European Journal of Human Genetics (2015) 23, 1042–1050

www.nature.com/ejhg

ARTICLE

Genetic epidemiology, prevalence, and genotype–phenotype correlations in the Swedish population with osteogenesis imperfecta

Supplementary Tables 1 and 2 have been replaced since online publication. A corrigendum appears in this issue

Katarina Lindahl*[,1], Eva Åström[2,3], Carl-Johan Rubin[4], Giedre Grigelioniene[5,6], Barbro Malmgren[7,8], Östen Ljunggren[1] and Andreas Kindmark[1,9]

Osteogenesis imperfecta (OI) is a rare hereditary bone fragility disorder, caused by collagen I mutations in 90% of cases. There are no comprehensive genotype–phenotype studies on >100 families outside North America, and no population-based studies determining the genetic epidemiology of OI. Here, detailed clinical phenotypes were recorded, and the *COL1A1* and *COL1A2* genes were analyzed in 164 Swedish OI families (223 individuals). Averages for bone mineral density (BMD), height and yearly fracture rate were calculated and related to OI and mutation type. N-terminal helical mutations in both the α1- and α2-chains were associated with the absence of dentinogenesis imperfecta ($P<0.0001$ *vs* 0.0049), while only those in the α1-chain were associated with blue sclera ($P=0.0110$). Comparing glycine with serine substitutions, α1-alterations were associated with more severe phenotype ($P=0.0031$). Individuals with type I OI caused by qualitative *vs* quantitative mutations were shorter ($P<0.0001$), but did not differ considering fractures or BMD. The children in this cohort were estimated to represent >95% of the complete Swedish pediatric OI population. The prevalence of OI types I, III, and IV was 5.16, 0.89, and 1.35/100 000, respectively (7.40/100 000 overall), corresponding to what has been estimated but not unequivocally proven in any population. Collagen I mutation analysis was performed in the family of 97% of known cases, with causative mutations found in 87%. Qualitative mutations caused 32% of OI type I. The data reported here may be helpful to predict phenotype, and describes for the first time the genetic epidemiology in >95% of an entire OI population.
European Journal of Human Genetics (2015) **23,** 1042–1050; doi:10.1038/ejhg.2015.81; published online 6 May 2015

INTRODUCTION

Osteogenesis imperfecta (OI) is a rare inherited heterogeneous connective tissue disorder; the cardinal symptom being bone fragility causing fractures, bone deformity, and short stature. Low bone mineral density (BMD) is generally observed, and extra-skeletal manifestations, such as blue sclera, dentinogenesis imperfecta (DI), and hearing impairment are common. Dominant mutations in collagen type I are generally stated to be responsible for 90% of cases, while a plethora of other genes have been associated with non-collagen OI in recent years.[1–13]

Collagen type I, encoded by *COL1A1* and *COL1A2,* constitutes 85% of the organic matrix in skeletal tissue, and forms a framework for mineral deposition, rendering bone the tensile properties needed to withstand torsion and bending powers. Procollagen is a heterotrimer, with a helical 1014-amino acid-long central stretch of two α1- and one α2-chains, which is flanked by globular N- and C-terminal regions. The helical portions of the α1- and α2-chain consist of Gly-X-Y triplicate repeats, where glycine is obligate at every third position, as this is the only residue that will fit in the sterically confined helical center.

Traditionally individuals with OI are classified according to Sillence types I–IV, based on clinical, radiological, and hereditary findings.[14] The most common is the mild type I OI, which is associated with blue sclera, fractures during childhood, and sometimes DI. OI type II is defined as pre- or perinatal lethal. These fetuses and infants have extreme osteopenia, pulmonary hypoplasia, short limbs due to crumpling fractures of the long bones, and soft under-mineralized skulls. OI type III is the most severe type compatible with surviving the neonatal period. Untreated individuals often suffer from hundreds of fractures, and have a severe short stature due to bowed long tubular bones, vertebral compression fractures, and scoliosis. OI type IV is a moderately deforming category, with a variable phenotypic presentation, but according to the original classification sclera are white after the first years of life.

OI type I is often caused by a quantitative mutation (*COL1A1* null allele), while OI types II–IV commonly are caused by qualitative

[1]Department of Medical Sciences, Uppsala University, Uppsala, Sweden; [2]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden; [3]Neuropediatric Unit, Astrid Lindgren's Children's Hospital, Karolinska University Hospital, Stockholm, Sweden; [4]Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden; [5]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; [6]Department of Clinical Genetics, Karolinska University Hospital Stockholm, Stockholm, Sweden; [7]Division of Pediatric Dentistry, Department of Dental Medicine, Karolinska Institutet, Huddinge, Sweden; [8]Division of Paediatrics, Department of Clinical Science, Intervention and Technology, Karolinska University Hospital, Huddinge, Sweden; [9]Science for Life Laboratory, Department of Medical Sciences, Uppsala University Hospital, Uppsala, Sweden
*Correspondence: Dr K Lindahl, Department of Medical Sciences, Uppsala University, MMS, ing. 40, Akademiska sjukhuset, Uppsala 75185, Sweden; Tel: +46 707487875; Fax: +46 18 55 36 01. E-mail: katarina.lindahl@medsci.uu.se
Received 10 December 2014; revised 17 March 2015; accepted 20 March 2015; published online 6 May 2015

collagen mutations (about 80% helical glycine substitutions and 20% splice mutations).[15] The phenotypic severity has been correlated to affected gene, helical location, resulting residue, and predicted final protein product; however, phenotypic variability is broad and the full extent of genotype–phenotype correlations remains to be elucidated.[15,16] Over 1300 collagen I mutations have been described in the OI variant database,[17–19] but *COL1A1* and *COL1A2* are comparatively large genes. In the latest large-scale review only 395 of 3765 theoretical helical glycine substitutions had been reported, with the only clinical data presented being whether the mutation was lethal or not.[15] Mutations in the majority of the newly described OI genes are associated with phenotypes that can clinically not be distinguished from those described by traditional classification.

No extensive study of mutations causing OI has been performed in Sweden. To our knowledge, there are no population-based studies on the genetic epidemiology of collagen I mutations in the majority of any OI population, as the available comprehensive prevalence studies predate molecular analyses or are register studies. The point prevalence of OI at birth is often approximated to 5–6.7/100 000, and although experts in the field anticipate the figure to be closer to 10/100 000, this has proven challenging to establish unequivocally.[20]

In the present study, we examined the *COL1A1* and *COL1A2* genes, and collected clinical data on 223 patients from 164 unrelated families with OI, to further illuminate genotype–phenotype correlations, and to describe the spectrum of OI mutations in Sweden. Furthermore, to establish the prevalence of OI and the distribution of collagen mutations in this population, we performed sub-group analyses based on the pediatric part of this cohort, which is estimated to represent over 95% of the entire Swedish clinically relevant pediatric OI population.

