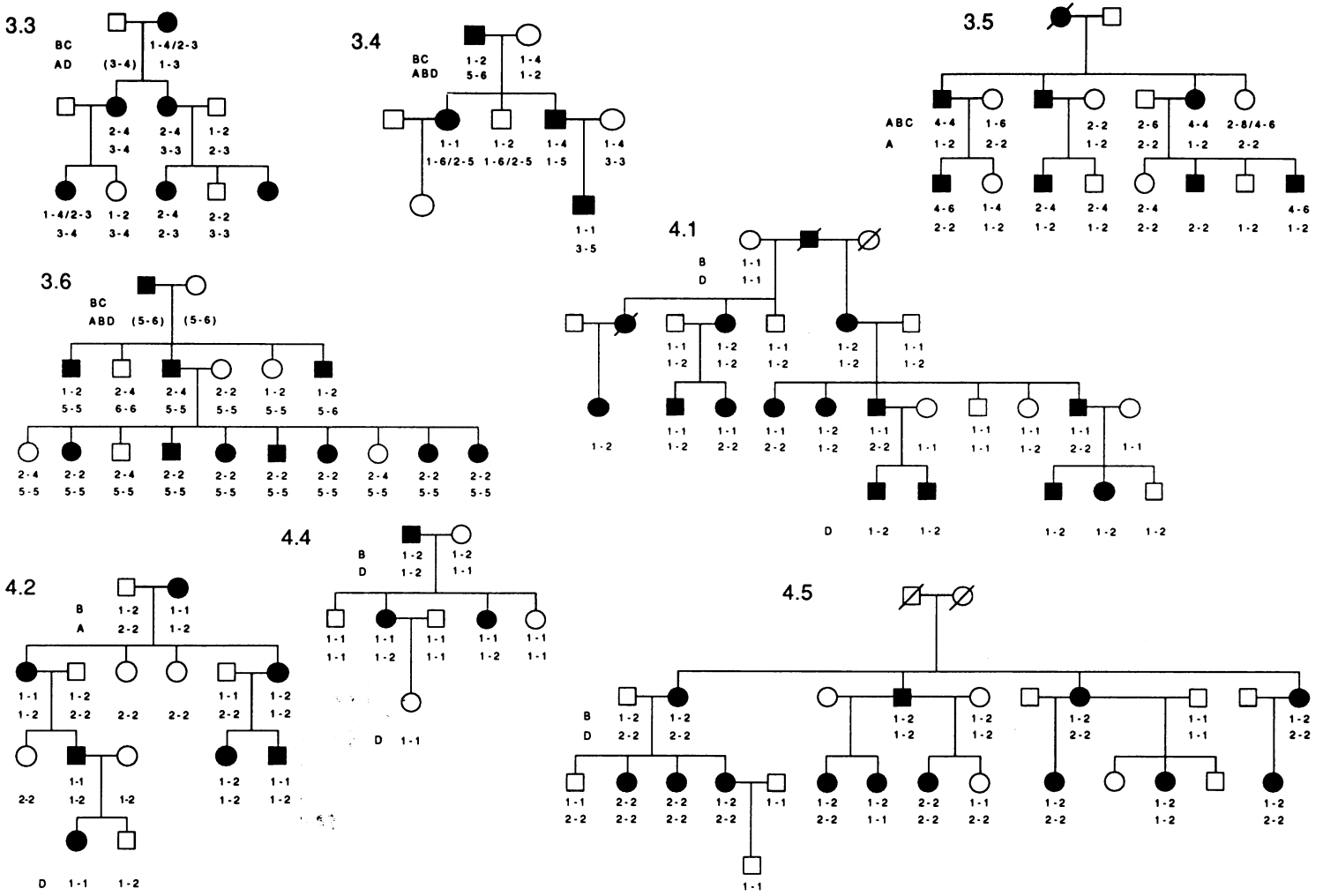

2.5
AB 1-2 2-3
A 1-2 2-2
1-4/2-3 1-3 1-3 1-3
1-1 1-2 2-2 2-2
1-3
1-2

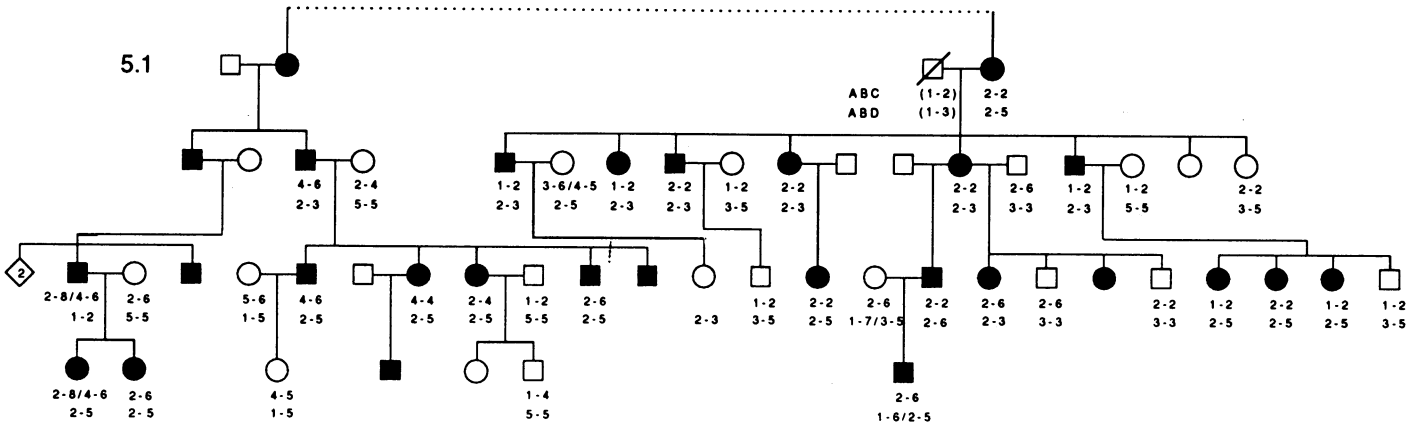

3.1
BC 2-2 1-4
ABD 5-6 6-6
2-2 2-4 1-2 1-2
3-7 6-6 6-6 5-6
3-6 2-4
3-6

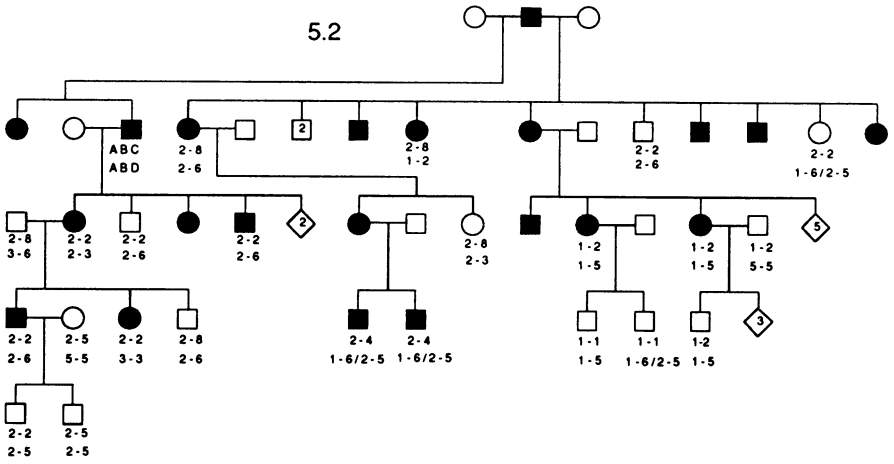

2.6
AB (2-3) 3-4
AB 2-3
3-3 2-4 1-2 2-3 3-3 2-3 2-3
1-2 2-3 3-3 2-3 2-3 2-3 1-2
2-4
2-3

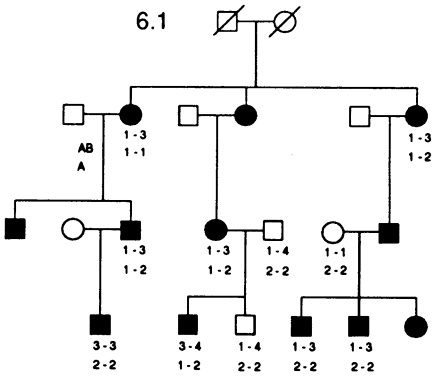

2.7
AB (1) 3-3
D (2) 1-1
1-1 1-3 1-3 1-2
1-1 1-2 1-2 1-2
1-1 1-2 1-2 1-2 1-1 1-1
