CrossMark

ORIGINAL ARTICLE

Gene mutation spectrum and genotype-phenotype correlation in a cohort of Chinese osteogenesis imperfecta patients revealed by targeted next generation sequencing

Y. Liu[1] · Asan[2,3] · D. Ma[1] · F. Lv[1] · X. Xu[1] · J. Wang[1] · W. Xia[1] · Y. Jiang[1] · O. Wang[1] ·
X. Xing[1] · W. Yu[4] · J. Wang[2,3] · J. Sun[2,3] · L. Song[2,3] · Y. Zhu[2,3] · H. Yang[5,6] ·
J. Wang[5,6] · M. Li[1]

Abstract

Summary The achievement of more accurate diagnosis would greatly benefit the management of patients with osteogenesis imperfecta (OI). In this study, we present the largest OI sample in China as screened by next generation sequencing. In particular, we successfully identified 81 variants, which included 45 novel variants. We further did a genotype-phenotype analysis, which helps make a better understanding of OI.

Introduction This study aims to reveal the gene mutation spectrum and the genotype-phenotype relationship among Chinese OI patients by next generation sequencing (NGS).

Methods We developed a NGS-based panel for targeted sequencing of all exons of 14 genes related to OI, and performed diagnostic gene sequencing for a cohort of 103 Chinese OI patients from 101 unrelated families. Mutations identified by NGS were further confirmed by Sanger sequencing and co-segregation analysis.

Results Of the 103 patients from 101 unrelated OI families, we identified 79 mutations, including 43 novel mutations (11 frameshift, 17 missense, 5 nonsense, 9 splice site, and 1 chromosome translocation) in 90 patients (87.4%). Mutations in genes encoding type I collagen, *COL1A1* ($n = 37$), and *COL1A2* ($n = 29$) accounts for 73.3% of all molecularly diagnosed patients, followed by *IFITM5* ($n = 9$, 10%), *SERPINF1* ($n = 4$, 4.4%), *WNT1* ($n = 4$, 4.4%), *FKBP10* ($n = 3$, 3.3%), *TMEM38B* ($n = 3$, 3.3%), and *PLOD2* ($n = 1$, 1.1%). This corresponds to 75 autosomal dominant inherited (AD) OI patients and 15 autosomal recessive (AR) inherited patients. Compared with AD inherited OI patients, AR inherited patients had lower bone mineral density (BMD) at spine ($P = 0.05$) and less frequent blue sclera ($P = 0.001$). Patients with type I collagen qualitative defects had lower femoral neck BMD Z-score ($P = 0.034$) and were shorter compared with patients with type I collagen quantitative defects ($P = 0.022$).

Yi Liu and Asan contributed equally to this work.

✉ M. Li
  limeilzh@sina.com

  Asan
  asan@genomics.cn

[1] Department of Endocrinology, Key Laboratory of Endocrinology, Ministry of Health, Peking Union Medical College Hospital, Chinese Academy of Medical Science, Beijing 100730, China

[2] Tianjin Medical Laboratory, BGI-Tianjin, BGI-Shenzhen, Tianjin 300308, China

[3] Binhai Genomics Institute, BGI-Tianjin, Tianjin 300308, China

[4] Department of Radiology, Peking Union Medical College Hospital, Chinese Academy of Medical Science, Beijing 100730, China

[5] BGI-Shenzhen, Shenzhen 518083, China

[6] James D. Watson Institute of Genome Sciences, Hangzhou 310058, China

Conclusion We revealed the gene mutation spectrum in Chinese OI patients, and novel mutations identified here expanded the mutation catalog and genotype and phenotype relationships among OI patients.

Introduction

Osteogenesis imperfecta (OI) is a rare congenital skeletal dysplasia that is characterized by increasing bone fragility, recurrent fracture, and subsequent growth retardation. OI patients also share some extraskeletal symptoms, such as blue sclera, hearing deficits, dentinogenesis imperfecta (DI), and valvular heart disease [1].

The widely used Sillence classification divides OI into four types on the basis of clinical and radiological features [2]. In 2014, van Dijk then made some modification to the phenotype assessment [3]. Type I OI is the mildest type, and patients usually have variable fractures and a normal stature. Type II OI is fatal, wherein patients with this type of OI have difficulty surviving perinatal period due to skeletal deformities and respiratory compromise. Type III is the most severe type in patients, in terms of those that are able to survive, as patients often sustain from multiple fractures and have a severely short stature. Patients with moderate skeletal deformities are classified into type IV OI. Finally, type V is distinguished by apparent hypercallus formation and intraosseous membrane calcification [3].