MATERIALS AND METHODS

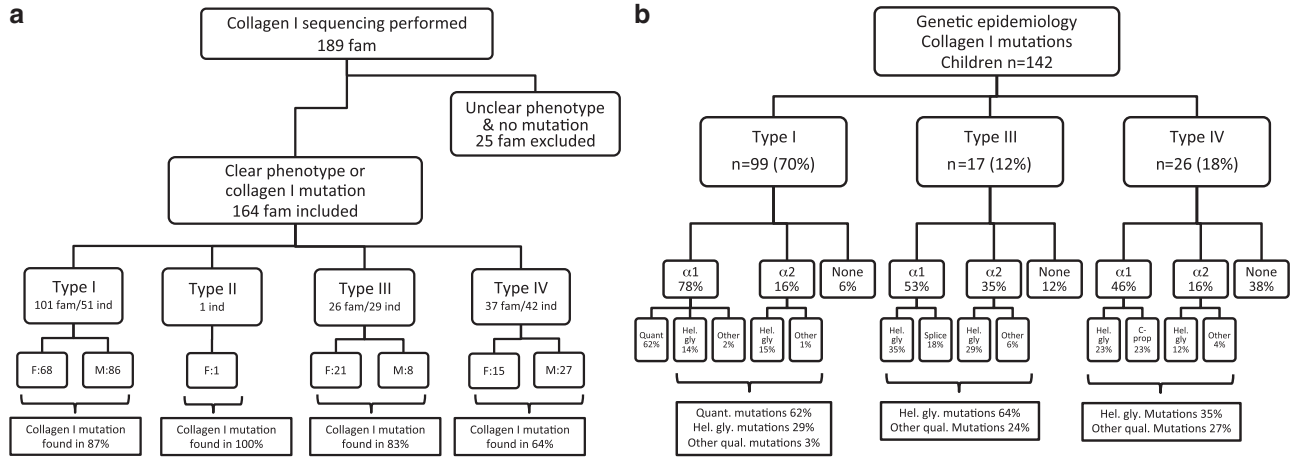
Subjects

Patients with typical or suspected OI cared for at the Uppsala University Hospital Osteoporosis Unit (adult OI center), Uppsala, or at the Astrid Lindgren Children's Hospital at Karolinska University Hospital (Sweden's national OI multi-disciplinary pediatric team), Stockholm, between 2005 and 2014 were offered to enter the study. Previous pediatric patients from a long-term observational study were asked to participate.[21] Patients with OI referred for collagen type I sequencing from other regions of Sweden were also included. Individuals with an uncertain clinical diagnosis ($n=25$) were excluded from the study if no typical mutation was found in *COL1A1* or *COL1A2*; however, all with a definite clinical OI phenotype remained in the study despite a negative sequencing result of these genes ($n=34$ families). The study was approved by the Swedish regional ethics committee at Uppsala University (Ups 06-212), and written consent was obtained from all participants, a parent, or legal guardian signing for children (Flow Chart 1a).

For the prevalence assessment and collagen type I mutation epidemiological analyses, all children and their relatives with OI born between 1 January 1993 and 31 December 2010 were included. These dates were chosen based on that the number of known patients born each year was relatively constant during this time period. Furthermore, it was assumed that the majority of children with *de novo* mutations causing OI, and all children with familial OI, would have been referred to the national OI team for assessment by the age of 4. Very mild *de novo* cases may not have been detected by this age, and may even go without diagnosis throughout life, and will thus not be represented here. Detailed pedigrees were created to ensure individuals and families were only counted once. A total of 142 children in 112 unrelated kindred born during this time period were diagnosed with OI types I, III, or IV, and 138 of these belonged to families in which the *COL1A1* and *COL1A2* genes were sequenced (Flow Chart 1b). The majority of these families ($n=100$) were referred for sequencing in this study, while a minority ($n=9$) already had results of genetic tests from other analysis facilities (Odense, Gent, and Erlanden). DNA samples from four children from three families (all OI type I) were not available for genetic analysis. The number of children younger than 18 years living in Sweden on 31 December 2010 was 1.919.094 according to official Swedish statistics.[22] According to our data request to the Swedish National Patient Register,[23] the number of children with the ICD-10 diagnosis code for OI (Q78.0), that had consulted specialist care at any given time point 2000–2010, was on average 132 individuals. This is close to the number of definite cases identified here.

COL1A1 and *COL1A2* mutation analyses

PCRs, using primers previously described,[24] were run on a GeneAmp PCR system 9800 using AmpliTaq Gold kits and standard reagents. Sequencing reactions were performed using adjusted Big Dye Terminator 3.1 sequencing protocol. Products were run on 16-capillary ABI 3130xl Genetic Analyzer automated sequencer, and thereafter analyzed in software SeqScape v2.5, using the GenBank reference sequences of *COL1A1* (genomic DNA NG_007400.1 and cDNA NM_000088.3) and *COL1A2* (genomic DNA NG_007405.1 and cDNA NM_000089.3). All reagents, equipment, and software were purchased from Applied Biosystems, Austin, TX, USA.



Flow Chart 1 (**a**) The composition of the entire cohort. Flow chart picturing the entire cohort (223 individuals). Data on OI type, gender,and frequency of collagen I mutations provided. (**b**) The genetic epidemiology of OI in Sweden. Flow chart picturing the pediatric cohort (142 individuals), divided by OI and mutation type to illustrate the genetic epidemiology of this disorder in an entire population. The included 142 children were predicted to represent >95% of all children with clinically relevant OI in Sweden.

Table 1a Clinical characteristics of entire cohort by OI type and subdivided by gender