1-2 1-2 2-2 1-2 2-2 2-2


3.2
BC
A (1-2) 2-2
(1-2) 1-2 1-2 2-4
2-2 1-2 1-2
1-2 1-2 2-2 1-2
1-2 2-2 2-2 1-2 2-2


3.3
BC
AD
(3-4)
1-4/2-3
1-3
2-4
3-4
2-4
3-3
1-2
2-3
1-4/2-3
3-4
1-2
3-4
2-4
2-3
2-2
3-3
3.4
BC
ABD
1-2
5-6
1-4
1-2
1-1
1-6/2-5
1-2
1-6/2-5
1-4
1-5
1-4
3-3
1-1
3-5
3.5
ABC
A
4-4
1-2
1-6
2-2
2-2
1-2
2-6
2-2
4-4
1-2
2-8/4-6
2-2
4-6
2-2
1-4
1-2
2-4
1-2
2-4
1-2
2-4
2-2
2-2
1-2
4-6
1-2
3.6
BC
ABD
(5-6)
(5-6)
1-2
5-5
2-4
6-6
2-4
5-5
2-2
5-5
1-2
5-5
1-2
5-6
2-4
5-5
2-2
5-5
2-4
5-5
2-2
5-5
2-2
5-5
2-2
5-5
2-4
5-5
2-2
5-5
2-2
5-5
4.1
B
D
1-1
1-1
1-2
1-1
1-2
1-2
1-1
1-2
1-2
1-1
1-2
1-1
1-2
1-2
1-1
1-2
2-2
1-1
1-2
1-2
1-1
1-2
2-2
1-1
1-1
1-2
2-2
1-1
D
1-2
1-2
1-2
1-2
1-2
4.2
B
A
1-2
2-2
1-1
1-2
1-1
2-2
1-2
2-2
2-2
2-2
1-1
2-2
1-2
1-2
1-2
1-2
1-1
2-2
1-2
1-2
D
1-1
1-2
4.4
B
D
1-2
1-2
1-2
1-1
1-1
1-1
1-1
1-2
1-1
1-1
1-1
1-2
1-1
D
1-1
4.5
B
D
1-2
2-2
1-2
2-2
1-2
1-2
1-2
1-2
1-2
2-2
1-1
1-1
1-2
2-2
1-1
2-2
2-2
2-2
2-2
2-2
1-1
1-2
2-2
1-1
2-2
2-2
2-2
1-2
2-2
1-2
1-2
1-2
2-2
1-2
2-2
1-1