Although the diagnosis of OI largely relies on clinical evaluation and pedigree analysis, molecular analysis can help achieve higher levels of diagnostic accuracy. Currently, it is known that about 85 to 90% OI patients are caused by autosomal dominant (AD) mutations of genes that encode type I collagen, *COL1A1*, and *COL1A2* [1]. Now, with the development of molecular diagnosis, a total of 20 genes have been identified as the causative candidate gene of OI, including *IFITM5, SERPINF1, CRTAP, P3H1, PPIB, SERPINH1, FKBP10, PLOD2, WNT1, CREB3L1, WNT1, SP7, CREB3L1,* and *MBTPS2* [4–14]. However, the absence of gene-specific symptoms and the large number of candidate genes hamper the molecular diagnosis. In addition, the current gold standard method, Sanger sequencing, is time-consuming and costly, especially when applied to large genes such as *COL1A1* and *COL1A2* [15]. As a result, compared to traditional methods, next generation sequencing (NGS) is more cost-effective and more efficient in its ability to identify the causal mutation of diseases with genetic heterogeneity [16].

Additionally, although over 1600 pathogenic variants had been identified, the genotype-phenotype correlations in OI patients have only partially been elucidated. Previous studies revealed that phenotypic severity was associated with affected collagen helical location and types of amino acid substitution. However, these genotype-phenotype studies focused more on AD inherited OI and less is known about autosomal recessive (AR) inherited OI. Moreover, genotype-phenotype studies are elusive and are often limited to a relatively small size in China.

Herein, we developed a novel panel of genes related to OI and other skeletal dysplasia disorders. Given that most reports on the mutation spectrum of OI have been restricted to Western populations and only a few OI-related genetic studies have been conducted in China, we applied the panel in a cohort of Chinese OI patients not only to prove the efficiency of NGS in the genetic analysis of OI but also to gain more population-specific insight into OI. Furthermore, by analyzing the mutation spectrum and bone morphologic findings of OI patients, we also aim to investigate the genotype-phenotype relationship.

Materials and methods

Patients

The study includes 103 patients with OI who were evaluated at Peking Union Medical College Hospital (PUMCH) from 2010 to 2014. Clinical diagnoses of OI were based on the following: a history of more than one fracture under minor trauma and an age-adjusted and sex-adjusted BMD Z-score of less than −1.0 or more for either lumbar spine or femoral neck or with a family history of OI or an adjusted areal BMD Z-score of −2·0 or less irrespective of a history of fractures [2]. Patients were further classified into OI subtypes based on Sillence classification[2]. Patients were regarded as type V OI if they had history of hyperplastic callus formation or had radiological apparent calcification of forearm intraosseous membrane [17]. Physical examination revealed hypermobile joints and mild skin hyperextensibility. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethic Committee of PUMCH. Written informed consent was obtained from parents or legal guardians of children younger than 18 years old.

Design of targeted capture panel

A capture panel of targeted DNA was designed by selecting 14 well-known OI-related genes from previous studies and OMIM, which are *COL1A1, COL1A2, IFITM5, SERPINF1, CRTAP, LEPRE1, PPIB, FKBP10, SERPINH1, SP7, PLOD2, TMEM38B, BMP1,* and *WNT1*. The panel also included a cluster of 708 possible candidate genes with expression patterns that have been associated with congenital rickets, such as *FGF23, PHEX, DMP1, ENP11, CLCN5,* and *SLC34A3* [18]. Genes related to cartilage problems were also included. The targeted panel included a total of 722 genes, which compromised 9712 exons.

Paneled exome sequencing

A minimum of 3 μg of genomic DNA was extracted from each blood sample and was sheared into 200-bp fragments to construct the DNA library. The DNA fragment was ligated to illumine adaptors with barcoded oligonucleotides using Klenow exonuclease. Pooled, barcoded libraries were clonally amplified using the Illumina systems. The targeted DNA was captured, washed, and recovered using streptavidin-coated magnetic beads. Captured libraries were then subjected to NGS using an Illumina Solexa HiSeq 2000 sequencer as 100-bp paired-end reads, following the manufacturer's protocol.