	OI type I	P-*value gender*	*OI type III*	P-*value gender*	*OI type IV*	P-*value gender*	P-*value OI types*
Individuals (n)	151	—	29	—	42	—	—
Gender (F/M)	65/86	—	21/8	—	15/27	—	0.0049
Gender (children F/M)	32/39	—	11/4	—	7/9	—	NS
Gender (adults, F/M)	33/46	—	10/4	—	8/18	—	0.0425
Age baseline (mean ± SD)	17.51 ± 13.94	—	17.72 ± 14.08	—	17.32 ± 10.61	—	NS
Age baseline (gender F/M, mean)	17.85/17.26	NS	18.50/15.50	NS	16.53/17.62	NS	—
BMD Lumbar Spine baseline (*Z*-score, mean ± SD)	−2.66 ± 0.16	—	−5.05 ± 1.90	—	−4.28 ± 1.66	—	<0.0001
BMD baseline (gender F/M, mean *Z*-score)	−2.32/−2.92	0.0482	−4.93/−5.35	NS	−4.36/−4.25	NS	—
Age database lock (mean ± SD)	22.37 ± 16.86	—	22.55 ± 18.73	—	19.64 ± 12.04	—	NS
Age database lock (gender F/M, mean)	22.71/22.11	NS	23.90/19.00	NS	20.13/19.37	NS	—
Height database lock (mean height *Z*-score ± SD)	−1.08 ± 1.38	—	−6.73 ± 3.04	—	−2.59 ± 2.61	—	<0.0001
Height database lock (gender F/M, mean height *Z*-score)	−1.40/−0.84	0.0265	−6.88/−6.25	NS	−3.23/−2.33	NS	—
Fractures database lock (mean per year ± SD)	0.57 ± 0.68	—	3.83 ± 9.32	—	1.33 ± 1.38	—	0.0003
Fractures database lock (gender F/M, mean per year)	0.50/0.64	NS	4.39/2.23	NS	1.04/1.44	NS	—
Fractures database lock (total no. ± SD)	10.59 ± 12.09	—	46.15 ± 51.58	—	21.09 ± 17.32	—	<0.0001
Fractures database lock (gender F/M, total no.)	9.29/11.6	NS	48.25/40.14	NS	17.22/22.61	NS	—
Vertebral fractures database lock (yes)	71%		100%	—	90%		0.0016
Vertebral fractures database lock (gender F/M, yes)	67%/75%	NS	100%/100%	NS	89%/91%	NS	—
DI database lock (yes)	13%	—	58%	—	43%	—	<0.0001
DI database lock (gender F/M, yes)	22%/7%	0.0130	61%/50%	NS	55%/38%	NS	—
Blue sclera database lock (yes)	98%	—	67%	—	33%	—	<0.0001
Blue sclera database lock (gender F/M, yes)	98%/99%	NS	60%/86%	NS	44%/27%	NS	—

Abbreviation: NS, not significant. Clinical characteristics of all 222 individuals with OI types I, III, and IV, irrespective of collagen I mutation status. Baseline BMD values are presented to avoid including bisphosphonate effects, while remaining characteristics are registered at database lock (31 Decembert 2013), average age by OI type stated for both time points. Figures stated are averages or frequencies for each group, as appropriate. Children defined as 18 years or younger at database lock (31 December 2013).

Table 1b Clinical characteristics of entire cohort by OI type and subdivided by age

	OI type I	*OI type III*	*OI type IV*
Individuals (*n*)	151	29	42
Gender (children, F/M)	32/39	11/4	7/9
Gender (adult, F/M)	33/46	10/4	8/18
Age baseline (children, average)	4.49 ± 3.28	0.69 ± 0.82	1.13 ± 1.06
Age baseline (adult, average)	18.77 ± 14.64	17.63 ± 19.37	13.63 ± 10.25
BMD baseline (SD/*Z*-scores, children)	−2.39 ± 1.50	−5.02 ± 1.35	−4.18 ± 1.40
BMD baseline (SD/*Z*-scores, adult))	−2.96 ± 1.35	−5.08 ± 2.61	−4.33 ± 1.81
P-value	NS	NS	NS
Age database lock (average)	22.37 ± 16.86	22.55 ± 18.73	19.64 ± 12.04
Age database lock (children)	9.22 ± 4.71	9.27 ± 4.82	9.25 ± 5.64
Age database lock (adults)	34.18 ± 14.98	36.79 ± 17.58	26.04 ± 10.35
Height database lock (SD/*Z*-scores, children)	−0.95 ± 1.39	−4.97 ± 1.58	−1.77 ± 1.25
Height database lock (SD/*Z*-scores, adult)	−1.18 ± 1.38	−9.38 ± 2.79	−3.04 ± 3.06
P-value	NS	<0.0001	NS
Fractures database lock (per year, children)	0.69 ± 0.85	4.94 ± 12.50	1.91 ± 2.02
Fractures database lock (Per year, adult)	0.48 ± 0.47	2.44 ± 1.66	1.00 ± 0.71
P-value	0.0983	NS	0.0691
Vertebral fractures database lock (yes, children)	65%	100%	82%
Vertebral fractures database lock (yes, adult)	78%	90%	95%
P-value	NS	NS	NS

Abbreviation: NS, not significant. Clinical characteristics of all 222 individuals with OI types I, III, and IV, irrespective of collagen I mutation status. Baseline BMD values are presented to avoid including bisphosphonate effects, while remaining characteristics are registered at database lock (31 December 2013), average age by OI type stated for both time points. Figures stated are averages or frequencies for each group, as appropriate. Children defined as 18 years or younger at database lock (31 December 2013).

Table 2 Clinical characteristics in relation to collagen I mutation

	Individuals	*BMD baseline*	*Height baseline*	*Number of fractures*	*Number of fractures children*	*Vertebral fractures*	*Vertebral fractures*	*Blue sclera*	*Dentinogenesis*
	(n)	*(Z-score)*	*(Z-score)*	*(per year)*	*(per year)*	*(yes)*	*children (yes)*	*(yes)*	*imperfect (yes)*
Type I quant.	83	−2.57	−0.53	0.59	0.72	79%	72%	100%	0%
Type I qual.	40	−2.99	−1.67	0.58	0.68	62%	64%	100%	42%
Type I quant. *vs* qual. *P*-value	—	NS	<0.0001	NS	NS	NS	NS	By definition	<0.0001
All qual. α1-chain	40	−4.13	−3.41	2.55 (1.36[II])	3.63	87%	91%	68%	57%
All qual. α2-chain	49	−3.91	−2.93	0.97	1.07	74%	75%	89%	54%
All qual. α1- *vs* α2-chain *P*-value	—	NS	NS	NS	NS	NS	NS	0.0511	NS
Helical gly α1-chain	38	−4.43	−3.61	1.30	1.31	93%	93%	70%	65%
Helical gly α2-chain	36	−3.94	−2.87	0.92	1.01	72%	73%	88%	53%
Helical gly α1- *vs* α2-chain *P*-value	—	NS	NS	NS	NS	NS	NS	NS	NS
Serine α1-chain	20	−5.0	−3.62	1.53	1.29	93%	89%	56%	79%
Serine α2-chain	20	−3.5	−1.61	0.64	0.63	57%	25%	89%	39%
Serine α1- *vs* α2-chain *P*-value	—	0.0146	0.0154	0.0351	NS	0.0352	0.0212	0.0523	0.0201
Collagen I mutation	42/155[II]	Not relevant	Not relevant	Not relevant	Not relevant	Not relevant	Not relevant	60%	30%
No collagen I mutation	17/32[II]	Not relevant	Not relevant	Not relevant	Not relevant	Not relevant	Not relevant	24%	0%
Collagen *vs* no collagen mutation *P*-value	—	—	—	—	—	—	—	0.0204	<0.0001

Abbreviation: NS, not significant. Comparison of different clinical characteristics in individuals with/without various collagen I mutation types defined for each column. Children defined as 18 years or younger at database lock (31 December 2013).

Thirty-three families negative for mutations by Sanger sequencing were subjected to multiplex ligation-dependent probe amplification (MLPA) analysis of *COL1A1* and *COL1A2*, to detect exonic and larger deletions or duplications. SALSA-MLPA Kits P271 *COL1A1* and P272 *COL1A2* (MRC-Holland, Holland) were utilized according to the manufacturer's recommendations.