5.1
ABC
ABD
(1-2)
(1-3)
2-2
2-5


5.2
ABC
ABD


6.1
AB
A

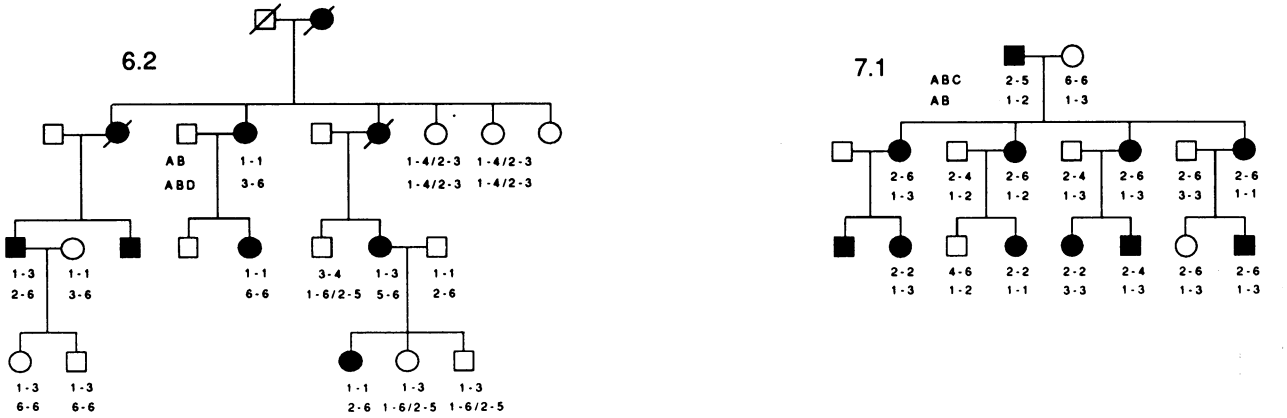


Figure 2 Pedigrees showing COL1A1 (*upper*) and COL1A2 (*lower*) genotypes. The haplotypes at each locus are constructed from the combination of results from the restriction-site dimorphisms indicated by letters at the top left of each pedigree. These letters refer to the sites on fig. 1 and are combined according to the notation in table 2. Genotypes separated by a slash indicate two indistinguishable alternatives, and those in brackets are deduced from relatives. The first digit of each pedigree number indicates the center (see list of authors) primarily involved as follows: 1 = Oxford; 2 = Cape Town; 3 = Harrow; 4 = Connecticut; 5 = Reykjavík; 6 = Oslo; 7 = Zurich. Hatched symbols represent individuals with uncertain phenotype.

Table 3

Lod Scores at Each Locus and the Linkage Constant for Each Family

Pedigree	Lod Score (θ = .00)		Linkage Constant
	COL1A1	COL1A2	
1.1	1.2	$-\infty$	.88
1.2	1.2	$-\infty$	.88
1.3	2.1	$-\infty$	.98
1.4	.8	.6	.70
1.5	$-\infty$	.9	.75
1.6	1.6	$-\infty$	.95
1.7	$-\infty$	.9	.75
1.8	$-\infty$	4.3	1.00
1.9	$-\infty$	2.0	.98
1.10	$-\infty$	1.5	.94
1.11	$-\infty$	2.2	.99
1.12	1.7	$-\infty$	.96
1.13	NI	0.90	.75
2.1	ND	3.9	1.00
2.2	ND	2.7	1.00
2.3	$-\infty$	1.5	.94
2.4	$-\infty$	.6	.50
2.5	.6	$-\infty$	.50
2.6	1.8	$-\infty$	.97
2.7	$-\infty$	1.5	.94
3.1	.9	NI	.75
3.2	.7	$-\infty$	.60
3.3	.9	.6	.75
3.4	.9	.6	.75
3.5	.3	$-\infty$	.00
3.6	3.1	$-\infty$	1.00
3.7	.8	$-\infty$	.94
4.1	$-\infty$	1.5	.94
4.2	$-\infty$	2.4	.99
4.3	1.2	$-\infty$	.88
4.4	$-\infty$	1.2	.88
4.5	2.0	$-\infty$	.98
5.1	$-\infty$	7.7	1.00
5.2	.8	$-\infty$	.70
6.1	1.1	$-\infty$	.85
6.2	1.6	$-\infty$	.95
6.3	$-\infty$	NI	...
7.1	2.7	$-\infty$	1.00

Note.—A lod score of $-\infty$ denotes discordance. NI = Noninformative; ND = not done.

(Ott 1986), the component α being the probability of sampling a linked family and the component $(1-\alpha)(1/2)^m$ being the probability of chance concordance in an unlinked family. The likelihoods ratio in favor of linkage (L), a direct derivative of the computed lod score, can be used as a convenient alternative to derive an approximation of the second component of equation (1).

$$p = \alpha + (1-\alpha)(1/L) . \quad (2)$$

This equation can now be used for dealing with more than one marker locus, as in the present study. Under these circumstances families will segregate in one of three ways: (1) with one locus and not the other, (2) with both loci (in which case segregation with one must be fortuitous), or (3) with neither (indicating a third disease locus). The first two ways are compatible with linkage to one or other of the loci.