In-depth bioinformatics analysis

Sequence reads acquired from raw data were converted to qseq, which were aligned to the reference human genome using the SOAPaligner program. Single-nucleotide polymorphisms were initially identified by the SOAPsnp program (http://soap.genomics.org.cn/) after filtering polymerase chain reaction duplicates using SAMtools (http://samtools.sourceforge.net/). Sequence variants were reported following the Human Genome Variation nomenclature guidelines (http://www.hgvs.org/mut-nomen/). Insertions or deletions were analyzed using BWA (http://bio-bwa.sourceforge.net/) and GATK programs (https://www.broadinstitute.org/gatk/). Variants were filtered with dbSNP135, 1000 Genome, and HapMap databases as well as an internal control database from the Beijing Genomics Institute. PoplyPhen-2, SIFT (score <0.05), and PhyloP (score >2.5) were used to assess missense variants. Mutations known to cause osteogenesis imperfecta were identified by searching against collagen-related disease database (http://www.le.ac.uk/ge/collagen).

Sanger sequencing validation

The genomic DNA was extracted from peripheral blood using Qiagen Midi Kit (Qiagen, Hilden, Germany). All of the candidate mutations were confirmed by Sanger sequencing following standard protocol. Primers were designed using Primer3 (http;//www.primer3.ut.ee, version 4.0). To ensure the quality of Sanger sequencing, the amplicons were designed to have a boundary around 100 bp away from the mutation. Then, the amplicons (~400 bp) were Sanger sequenced on Applied Biosystems 3730xl capillary sequencer (Applied Biosystems Inc., Foster City, CA, USA) using standard method.

Evaluation of phenotype

The medical history was collected, and physical examination was completed in detail. The previous fractures were documented including site, degree of trauma, frequency, and time of the initiation. Body height was measured with a Harpenden stadiometer and was adjusted to age- and sex-specific Z-scores on the basis of reference data of the Chinese National Centers for Disease Control and Prevention [19]. BMD at lumbar spine and proximal femur was measured by dual-energy x-ray absorptiometry (DXA) (Lunar, GE, USA). BMD phantom scan was measured daily by a DXA instrument and detected no significant machine drifts during the 5-year study. Areal BMD values were converted to age and sex-specific Z-scores using data of previous studies [20, 21].

Statistical analysis

Data that was normally distributed such as age, BMD, and height were presented as mean ± standard deviation, while those of abnormal distribution were expressed as median and quartiles. Difference between two groups (for example, age, gender, inheritance pattern, quality versus quantity defect of type I collagen) was analyzed by Student's t test. Differences between the three groups (for example, Sillence classification, and genotype of OI) were evaluated using analysis of variance. Qualitative data was compared by the chi-square test. Analysis was also performed by Mann-Whitney U test as appropriate. The statistical analyses were performed using SPSS 21.0 (SPSS Inc., Chicago, IL, USA). $P < 0.05$ was considered statistically significant.

Results

Phenotypes of OI patients

We included 103 patients from 101 unrelated non-consanguineous families. The patients ranged from neonates to 40 years old at assessment, and their average age was 9.9 years old at diagnosis of OI. According to Sillence classification, our cohort includes 29 type I, 39 type III, and 35 type IV OI patients. Sixty-six patients were male, 37 patients were female (Table 1). A positive family history was reported in 23 patients. Type I OI patients had the highest BMD while type III OI patients had the lowest BMD at both lumbar spine and femoral neck (Table 1). Compared with children OI, height Z-scores of adult patients were significantly lower (Z-score −13.6 ± 11.4 in adult versus −2.9 ± 4.1, $P = 0.012$, respectively). In type III OI, male patients were found to have lower lumbar spine BMD (Z-score −4.4 ± 1.2) than female patients (Z-score − 3.3 ± 1.7, $P = 0.046$, respectively). Femoral BMD Z-scores were higher in adult than in children (Z-score −6.6 ± 3.5 in children versus −3.5 ± 2.1 in adults, $P = 0.041$, respectively).