Sequence variants were recorded following convention (http://www.hgvs.org). Exons were numbered as in NG_007400.1 (*COL1A1*) and NG_007405.1 (*COL1A2*). All described variants were entered into the OI variant databases[17–19] for *COL1A1* (https://oi.gene.le.ac.uk/home.php?select_db=COL1A1) and *COL1A2* (https://oi.gene.le.ac.uk/home.php?select_db=COL1A2), with patient IDs, AN_004901 to AN_005123 and database ID COL1A1_00265 (previously reported[25]). Variants were defined as novel if not previously reported here. When necessary, segregation analyses were performed if possible in the families to determine the nature of a mutation.

Clinical characteristics

Detailed clinical characteristics were collected through examination of available charts (Tables 1a, b and 2). Patients were clinically classified according to Sillence (types I–IV), and the original classification was thereafter used for further analyses, irrespective of whether a collagen I mutation was found or not. In calculations where children and adults were compared as groups, the cutoff 18 years or younger at database lock (on 31 December 2013) was used to separate adults and children. Due to incomplete phenotypic data, the number of observations varies for the different clinical characteristics.

All height measurements were recorded at database lock as standard deviations (SD) in relation to age- and sex-specific Swedish reference data (*Z*-score). In children, the body length was measured in the supine position using a wooden height measurement board, and height registered was before bisphosphonate exposure in relevant cases.

Data on fractures were collected as the total number of fractures at database lock, without information on relation to bisphosphonate treatment. Digit and rib fractures were not included. The average number of fractures per year was calculated to enable comparison between individuals of all ages. Data on when fractures occurred were not available. Vertebral compression fractures were recorded as present or absent at database lock based on conventional radiology of the spine.

Scleral hue was judged clinically, and all hues in the blue-gray scale were digitally recorded as 'blue', as compared with 'white'. The diagnosis of DI was based on clinical examination, performed by a dentist specialized in this condition.

BMD of the lumbar spine (L1–L4) was measured using a Lunar prodigy dual-energy X-ray absorptiometry (DXA) machine (Uppsala, Sweden) and a Hologic QDR 4500 system (Stockholm, Sweden). *Z*-scores stated are baseline values, based on age, and gender matched reference populations supplied by the manufacturers, and are before bisphosphonate treatment in relevant cases.

Statistical analysis

All statistical analyses were performed in the software JMP, version 10.0 (SAS Institute Inc., JMP, Cary, NC, USA). One-way ANOVA and Student's *t*-test was used for averages. Fisher's exact test was used for two-by-two contingency tables with small sample sizes, and Pearson's chi-squared test for larger contingency tables with expected cell counts >5.

RESULTS

A total of 223 individuals (164 unrelated families) were included in the study: Sillence type I $n=151$ (101 families), type II $n=1$, type III $n=29$ (26 families), and type IV $n=42$ (37 families) (Flow Chart 1a). An additional 180 relatives with OI were mentioned in referrals and charts. Based on the prevalence calculated for the pediatric cohort below, this corresponded to 56% of Swedish OI patients. In 130 families, typical OI mutations were discovered in collagen I; 56 quantitative and 74 qualitative (39 *vs* 35 qualitative in the α1- and α2-chain, respectively) (Supplementary Table 1). Out of these, 11 qualitative and 30 quantitative mutations had not previously been reported in the OI database.[17–19] Two additional mutations were

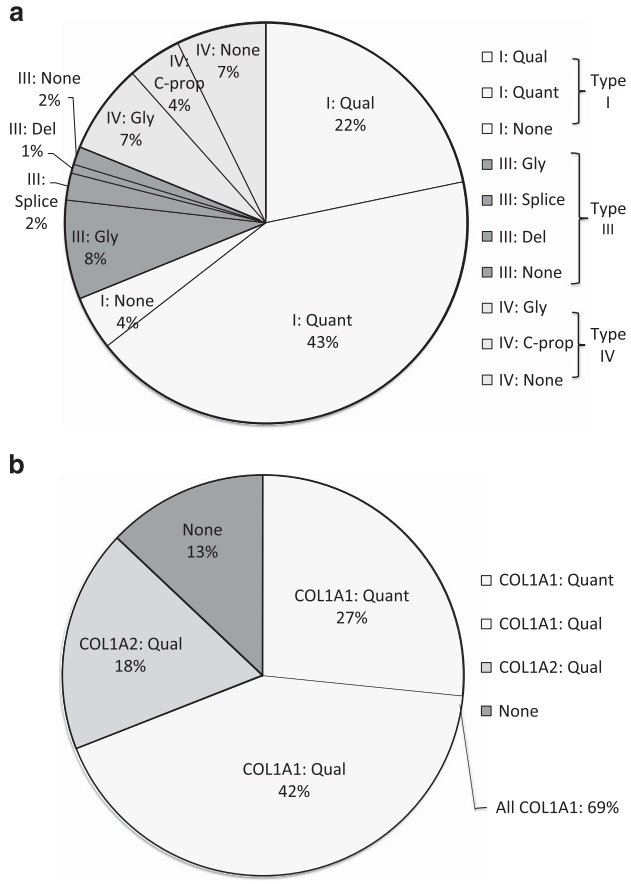


Figure 1 Genetic epidemiology of Swedish children with OI. (**a**) OI types represented by gray scale and in subsections by mutation type. (**b**) Collagen I mutation distribution by gene *vs* no mutation found. Qual = Qualitative mutations; Quant = Quantitative mutations; None = No collagen I mutation found; Gly = Helical glycine substitutions; Splice = Splice mutations; Del = Large deletion; C-prop = C-propeptide mutations.

determined through MLPA analysis of collagen I; a deletion of one copy of the entire *COL1A1* gene in an individual with OI type IA (deletion confirmed by array CGH analysis), and a partial deletion of *COL1A2* exons 1 through 15 in an individual with OI type IV. To date, exact genomic positions are not available for the mutations found through MLPA analysis. No mutation was found in 32 families.

Prevalence and genetic epidemiology of OI in the Swedish pediatric population

The data presented here on 142 children (62 girls, 80 boys) belonging to 112 families were estimated to represent the molecular background of more than 95% of the Swedish pediatric OI population (Flow Chart 1b). The gender distribution was not skewed ($P=0.14$). The prevalence of OI was 7.40/100 000, that is, 1/13500. The prevalence for types I, III, and IV was 5.16, 0.89, and 1.35/100 000, respectively. The distribution was thus 70% type I, 12% type III, and 18% type IV (Figure 1a).

Mutation analysis of collagen type I was performed in the family of 97% of these cases, and an alteration was found in 87%. Qualitative *vs* quantitative collagen I mutations were the cause in 46 *vs* 41%, and no mutations were discovered in 13%. Divided by gene, *COL1A1 vs COL1A2* harbored 69 *vs* 18% of causative mutations. Qualitative C-terminal propeptide mutations were present in 4%, and an OI-Ehler Danlos syndrome (OI-EDS) overlap phenotype in 1% (Figure 1a and b).