The probability (p_i) that a given family belongs to

segregation types 1 or 2, (i.e., no recombinant offspring) is given by

$$p_i = \alpha + (1-\alpha)(2/L_i^{max}) = \alpha\,(1-2/L_i^{max}) + 2/L_i^{max}, \quad (3)$$

where α is now the proportion of the families in the population that are linked to *either* locus and where L_i^{max} is the higher of the two likelihood ratios. In families of segregation type 1, L^{max} will be derived from the nondiscordant locus. In type 2 families, usually small, where there is concordance at both loci but only in one can this be due to linkage, L^{max} is taken to be the higher of the two ratios. As two unlinked loci are being considered, the probability of concordance by chance at one of them is doubled compared with that for the single locus (equation [2]), and this is accounted for by the numerator in the second part of equation (3).

Table 3 gives the value of $(1-2/L^{max})$ for each family. This is the linkage constant shown in the table. The overall probability (P) of detecting one or more unlinked families in the 38-family series is given by

$$P = 1 - (p_1 \cdot p_2 \cdot p_3 \ldots, p_{38}),$$

where p_1, p_2, etc., are the values for families 1, 2, etc. and may be calculated for different values of α by using equation (3). These values are given in table 4. These probabilities can be seen as defining the approximate confidence levels that the true value of α is at or above that used in the calculation. Thus, we are 97% confident that α is .90 or higher. The 95% confidence interval places α between .91 and 1.00.

Table 5 records the important phenotypic characteristics of each pedigree that have been assessed as the most severe expression of the OI gene in that pedigree. Table 6 compares the mutant locus in a family, as determined by segregation analysis, and features of the phenotype shown in table 5.

Discussion

The principal aim of the present study has been to provide essential information for the application of prenatal diagnosis in dominantly inherited OI. Unlike inherited disorders in which the genes responsible have a low new mutation rate, OI is most probably caused by a different mutation in each separate family. This level of allelic heterogeneity in OI rules out widespread prenatal diagnosis by identification of specific mutations, at least until techniques for doing so improve considerably. Nevertheless, if these mutations occur at a very limited number of loci, then indirect identification, using genetic markers, of the mutant allele in a fetus can provide an adequate basis for such a test. The results presented here indeed show that the genetic heterogeneity in dominantly inherited OI does appear to be limited to the two type I collagen loci COL1A1 and COL1A2.

Table 4

Lower Confidence Levels for Different Values of the Proportion of OI Families Linked to Either COL1A1 or COL1A2

α	.80	.85	.90	.95
P	.999	.993	.966	.808

NOTE.—Data are the probabilities (P) of observing one or more unlinked families in the study sample at different values of the true proportion of linked families in the population (α).

There are two important caveats. The first is that the conclusion is based on OI with clear dominant inheritance. Other work has shown that recessively inherited OI can be caused by mutations at loci unlinked to either COL1A1 or COL1A2 (Aitchison et al. 1988; Daw et al. 1988; Wallis et al., submitted). The second caveat is that the pedigrees studied here were extensive, and this selection bias should be borne in mind when applying the conclusions to smaller families, who will form the majority of those requesting prenatal diagnosis. However, there is no reason to think that the disease is fundamentally different in small families.

The study demonstrates a 95% probability that the proportion of OI families linked to either COL1A1 or COL1A2 is .91 or higher. There is an 80% chance that this proportion is higher than .95. Further improvements in this figure are subject to the law of diminishing returns. For instance, to increase the lower limit for 95% confidence from .91 to .95 would require at least another 30 large families to be analyzed—a formidable and probably unnecessary task. Other sources of error, such as misdiagnosis, sampling, and technical problems, would become more significant than the diminishing possibility of a third OI locus.