Table 1 Phenotypes of 103 Chinese OI patients

	Number	Fracture frequency (number/year)	LS BMD Z-score (mean ± SD)	FN BMD Z-score (mean ± SD)	Height Z-score (mean ± SD)	Blue sclera *n* positive (percentage)	DI *n* positive (percentage)
Type I	29	1.23 ± 1.26	−0.8 ± 1.4	−1.9 ± 1.5	−0.5 ± 1.7	17 (58.6%)	3 (10.3%)
Other type	74	2.00 ± 2.17	−3.3 ± 1.7	−5.2 ± 1.7	−5.3 ± 6.9	46 (62.2%)	13 (17.6%)
P value		0.09	0.151	**0.003**	**<0.0001**	0.658	0.279
Type I	29	1.23 ± 1.26	−0.8 ± 1.4	−1.9 ± 1.5	−0.5 ± 1.7	17 (58.6%)	3 (10.3%)
Type III	39	2.54 ± 2.60	−4.0 ± 1.4	−6.1 ± 3.5	−8.0 ± 8.2	27 (69.2%)	9 (23.1%)
Type IV	35	1.39 ± 1.46	−2.6 ± 1.7	−4.4 ± 2.6	−2.6 ± 3.3	19 (54.2%)	6 (17.1%)
P value		**0.01**	**<0.0001**	**<0.0001**	**<0.0001**	0.357	0.463
Female	36	1.97 ± 1.55	−2.3 ± 1.8	−4.3 ± 4.0	−4.7 ± 4.0	22 (61.1%)	7 (19.4%)
Male	67	1.69 ± 2.20	−2.7 ± 2.0	−4.3 ± 2.7	−4.2 ± 2.5	41 (61.2%)	11 (16.4%)
P value		0.517	0.354	0.940	0.317	1.000	0.785
Adult	11	1.28 ± 1.09	−2.7 ± 1.7	−3.0 ± 1.9	−13.6 ± 11.4	6 (54.5%)	4 (36.3%)
Children	92	1.85 ± 2.08	−2.6 ± 2.0	−4.5 ± 3.3	−2.8 ± 4.0	57 (62.0%)	14 (15.2%)
P value		0.367	0.904	0.184	**0.012**	0.745	0.187
Type I							
Female	10	1.69 ± 1.70	−0.6 ± 1.1	−1.7 ± 1.7	−0.8 ± 0.8	8 (80.0%)	1 (10.0%)
Male	19	0.95 ± 0.87	−1.0 ± 1.5	−2.0 ± 1.4	−0.3 ± 2.0	9 (47.3%)	2 (10.5%)
P value		0.146	0.533	0.569	0.479	0.126	1.000
Adult	2	0.34 ± 0.35	−0.5 ± 0.7	−2.1 ± 1.7	−1.6 ± 0.9	1 (50.0%)	0 (0%)
Children	27	1.29 ± 1.29	−0.9 ± 1.4	−1.9 ± 1.5	−0.4 ± 1.7	16 (59.2%)	3 (11.1%)
P value		0.317	0.706	0.876	0.375	1.000	1.0000
Type III							
Female	16	2.56 ± 1.56	−3.3 ± 1.7	−6.5 ± 4.9	−7.7 ± 9.0	10 (62.5%)	4 (25.0%)
Male	23	2.46 ± 3.13	−4.4 ± 1.2	−5.9 ± 2.2	−8.3 ± 7.0	17 (73.9%)	5 (21.7%)
P value		0.814	**0.046**	0.668	0.826	0.722	1.000
Adult	7	1.76 ± 1.10	−3.5 ± 1.5	−3.5 ± 2.1	−18.3 ± 11.6	5 (71.4%)	4 (57.1%)
Children	32	2.72 ± 1.81	−4.1 ± 1.4	−6.6 ± 3.5	−5.4 ± 4.7	22 (71.8%)	5 (15.6%)
P value		0.386	0.322	**0.041**	**0.026**	1.000	**0.041**
Type IV							
Female	10	1.12 ± 0.74	−2.9 ± 1.0	−3.8 ± 2.0	−4.1 ± 3.6	4 (40%)	2 (20%)
Male	25	1.49 ± 1.52	−2.5 ± 1.9	−4.6 ± 2.8	−2.0 ± 3.0	15 (60%)	4 (16%)
P value		0.497	0.497	0.440	0.497	0.454	1.000
Adult	2	0.52 ± 0.17	−2.6 ± 0.8	−2.6 ± 1.8	−8.9 ± 2.3	0 (0%)	0 (0%)
Children	33	1.45 ± 1.38	−2.6 ± 1.8	−4.5 ± 2.6	−2.2 ± 2.9	19 (57.6%)	6 (18.2%)
P value		0.356	1.000	0.344	**0.030**	0.202	1.000

The *P* values refer to the difference in BMD, height, blue sclera, and dentinogenesis imperfecta between each subgroup Bold values indicate significance.