Type I OI ($n=99$) was caused by a quantitative mutation in 62%, a qualitative mutation in 32% (divided 50–50 for affected chain), and no mutations were found in collagen I in 6%. Type III OI ($n=17$) was associated with a glycine substitution in 65%, a splice site mutation in 18%, and no mutation was found in collagen I in 12%. OI type IV ($n=26$) was a heterogeneous group including helical glycine substitutions in 37%, C-propeptide mutations in 22%, and no mutations were found in collagen I in 37% (Figure 1a).

Coding collagen variants of unknown significance in addition to a certain OI-causing mutation

In 13 families a coding variant of unknown significance (VUS) was discovered in collagen I, and these were compared with reported variants in the ExAC, OI, and NCBI SNP databases (Supplementary Table 2) (http://www.ncbi.nlm.nih.gov/snp/).[17-19, 26] To assess the potential damaging effect of these coding alterations, the SIFT score, Polyphen score, and PhyloP score were determined for all substitutions (Supplementary Table 2).[27,28] Only one VUS, c.1601T>A (p.(Leu534Gln)) in *COL1A1*, was not previously reported in any of these databases, and in this individual a 32 nucleotide *COL1A1* deletion including a splice site was determined to cause the reported OI type I phenotype. In 10 of the families the coding VUS was found in addition to a typical OI-causing collagen I mutation. The coding VUS c.1268G>A (p.(Arg423His)) in *COL1A2*, previously reported in severe OI, was the only coding variant found in collagen I in a severely affected adopted child, and no parental DNA was thus available to assess whether the mutation was *de novo* (Individual 8, Supplementary Table 2). However, this VUS was also found in an individual that harbored a typical OI-causing mutation, and furthermore the VUS was found in this individual's healthy mother (Individual 9, Supplementary Table 2). In the remaining two families with no certain OI-causing mutation, the discovered VUS was not judged to cause an OI phenotype since the alterations were predicted to be benign.[27,28] In the three cases of a *COL1A1* null allele and an additional *COL1A1* coding VUS, the mutations were *in cis* in two instances, while it was not possible to perform segregation analysis in the third individual.

Noteworthy mutations

A child with repeated fractures and a high bone mass phenotype was found to have a *de novo COL1A1* C-propeptide cleavage site mutation, as previously reported.[25] Seven individuals, from five families, harbored *COL1A1* C-propeptide mutations. Two unrelated individuals had OI-EDS caused by mutations in *COL1A1* ($n=1$) and *COL1A2* ($n=1$). In four families, mutations were located in the eight lethal clusters of *COL1A2* described by Marini *et al*:[15] c.1937G>T (p.(Gly646Val)), c.2918G>T (p.(Gly973Val)), c.3089G>C (p.(Gly1030Ala)), and c.3106G>C (p.(Gly1036Arg)). These clusters have been reported to predict phenotype correctly in 86% of cases (Supplementary Table 1).[15]

Genotype–phenotype correlations

BMD. As expected, individuals with OI type I averaged the highest lumbar spine BMD *Z*-score at −2.66, compared with −5.05 in OI type III, and −4.28 in type IV ($n=148$, overall $P<0.0001$). The only association between BMD and gender was seen in OI type I; males had a lower average BMD *Z*-score than females (−2.90 *vs* −2.32, $P=0.0482$). BMD *Z*-score did not differ between children and adults. *Z*-scores neither differed between individuals with helical glycine mutations in α1- *vs* α2-chain (−4.43 *vs* −3.94), nor between individuals with any

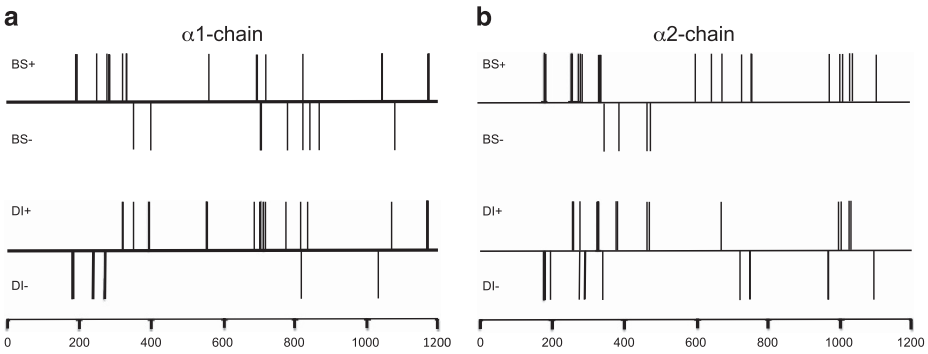


Figure 2 Dentinogenesis imperfecta and blue sclera *vs* helical location. Location of helical glycine substitutions in relation to presence (+) or absence (−) of blue sclera (BS) and dentinogenesis imperfecta (DI), from N- to C-terminal, in the (**a**) α1-chain and (**b**) α2-chain. Affected residues numbered from translation initiation.

qualitative mutation in the α1- *vs* α2-chain (−4.13 *vs* −3.91). Type I OI caused by quantitative *vs* qualitative mutations did not differ in BMD *Z*-score (−2.57 *vs* −2.99) (Tables 1a, b, and 2).

Height. Individuals with OI type I measured on average −1.08SD, compared with type III at −6.73SD and type IV at −2.59SD ($n=173$, overall $P<0.0001$). The only association between height and gender was seen in OI type I; males had a higher average height *Z*-score than females (−0.84 *vs* −1.40SD, $P=0.0265$). The only association between height and age was seen in OI type III; children had a higher average height *Z*-score than adults (−4.97 *vs* −9.38SD, $P<0.0001$). Height neither differed between individuals with helical glycine substitutions in the α1- *vs* α2-chain (−3.61 *vs* −2.87SD), nor between individuals with any qualitative mutation in the α1- *vs* α2-chain (−3.41 *vs* −2.93SD). However, individuals with type I OI caused by quantitative mutations were taller than those with qualitative mutations (−0.53 *vs* −1.67SD, $P<0.0001$) (Tables 1a, b, and 2).

Peripheral and vertebral fractures. The number of peripheral fractures per year was 0.57, 3.83, and 1.33 for types I, III, and IV respectively ($n=177$, overall $P=0.0003$). Fractures neither differed between patients with helical mutations in the α1- *vs* α2-chain (1.30 *vs* 0.92/year), nor between patients with any qualitative mutation in the α1- *vs* α2-chain (2.55 *vs* 0.97/year). OI type I caused by quantitative *vs* qualitative mutations did not differ in fracture rate ($n=0.59$ *vs* 0.58/year). Vertebral fractures were present in 71, 100, and 90% for types I, III, and IV, respectively ($n=157$, overall $P<0.0001$). There were no associations between peripheral or vertebral fractures and gender or age (children *vs* adults). However, there was a trend for peripheral fractures per year to be more frequent in children with OI types I and IV than in adults ($P=0.983$ and 0.0691), and for vertebral fractures to be more common in helical α1- *vs* α2-chain mutations ($P=0.0707$) (Tables 1a, b, and 2).