Although we did not find any evidence of a third unlinked OI locus, it can never be ruled out by this sort of study. Nonetheless, were one to be demonstrated later, the present study suggests that its frequency in the OI population is likely to be low. Clearly, prenatal diagnosis in such a family on the basis of identification of a concordant type I collagen allele in the fetus would be misleading. Therefore, we feel it would be prudent at

Table 5

Phenotype Features of OI Families and the Mutant Locus

Pedigree	Phenotypic Feature							
	Mutant Locus[a]	Sillence Type[b]	Scleral Color[c]	Dentinogenesis Imperfecta[d]	Hearing Loss[d]	Fractures at Birth[d]	Fracture Number[e]	Deformity[f]
1.1	1	1	2	0	1	0	3	0
1.2	1	1	2	0	1	1	4	0
1.3	1	1	2	0	1	0	3	0
1.4	3	1	3	0	0	0	1	0
1.5	2	1	2	0	0	0	2	0
1.6	1	1	2	0	0	0	3	2
1.7	2	1	2	0	0	0	1	2
1.8	2	4	1	1	1	1	3	0
1.9	2	4	0	1	1	0	3	1
1.10	2	4	0	1	1	0	4	3
1.11	2	4	0	1	1	1	3	3
1.12	1	1	2	0	0	0	2	0
1.13	3	1	2	0	1	0	2	0
2.1	2	1	2	1	0	0	3	1
2.2	2	1	1	1	0	1	1	0
2.3	2	1	2	1	0	0	1	1
2.4	2	1	2	1	0	0	1	1
2.5	1	1	2	0	1	0	2	0
2.6	1	1	2	1	1	0	1	0
2.7	2	1	2	1	0	0	1	0
3.1	3	1	2	0	0	0	3	1
3.2	1	1	2	0	0	0	4	1
3.3	3	1	2	1	0	0	3	1
3.4	3	1	2	0	1	0	2	0
3.5	1	1	2	0	1	0	1	0
3.6	1	1	2	0	1	1	2	1
3.7	1	1	2	1	1	0	2	0
4.1	2	4	1	1	1	1	3	2
4.2	2	4	0	1	1	1	4	2
4.3	1	1	0	0	1	0	4	2
4.4	2	4	0	0	1	1	3	0
4.5	1	1	3	0	1	0	2	0
5.1	2	4	0	1	0	0	3	0
5.2	1	1	2	0	1	0	3	1
6.1	1	1	2	0	0	0	2	0
6.2	1	1	3	0	1	0	3	0
6.3	3	1	2	0	1	1	2	0
7.1	1	1	2	0	1	0	2	0

NOTE.—A summary of the phenotypic features of each pedigree and the concordant locus (the score for each feature records the *most severe expression* of the OI gene in that pedigree, not necessarily in the same individual) is as follows: mutant locus—1 = COL1A1, 2 = COL1A2, 3 = uncertain; Sillence type—1 = type 1, 4 = type IV; scleral color—0 = white, 1 = light blue/grey, 2 = mid blue, 3 = dark blue; fracture number—1 = less than 10, 2 = 10–19, 3 = 20–50, 4 = more than 50; deformity—0 = absent, 1 = mild, 2 = moderate, 3 = severe; other features—0 = absent, 1 = present.

this stage to include a 5% probability of this being the case when calculating the risk modification for those families requesting prenatal diagnosis in cases where there is no independent evidence of linkage to type I collagen genes from segregation analysis.

At present, accurate prenatal diagnosis of OI in any family depends on information from not one but both type I collagen loci. The first step must be to distinguish whether the disease in that family is linked to COL1A1 or to COL1A2, by analyzing the segregation

Table 6

Analysis of Phenotype Characters by Concordant Locus

Category	Total No.	Locus		
		COL1A1	COL1A2	Uncertain
All OI:				
Sillence type				
I	30	17	7	6
IV	8	0 (.00)[a]	8 (1.00)[a]	0
Sillence type I:				
Scleral color:				
0	1	1	0	0
1	1	0	1	0
2	25	14	6	5
3	3	2	0	1
Dentinogenesis imperfecta:				
0	22	15	2	5
1	8	2 (.29)[b]	5 (0.71)[b]	1
Presenile hearing loss:				
0	14	4 (.36)[b]	7 (.64)[b]	3
1	16	13 (1.00)[a]	0 (.00)[a]	3
Fractures at birth:				
0	26	15	6	5
1	4	2	1	1
Fracture number:				
1	8	2 (.29)[b]	5 (.71)[b]	1
2	11	7	1	3
3	8	5	1	2
4	3	3	0	0
Deformity:				
0	20	12	3	4
1	8	3	3	2
2	2	2	1	0

Note.—Data are summaries of phenotypic features from table 5, grouped according to the concordant locus in that pedigree. Figures in parentheses are the proportion of *locus-assigned* pedigrees in each column only where there is a significant distortion from the expected distribution.