BMD, bone mineral density; LS, lumbar spine; FN, femoral neck DI dentinogenesis imperfecta

Next generation sequencing summary

The sequencing yield was on average 288× per sample. On average, 547,696 of the reads mapped to the intended targets and 98.79% of the targets had at least 20 reads per base, with an average of 200 reads per base. In order to test the validity, we also tested ten patients with previous sequencing by Sanger analysis. The variation in these patients was identical between NGS and Sanger sequencing.

Identification of mutation

Definite genetic diagnosis was achieved in 90 out of 103 patients. Among the 90 patients, a total of 79 mutations were identified (15 frameshift, 6 nonsense, 40 missense, 15 splice site, 1 new start codon, 1 chromosome translocation, 1 whole gene deletion), 43 of which were novel variants that had not previously been reported (11 frameshift, 5 nonsense, 17 missense, 9 splice site, 1 chromosome translocation; shown

Table 2 Novel variants detected by next generation sequencing

Family no.	Mutated gene	Transcript	Sillence classification	Nucleotide change	Predicted amino acid change	Mutation type	Mutation effect	Exon or intron	ExAC frequency	PhyloP primates	PhyloP vertebrates	PhyloP placental mammals	SIFT score	Polyphen2HumVar score
1	*COL1A1*	NM_000088.3	I	c.67C>T	p.Q23*	Substitution	Nonsense	Exon 1	—	0.313	243	2.263	—	—
2	*COL1A1*	NM_000088.3	I	c.143dupA	p.H48Qfs*2	Duplication	Frameshift	Exon 2	—	—	—	—	—	—
3	*COL1A1*	NM_000088.3	I	c.148C>T	p.R50*	Substitution	Nonsense	Exon 2	—	0.555	0.46	0.025	—	—
4	*COL1A1*	NM_000088.3	IV	c.299-2A>G		Substitution	Splice	Intron 2	—	0.454	1.612	2.08	—	—
5	*COL1A1*	NM_000088.3	IV	c.588+1G>T		Substitution	Splice	Intron 7	—	0.655	5.293	2.547	—	—
6	*COL1A1*	NM_000088.3	I	c.643-13_662delCTATCTTTTCTAGGGTCCCATGGGTCCCC	p.A362fs*?	Deletion	Frameshift	Exon 9	—	—	—	—	—	—
7	*COL1A1*	NM_000088.3	IV	c.695delA	p.D232Vfs*33	Deletion	Frameshift	Exon 9	—	0.459	4.192	1.856	—	—
8	*COL1A1*	NM_000088.3	I	c.1661delG	p.G54Afs*26	Deletion	Frameshift	Exon 24	—	0.655	5.259	2.22	—	—
9	*COL1A1*	NM_000088.3	I	c.1667delC	p.556Lfs*24	Deletion	Frameshift	Exon 24	—	0.655	2.363	2.22	—	—
10	*COL1A1*	NM_000088.3	III	c.1822-2delA		Substitution	Splice	Intron 26	—	0.454	4.501	1.795	—	—
11	*COL1A1*	NM_000088.3	III	c.1822-2A>T		Substitution	Splice	Intron 26	—	0.454	4.501	1.795	—	—
12	*COL1A1*	NM_000088.3	IV	c.1922G>C	p.G641A	Substitution	Missense	Exon 28	—	0.65	5.58	2.343	0.01	0.998
13	*COL1A1*	NM_000088.3	IV	c.2250delC	p.G752Vfs*14	Deletion	Frameshift	Exon 33	—	0.563	1.409	2.424	—	—
14	*COL1A1*	NM_000088.3	I	c.2280T>A	p.D760E	Substitution	Missense	Exon 33	—	0.46	1.406	0.204	0.07	0.991