Blue sclera. Scleral hue ($n=180$) was reported as blue in 82%. There was no association between blue sclera and gender. All with a *COL1A1* null allele ($n=73$) had blue sclera, compared with 58/75 individuals with qualitative mutations ($P<0.0001$). No difference was observed concerning scleral hue when comparing helical, or any qualitative mutation in the α1- *vs* α2-chain; however, there was a trend for blue sclera being more common for α2 alterations ($P=0.120$ and 0.0511). Individuals with collagen I mutations had blue sclera more often than those without in OI types III and IV (25/42 *vs* 3/17 $P=0.0094$).

Concerning N- to C-terminal location, all 11 mutations located N-terminally of p.Gly332 were associated with blue sclera, whereas C-terminal of this point only 9 of 19 exhibited this phenotype ($P=0.0110$) (Figure 2a). In *COL1A2*, there was no correlation for N- to C-terminal helical location (Figure 2b) (Tables 1a and 2).

Dentinogenesis imperfecta. Dental status ($n=187$) was reported as clinical DI in 25%, and 72/73 individuals with DI had qualitative collagen I mutations ($P<0.0001$). An association between DI and gender was seen in OI type I; DI was more common in females than in males (22 *vs* 9%, $P=0.0258$). When comparing only OI type I cases caused by qualitative mutations, DI was also more common in females than in males (58 *vs* 26%, $P=0.0487$). No difference was observed concerning the presence of DI when comparing all helical, or any qualitative mutation the α1- *vs* α2-chain. DI was more common in severe (type III and IV) than in mild (type I) OI caused by helical glycine substitutions ($P=0.0137$). None of the 32 patients without collagen I mutations (but with data on dental status) had clinical DI, which strongly associates this phenotype with classical dominant OI ($P<0.0001$).

Concerning N- to C-terminal location, none of the nine individuals with helical glycine substitutions N-terminal of p.Gly305 in the α1-chain had DI, whereas C-terminal of this point 22/25 was affected by DI ($P<0.0001$) (Figure 2a). In the α2-chain, none of the six individuals with helical glycine substitutions N-terminal of p.Gly211 had DI, whereas C-terminal of this point 19/29 was affected by DI ($P=0.0049$) (Figure 2b) (Tables 1a and 2).

Serine mutations in α1- vs α2-chain. To compare the severity of helical mutations in the α1- *vs* α2-chains, while eliminating the effects of specific amino acid or substitution patterns in the two genes, the consequence of the most frequent amino-acid substitution, glycine to serine,[15] was analyzed (α1 $n=20$; α2 $n=20$). Individuals with serine mutations in α1-chain had a lower baseline lumbar spine BMD *Z*-score (−5.0 *vs* −3.5, $P=0.0146$) and were shorter (−3.62 *vs* −1.61SD, $P=0.0154$), compared to those with α2-chain mutations. A higher proportion of mutations caused moderate or severe phenotype (type IV or III) when located in the α1- *vs* α2-chain (13/20 *vs* 3/20; $P=0.0031$). Individuals with α1-serine substitutions had both a higher frequency of vertebral compression fractures ($P=0.0352$) and intramedullary rodding ($P=0.0069$), compared with α2 substitutions. The average number of fractures per year was 1.53 *vs* 0.64 for α1 *vs* α2 mutations ($P=0.0351$). DI was more common in α1 mutations ($P=0.0201$), and there was a trend for blue sclera to be more common in α2 mutations ($P=0.0523$) (Table 2).

DISCUSSION

Our study is the hitherto largest OI-cohort presenting genotype–phenotype correlations outside North America. Furthermore, it is the only population-based study of the genetic epidemiology encompassing the majority of an OI population. The gender distribution was slightly skewed, with a larger proportion of females than males with OI type III (Table 1a). However, this phenomenon was not seen for the pediatric cohort. OI type III is not generally considered to be more common in females, and this may well be due to a small sample size in type III OI. The spectrum of collagen I mutations causing OI was similar to other studied cohorts, and support previously presented regional models for genotype–phenotype correlations.[15,29,30]

Qualitative mutations affecting *COL1A1* are more often lethal than mutations located in *COL1A2* (36 *vs* 20%).[15] In accordance with this, our data confirmed that the clinical phenotype concerning lumbar spine BMD, height, peripheral fracture rate, the presence of vertebral compression fractures, intramedullary rodding, and OI type was more severe for serine substitutions in the α1- compared with the α2-chain.

OI-prevalence studies are often based on birth registers, and mild phenotypes will not necessarily be detected until later in life. An unknown proportion of very mild cases will most likely never come to medical attention. When the comprehensive studies by Sillence *et al*[14,31] were performed in the 1970s, molecular diagnostics was not available and the causative genes were not known. Later studies including molecular analysis are often skewed toward severe phenotypes, and cover only subsets of patients. In contrast, the pediatric prevalence and genetic epidemiological data presented here reflect the prevalence of all OI types compatible with surviving the neonatal period in modern Western society. OI types II and III have been estimated to each affect 1–2/100 000 at birth in an era before prenatal ultrasound, which has routinely been used in gestation week 18 since the early 1980s in Sweden.[14,20] Extrapolating this estimated incidence (1.50/100 000) unto the prevalence calculated for all types here, the prevalence of fetuses with any type of OI would be approximately 9.50/100 000 (ie, 1/10 500). This would be one of the higher frequencies reported in a population; however, a point prevalence at birth close to 10/100 000 is indeed what the research field has considered probable, but which has proven difficult to confirm.

There are several plausible reasons for the relatively high prevalence presented here. There is only one Swedish national pediatric OI center, and all types of OI as well as cases with bone fragility of unknown cause are assessed here, which increases chances of very mild cases being detected and included. There are detailed national registers concerning the number and age of people living in the country, and accessible national diagnosis registers of all patients seeking medical care. Furthermore, every Swedish citizen has a unique personal identification number received at birth and kept during life, which facilitates follow-up, and allows data to be extracted from registers. Finally, in the present study all patients have been molecularly analyzed, irrespective of OI phenotype.

The prevalence of OI types I, III, and IV determined here (7.40/100 000) is higher than in a nation-wide, repeatedly cited, register-based Finnish study from 2004 that estimated the prevalence to be 5.7/100 00 in the entire population. In the Finnish study, it was estimated that 41.1% of the OI population was molecularly analyzed, which to our knowledge is the most extensive study to date. However, this was a study investigating hearing loss, in which patients were invited to participate by returning a questionnaire, and the inclusion may have been skewed toward mutations causing this phenotype. In Norway, a national database estimated the prevalence of pediatric OI to be 5.9/100 000 in 2011 (personal communication).