[a] $.01 > p > .001$ (calculated from χ^2 distribution).

[b] $.1 > p > .05$ (calculated from χ^2 distribution).

patterns of the two loci until one is seen to be discordant. So long as the paternity and maternity of apparent recombinants are correct, this identifies the other locus as the mutant, within the confidence limits discussed above. Prenatal diagnosis then depends on being able to distinguish, in the usual way, the mutant from the normal allele in the fetus at risk. The requirement for informative meioses at two loci means that some families cannot be offered prenatal diagnosis because the mutant locus cannot be identified even though the parental genotypes of the fetus at risk are fully informative at one or even at both loci.

Alternative ways of identifying the mutant locus in these families would be valuable, and we have analyzed aspects of the clinical phenotypes of the pedigrees in this series to see whether any are likely to be useful for predicting the mutant locus. The results of this analysis are shown in tables 3 and 4 and are derived from data received from the collaborating centers, which were asked to record the *most severe* expression of the gene in the family. This introduces a bias in that, clearly, the larger number of affected individuals there are within a pedigree, the greater the chance of encountering more severe symptoms.

In 32 of the 38 pedigrees the mutant locus could be identified by segregation analysis. Of the remaining six families, three were concordant at both loci—though, clearly, had more informative meioses been available, one locus would have eventually become discordant. Two families were concordant at one locus and nonin-

formative at the other, while one family was discordant at one locus and noninformative at the other. The disease in eight of these 32 "locus-assigned" pedigrees was classified by the examining clinicians as Sillence OI type IV (short stature, deformity, white sclerae), and all eight segregated with COL1A2. Of the 24 remaining pedigrees, classified as Sillence OI type 1 (normal stature, no deformity, blue sclerae), seven segregated with COL1A2 and 17 with COL1A1. From this it emerges that the Sillence classification itself is a promising indicator of the mutant locus where the phenotype is OI IV but not where it is OI 1. The correlation between OI IV and COL1A2 has been noted before (Tsipouras et al. 1984; Sykes et al. 1986), but this series extends from four to eight the number of separate pedigrees following the pattern.

When the 24 OI 1 pedigrees were examined for other expression parameters, some emerged as surprisingly good predictors of the concordant locus (table 5). All 13 pedigrees with early-onset hearing loss were linked to COL1A1. Five of the seven pedigrees linked to COL1A2 had dentinogenesis imperfecta, while only two of the 17 linked to COL1A1 showed this complication. Fracture numbers were higher in COL1A1-linked families compared with COL1A2 segregants. On the other hand, deformity and the presence of fractures at birth were of no predictive value in the series.

It would be a mistake to attribute too much weight to these correlations, especially when the phenotype data collection was not particularly carefully controlled and refers to expression in pedigrees and not in individuals. The variability of expression in OI is well known, and in the same pedigree (e.g., 1.8) the same gene can be carried by some individuals with all the signs associated with severe OI IV and by others in whom even the diagnosis of OI is in doubt. In practice this information might be used to adjust the risk in such cases where there was a very pressing need for prenatal diagnosis and where the limitations of the precision were well understood by the parents.

When the causal mutations at both loci are fully described and when their effects on collagen and matrix behavior are understood, a sounder basis for these correlations between phenotype and mutant locus may emerge. Until then, by far the safest way of discriminating between the two loci is by a clear demonstration of recombination at one of them.

Acknowledgments

Our thanks are due to the patients and their families who took part in this study. We are also grateful to many physicians who allowed us access to their patients, especially Drs. Roger Smith, Colin Paterson, Ruth Wynne-Davies, L. Stefan Levin, and Jon Bjarnason, Lesley Watts typed the manuscript. The work involved was supported by grants from the following: Action Research for the Crippled Child (U.K.), Wellcome Trust, Medical Research Council of South Africa, University of Cape Town Staff Research Fund, Mauberger Foundation, Harry Crossley Foundation, March of Dimes (Clinical Research grant 6–534), Swiss National Foundation (grant 3.905.086). P.T. is the recipient of Clinical Investigator Award AR 01224 from the National Institutes of Health.