15	*COL1A1*	NM_000088.3	I	c.2450delC	p.P817Lfs*291	Deletion	Frameshift	Exon 36	—	0.462	5.595	2.357	—	—
16	*COL1A1*	NM_000088.3	IV	c.3262-1G>A		Substitution	Splice	Intron 45	—	0.462	5.703	2.456	—	—
17	*COL1A1*	NM_000088.3	III	c.3540_3541insC	p.G1181Rfs*39	Insertion	Frameshift	Exon 49	—	—	—	—	—	—
18	*COL1A2*	NM_000089.3	I	c.478G>A	p.G160R	Substitution	Missense	Exon 10	—	0.655	5.804	2.814	0	1.000
19	*COL1A2*	NM_000089.3	III	c.568G>C	p.G190R	Substitution	Missense	Exon 12	—	0.65	3.915	2.88	0	1.000
20	*COL1A2*	NM_000089.3	III	c.875G>A	p.G292D	Substitution	Missense	Exon 17	—	0.655	6.291	2.882	0	0.999
21	*COL1A2*	NM_000089.3	I	c.1586G>A	p.G529D	Substitution	Missense	Exon 27	—	0.655	6.128	2.7	0	1.000
22	*COL1A2*	NM_000089.3	IV	c.1748G>C	p.G583A	Substitution	Missense	Exon 30	—	0.46	5.984	2.564	0	1.000
23	*COL1A2*	NM_000089.3	III	c.2324G>A	p.G775E	Substitution	Missense	Exon 38	—	0.655	6.24	2.824	0	0.998
24	*COL1A2*	NM_000089.3	III	c.2330G>A	p.R777H	Substitution	Missense	Exon 38	0	0.655	4.415	2.824	0.04	0.986
25	*COL1A2*	NM_000089.3	I	c.2341G>A	p.G781S	Substitution	Missense	Exon 38	—	0.655	6.24	2.824	—	0.995
26	*COL1A2*	NM_000089.3	III	c.3814T>C	p.C1272R	Substitution	Missense	Exon 51	—	0.533	5.027	2.225	0	0.999
27	*COL1A2*	NM_000089.3	IV	chr5:146294373 .chr7:94038710		Deletion	Whole gene deletion		—	—	—	—	—	—
28	*TMEM38B*	NM_018112.2	IV	c.455-7T>G		Substitution	Splice	Intron 3	—	-0.333	0.165	0.325	—	—
		NM_018112.2	IV	c.507G>A	p.W169*	Substitution	Nonsense	Exon 4	—	0.655	3.621	2.683	—	—
29		NM_018112.2	IV	c.507G>A	p.W169*	Substitution	Nonsense	Exon 4	—	0.655	3.621	2.683	—	—
30	*FKBP10*	NM_021939.3	III	c.343C>T	p.R115*	Substitution	Nonsense	Exon 2	0	0.561	0.304	1.103	—	—
		NM_021939.3	IV	c.1085_1086delC	p.V363Cfs*2	Deletion	Frameshift	Exon 7	—	0.655	5.913	2.585	—	—
31	*FKBP10*	NM_021939.3	IV	c.918-3C>G		Substitution	Splice	Intron 5	—	0.561	5.926	2.709	—	—
32	*PLOD2*	NM_000935.2/NM_..._18	IV	c.879C>G	p.Y293*	Substitution	Nonsense	Exon 5	—	-0.258	0.871	0.724	—	—
				c.503-2A>G		Substitution	Splice	Intron 4	0	0.533	4.502	1.89	—	—
33	*SERPINF1*	NM_002615.5	IV	c.1138C>T	p.R380C	Substitution	Missense	Exon 11	0	0.655	5.819	2.746	0.04	0.999
			III	c.271_279dupGCCCTCT	p.Ala91_Ser93dup	Substitution	Frameshift	Exon3	0	—	—	—	—	—
34	*SERPINF1*	NM_002615.5	III	c.184G>A	p.G62S	Substitution	Missense	Exon 3	0	0.65	4.728	2.584	0	1.000
				c.283+1G>T		Substitution	Splice	Intron 3	—	0.65	4.728	2.584	—	1.000
35	*SERPINF1*	NM_002615.5	III	c.498_499delCA	p.R167Sfs*35	Deletion	Frameshift	Exon 5	—	—	—	—	—	—
			III	c.1202_1203delAC	p.T401Rfs*?	Deletion	Frameshift	Exon 8	—	0.533	3.091	2.023	0.01	0.020
36	*WNT1*	NM_005430.3	III	c.110T>C	p.I37T	Substitution	Missense	Exon 2	—					
			III	c.505G>T	p.G169C	Substitution	Missense	Exon 3	0	0.655	5.585	2.574	0	1.000