Based on these prevalence figures in Sweden's neighboring countries, on the number of children with the ICD-10 diagnosis code for OI according to the Swedish National Patient Register, and on the design of the study, it was concluded that data presented herein represent the molecular background of more than 95% of all Swedish children with clinically significant OI.

It has been proposed that a quantitative collagen I defect should be required for an OI type I diagnosis, although 9% of helical substitutions in either chain may cause this phenotype.[15] To our knowledge, there is neither population-based data on the proportion, nor on the clinical characteristics of qualitative *vs* quantitative mutations in type I OI. Unexpectedly, our study showed that 32% of mild cases were caused by qualitative mutations. This highlights that conclusions concerning genotype cannot be drawn based solely on phenotype in future situations where, for example, mutation type could influence treatment choice. For example, there is evidence that OI type I responds better to anabolic treatment with teriparatide than more severe types (III and IV), and a hypothesis is that this could be associated with if the patient has a quantitative defect.[32] Here, individuals with qualitative type I OI were shorter than patients with a quantitative defect, which may imply that also mild qualitative defects are more disruptive to the skeletal tissue than are the quantitative alterations.

When analyzing clinical characteristics by gender it was noted that males with OI type I averaged a lower BMD *Z*-score than females; however, males with OI type I were taller, while the fracture frequency did not differ. Although the sample size is small, this further supports that BMD does not account for bone quality, which is a very important factor in OI. Furthermore, it was noted that adults with OI type III average a distinctly lower height *Z*-score than children with type III. We hypothesize that this may in part be due to treatment in these children from an early age with bisphosphonates, which was not available for the adults at early age. However, most likely the greater part of this discrepancy is explained by a progressive deviation from the mean population height during growth in severe OI.

In this cohort five unrelated patients had inconclusive clinical phenotypes, and were diagnosed to have OI through genetic analysis; three typical glycine substitutions, a *COL1A1* null-allele, and a C-propetide cleavage site mutation.[25] Furthermore, in several cases parents were diagnosed via their children. This supports referral for genetic analysis of unclear bone fragility; however, cost of sequencing is still considerable in most clinical settings.

In a cohort of 83 English patients with OI types I–IV the mutation rate discovered by Sanger sequencing of collagen I was 75 or 84%, depending on inclusion criteria used.[30] This is similar to our study (80% of families *vs* 87% in the pediatric cohort). Furthermore, in the English cohort 5/83 patients were found to have a VUS, which is equivalent to the frequency detected in collagen I in our study. A VUS is sometimes reported to cause OI, presumably because it is the only variant found when sequencing type I collagen; however, without functional data or segregation analysis this is not a justifiable conclusion.

Most of the qualitative mutations (85%) found here have previously been reported, despite that only a minority of theoretically possible glycine substitutions have been described.[15] It has been shown that frequently repeated mutations often are associated with CpG dinucleotides, which have been described as mutational hot spots.[33] Here, seven mutations were found twice in *COL1A1* (ie, in 14 unrelated families total), and two mutations were found in three families each in *COL1A2* (ie, in 6 unrelated families total). In 10/14 of the families, the *COL1A1* mutations were *de novo* and caused severe OI. The families

with mild OI in several generations might have been unknowingly related; consanguinity was not determined as part of this study. It is not known why other locations on the contrary lack known substitutions. By chance, some substitutions may simply not have been reported yet, whereas others may cause a mild phenotype that is mistaken for ordinary osteoporosis. Sequencing is still costly and consequently there is a sequencing bias, causing severe forms to be more frequently referred for genetic testing. Furthermore, it has been hypothesized that some of the missing mutations may be embryonically lethal. In contrast to the qualitative mutations, a minority of the quantitative mutations (46%) found here had previously been reported to the OI database.[17–19] This is likely also due to a sequencing bias, and that there are virtually an infinite number of insertions and deletions that have a *COL1A1* null allele outcome.

Rauch *et al*[16] reported in 2010 that N-terminal helical mutations in *COL1A1,* but not *COL1A2,* were associated with blue sclerae, and that C-terminal helical mutations were associated with DI in both genes.[16] Here, the same cutoff points were used, since the distributions observed supported these boundaries, also enabling comparison between the studies. The same correlations reported previously were observed here; however, the frequency of DI in individuals with qualitative mutations was lower in the present study. A hypothesis is that this reflects the higher frequency of individuals with type I OI caused by a qualitative mutation here. Unexpectedly, DI was found to be more common in females than in males with mild type I OI. This may be an effect of small sample size, and larger studies are needed to support or reject this association. According to our study, DI is more common in severe (types III and IV) than in mild (type I) OI caused by helical glycine substitutions. It is noteworthy that none of the patients without collagen I mutations reported here have clinical DI, which is congruent with previous case reports of non-collagen related OI. To our knowledge, there is no publication on the prevalence of DI in a large cohort of non-collagenous OI. A primary collagen defect may be mandatory, but not sufficient to develop DI in OI.

Eleven types of non-collagen OI have been described since 2006, and these individuals are often reported to have a moderate-to-severe phenotype.[1–13] On the contrary, here 15/32 families with no mutation in collagen I had a typical type I phenotype, and 25 additional families were excluded on the basis of a negative sequencing result in combination with an unclear clinical phenotype. These individuals may have a *COL1A1* null allele caused by a non-exonic mutation missed here; however, another OI-related gene might be causative in some instances. Historically, there has been an inclination to investigate severe phenotypes more thoroughly, and the present view of non-collagen OI may thus change by future studies. Recently heterozygous *WNT1* mutations were associated with early onset osteoporosis, while homozygous mutations in the same gene were associated with severe OI.[11] Furthermore, a form of X-linked osteoporosis due to mutations in *PLS3* has been described in families with male osteoporosis, and in males with suspected OI type I tested negative for collagen I mutations.[13] Severe OI has also recently been associated with homozygous deletions of *CREB3L* in two affected sibs, whereas heterozygous family members exhibited certain features of OI type.[12] Analogous mutations in these and other genes may well cause mild OI in more cases than previously thought. On the basis of our data, 40 families (15 included and 25 excluded families) could be affected by this type of genetic alteration causing an OI-like skeletal phenotype, which would correspond to 21% (40/189) of all analyzed families.

CONCLUSION

In conclusion, there are no population-based studies including molecular analysis concerning the prevalence of OI. Here, the *COL1A1* and *COL1A2* genes were analyzed in the families of what is estimated to be more than 95% of Swedish children with OI, and an alteration was found in 87%. The prevalence of OI types I, III, and IV was determined to be 7.4/100 000, which is higher than in many previous studies, but close to the figure experts in the field have anticipated. Notably, 32% of type I OI was due to qualitative mutations. Genotype–phenotype analysis confirmed previous studies on severity of disease *vs* chain affected and mutation type, as well as solidified a previous report concerning N- to C-terminal chain position *vs* presence of DI and blue sclera.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all of the individuals with OI who participated in this study. Furthermore, we thank Anna-Lena Johansson, Elin Carlsson, and Catharina Kumlien for skillful technical assistance. Funding was received from the Swedish research council, and from the Crown-princess Lovisa, Norrbacka Eugenia, Promobilia, RBU, and Sunnerdahls foundations.