References

Aitchison K, Ogilvie DJ, Honeyman M, Thompson E, Sykes BC (1988) Homozygous osteogenesis imperfecta unlinked to collagen I genes. Hum Genet 78:233–236

Børresen A-L, Berg K, Tsipouras P, Dickson LA, Prockop DJ, Ramirez F (1985) DNA polymorphisms in collagen genes: potential use in the study of disease. Prog Clin Biol Res 177:37–51

Byers PH, Tsipouras P, Bonadio JF, Starman B, Schwartz RC (1988) Perinatal lethal osteogenesis imperfecta (OI type II): a biochemically heterogeneous disorder usually due to new mutations in the genes for type I collagen. Am J Hum Genet 42:237–248

Daw S, Nicholls AC, Williams EM, Sykes BC, Pope FM (1988) Autosomal recessive osteogenesis imperfecta: excess post-translational modification of collagen not linked to either COL1A1 or COL1A2. J Med Genet 25:275

Falk CT, Schwartz RC, Ramirez F, Tsipouras P (1986) Use of molecular haplotypes for the human proα_2 (I) collagen gene in linkage analysis of the mild autosomal dominant forms of osteogenesis imperfecta. Am J Hum Genet 38: 269–279

Gilchrist F (1987) Dissertation: part 2. Final honours school (biochemistry), University of Oxford, Oxford

Grobler-Rabie AF, Brebner DK, Vandenplas S, Wallis G, Dalgleish R, Kaufman RE, Mathew CG, et al (1985*a*) Polymorphism of DNA sequence in the pro-α_2(1) collagen gene. J Med Genet 22:182–186

Grobler-Rabie AF, Wallis G, Brebner DK, Beighton P, Mathew CG (1985*b*) Detection of a high frequency Rsa 1 polymorphism in the human pro-α_2(1) collagen gene which is linked to an autosomal dominant form of osteogenesis imperfecta. EMBO J 4:1745–1748

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73

McKusick VA (1972) Heritable disorders of connective tissue. Mosby, St. Louis

Ogilvie DJ, Aitchison K, Sykes BC (1987) An RFLP close to the human collagen gene COL1A1. Nucleic Acids Res 15:4699

Ott J (1986) The number of families required to detect or exclude linkage heterogeneity. Am J Hum Genet 39:159–165

Penttinen RP, Lichtenstein JR, Martin GR, McKusick VA

(1975) Abnormal collagen metabolism in cultured cells in osteogenesis imperfecta. Proc Natl Acad Sci USA 72: 586–589

Sillence DO, Senn A, Danks DM (1979) Genetic heterogeneity in osteogenesis imperfecta. J Med Genet 16:101–116

Sykes BC, Francis MKO, Smith R (1977) Altered relation of two collagen types in osteogenesis imperfecta. N Engl J Med 296:1200–1203

Sykes BC, Ogilvie DJ, Wordsworth BP, Anderson DJ, Jones N (1986) Osteogenesis imperfecta is linked to both type 1 collagen structural genes. Lancet 2:69–72

Tsipouras P, Børresen AL, Dickson LA, Berg K, Prockop DJ, Ramirez F (1984) Molecular heterogeneity in the mild autosomal dominant forms of osteogenesis imperfecta. Am J Hum Genet 36:1172–1179

Tsipouras P, Myers JC, Ramirez F, Prockop DJ (1983) Restriction fragment length polymorphism associated with the pro-α_2 (1) gene of human type 1 procollagen. J Clin Invest 72:1262–1267

Wallis G, Beighton P, Boyd C, Mathew CG (1986) Mutations associated with the pro-α_2 (1) collagen gene are responsible for several cases of osteogenesis imperfecta type 1. J Med Genet 23:411–416

Wallis G, Versfeld J, Sykes BC, Mathew CG, Beighton P. Osteogenesis imperfecta type III—mutations in the type 1 collagen structural genes are not necessarily responsible (submitted)