1 Barnes AM, Chang W, Morello R *et al*: Deficiency of cartilage-associated protein in recessive lethal osteogenesis imperfecta. *N Engl J Med* 2006; **355**: 2757–2764.
2 Baldridge D, Schwarze U, Morello R *et al*: CRTAP and LEPRE1 mutations in recessive osteogenesis imperfecta. *Hum Mutat* 2008; **29**: 1435–1442.
3 van Dijk FS, Nesbitt IM, Zwikstra EH *et al*: PPIB mutations cause severe osteogenesis imperfecta. *Am J Hum Genet* 2009; **85**: 521–527.
4 Christiansen HE, Schwarze U, Pyott SM *et al*: Homozygosity for a missense mutation in SERPINH1, which encodes the collagen chaperone protein HSP47, results in severe recessive osteogenesis imperfecta. *Am J Hum Genet* 2010; **86**: 389–398.
5 Alanay Y, Avaygan H, Camacho N *et al*: Mutations in the gene encoding the RER protein FKBP65 cause autosomal-recessive osteogenesis imperfecta. *Am J Hum Genet* 2010; **86**: 551–559.
6 Homan EP, Rauch F, Grafe I *et al*: Mutations in SERPINF1 cause osteogenesis imperfecta type VI. *J Bone Miner Res* 2011; **26**: 2798–2803.
7 Kelley BP, Malfait F, Bonafe L *et al*: Mutations in FKBP10 cause recessive osteogenesis imperfecta and Bruck syndrome. *J Bone Miner Res* 2011; **26**: 666–672.
8 Puig-Hervas MT, Temtamy S, Aglan M *et al*: Mutations in PLOD2 cause autosomal-recessive connective tissue disorders within the Bruck syndrome–osteogenesis imperfecta phenotypic spectrum. *Hum Mutat* 2012; **33**: 1444–1449.
9 Semler O, Garbes L, Keupp K *et al*: A mutation in the 5'-UTR of IFITM5 creates an in-frame start codon and causes autosomal-dominant osteogenesis imperfecta type V with hyperplastic callus. *Am J Hum Genet* 2012; **91**: 349–357.
10 Shaheen R, Alazami AM, Alshammari MJ *et al*: Study of autosomal recessive osteogenesis imperfecta in Arabia reveals a novel locus defined by TMEM38B mutation. *J Med Genet* 2012; **49**: 630–635.
11 Laine CM, Joeng KS, Campeau PM *et al*: WNT1 mutations in early-onset osteoporosis and osteogenesis imperfecta. *N Engl J Med* 2013; **368**: 1809–1816.
12 Symoens S, Malfait F, DH S *et al*: Deficiency for the ER-stress transducer OASIS causes severe recessive osteogenesis imperfecta in humans. *Orphanet J Rare Dis* 2013; **8**: 154.
13 van Dijk FS, Zillikens MC, Micha D *et al*: PLS3 mutations in X-linked osteoporosis with fractures. *N Engl J Med* 2013; **369**: 1529–1536.
14 Sillence DO, Senn A, Danks DM: Genetic heterogeneity in osteogenesis imperfecta. *J Med Genet* 1979; **16**: 101–116.
15 Marini JC, Forlino A, Cabral WA *et al*: Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum Mutat* 2007; **28**: 209–221.
16 Rauch F, Lalic L, Roughley P, Glorieux FH: Genotype-phenotype correlations in nonlethal osteogenesis imperfecta caused by mutations in the helical domain of collagen type I. *Eur J Hum Genet* 2010; **18**: 642–647.
17 Dalgleish R: The human type I collagen mutation database. *Nucleic Acids Res* 1997; **25**: 181–187.
18 Dalgleish R: The Human Collagen Mutation Database 1998. *Nucleic Acids Res* 1998; **26**: 253–255.
19 Dalgleish R. http://www.le.ac.uk/ge/collagen/).

20 Steiner RD, Adsit J, Basel D: COL1A1/2-related osteogenesis imperfecta; in: Pagon RA, Adam MP, Bird TD, Dolan CR, Fong CT, Stephens K (eds): *GeneReviews*. Seattle, USA: University of Washington, 1993.
21 Astrom E, Soderhall S: Beneficial effect of bisphosphonate during five years of treatment of severe osteogenesis imperfecta. *Acta Paediatr* 1998; **87**: 64–68.
22 http://www.scb.se/en_/Hitta-statistik/Statistik-efter-amne/Befolkning/Befolkningens-sammansattning/Befolkningsstatistik/25788/25795/Helarsstatistik—Riket/26040/, 2014.
23 http://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish, 2014.
24 Korkko J, Ala-Kokko L, De Paepe A, Nuytinck L, Earley J, Prockop DJ: Analysis of the COL1A1 and COL1A2 genes by PCR amplification and scanning by conformation-sensitive gel electrophoresis identifies only COL1A1 mutations in 15 patients with osteogenesis imperfecta type I: identification of common sequences of null-allele mutations. *Am J Hum Genet* 1998; **62**: 98–110.
25 Lindahl K, Barnes AM, Fratzl-Zelman N *et al*: COL1 C-propeptide cleavage site mutations cause high bone mass osteogenesis imperfecta. *Hum Mutat* 2011; **32**: 598–609.
26 Exome Aggregation Consortium (ExAC): Cambridge, MA. Available at http://exac.broadinstitute.org (last accessed March 2015).
27 Adzhubei I, Jordan DM, Sunyaev SR: Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013; Chapter 7: Unit7 20.
28 Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**: 1073–1081.
29 Hartikka H, Kuurila K, Korkko J *et al*: Lack of correlation between the type of COL1A1 or COL1A2 mutation and hearing loss in osteogenesis imperfecta patients. *Hum Mutat* 2004; **24**: 147–154.
30 Pollitt R, McMahon R, Nunn J *et al*: Mutation analysis of COL1A1 and COL1A2 in patients diagnosed with osteogenesis imperfecta type I-IV. *Hum Mutat* 2006; **27**: 716.
31 Sillence DO, Rimoin DL: Classification of osteogenesis imperfect. *Lancet* 1978; **1**: 1041–1042.
32 Orwoll ES, Shapiro J, Veith S *et al*: Evaluation of teriparatide treatment in adults with osteogenesis imperfecta. *J Clin Invest* 2014; **124**: 491–498.
33 Schorderet DF, Gartler SM: Analysis of CpG suppression in methylated and nonmethylated species. *Proc Natl Acad Sci USA* 1992; **89**: 957–961.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)