Table 2 (continued)

Family no.	Mutated gene	Transcript	Sillence classification	Nucleotide change	Predicted amino acid change	Mutation type	Mutation effect	Exon or intron	ExAC frequency	PhyloP primates	PhyloP vertebrates	PhyloP placental mammals	SIFT score	Polyphen2 HumVar score
37	WNT1	NM_005430.3	IV	c.385G>A	p.A129T	Substitution	Missense	Exon 3	0	0.655	5.635	2.63	0	0.922
				c.506G>A	p.G169D	Substitution	Missense	Exon 3	0	0.655	5.585	2.574	0	0.999
38	WNT1	NM_005430.3	IV	c.506G>A	p.G169D	Substitution	Missense	Exon 3	0	0.655	5.585	2.574	0	0.999
				c.506dupG	p.C170Lfs*6	Deletion	Frameshift	Exon 3	0	–	–	–	–	–

- Data is not available from the current database or software

in Table 2). The other 52 mutations have been reported in previous studies (4 frameshift, 1 nonsense, 23 missense, 6 splice site, 1 new start codon, and 1 whole gene deletion; shown in Supplemental materials). Nine mutations were found in more than two unrelated families.

Molecular diagnosis

The genetic etiologies found in this cohort were as follows: *COL1A1* (n = 37), *COL1A2* (n = 29), *TMEM38B* (n = 3), *FKBP10* (n = 3), *PLOD2* (n = 1), *IFITM5* (n = 9), *SERPINF1* (n = 4), and *WNT1* (n = 4). The details of the different mutation types are shown in Fig. 1.

The majority of the variants were due to *COL1A1* and *COL1A2* mutations, of which 37 patients had *COL1A1* variants, followed by 29 patients with *COL1A2* mutations. Fifteen patients had *COL1A1* missense mutations, 9 patients had *COL1A1* frameshift mutations, 9 patients had splice site mutations, 3 patients had *COL1A1* nonsense mutations, and 1 patient harbored whole gene deletions (family 39). A total of 25 patients had *COL1A2* missense mutations, with the glycine to serine being the most frequent amino acid substitutions. Three patients had *COL1A2* splice site mutations, and one patient had a deletion in chr5:146294373 and chr7:94038710 that resulted in chromosome translocation t(5;8)(q32;q21).

As the third most common cause of OI, *IFITM5* mutations have been recognized in nine patients. The recurrent *IFITM5* c.-14C>T mutation was found in eight patients, and *IFITM5* c.119C>T mutation was found in one patient.

For AR inherited OI, four individuals carried *SERPINF1* mutations, three of which were compound heterozygous and one was homozygous mutation. The majority of *SERPINF1* variants lead to the production of truncated protein, and the missense mutations were predicted to affect the secondary structure of the protein encoded by *SERPINF1*.

Mutations in *TMEM38B* were identified in three patients from two unrelated families. In two affected siblings from a non-consanguineous family, we identified a nonsense mutation that leads to premature termination codon. The c.455-7T>G in *TMEM38B* was found in another family which led to an insertion of two amino acids p. G152_A153insVL and affected a key domain of the TRIC-B protein.

Genes that are involved in collagen folding and cross-liking (*FKBP10*, *PLOD2*) had been identified to cause OI in four patients. Most of the mutations in *FKBP10* are predicted to result in loss of function. One patient was identified to be compound heterozygous for splice mutation and missense mutation in *PLOD2*.

Patients from four families were either reported homozygous (n = 1, c.506dupG) or compound heterozygous (n = 3, [c.110 T>C] + [c.505G>T] and [c.385G>A] + [c.506 G>A], [c.506 G>A] + [c.506dupG]) for alterations of the *WNT1*

Fig. 1 Mutation spectrum in Chinese OI patients. **a** Mutation spectrum in Chinese OI patients by causative genes. **b** Mutation spectrum of type I OI patients. **c** Mutation spectrum of type III OI patients. **d** Mutation spectrum of type IV OI patients. **e** Mutation spectrum of Chinese osteogenesis imperfecta patients by mutation effects
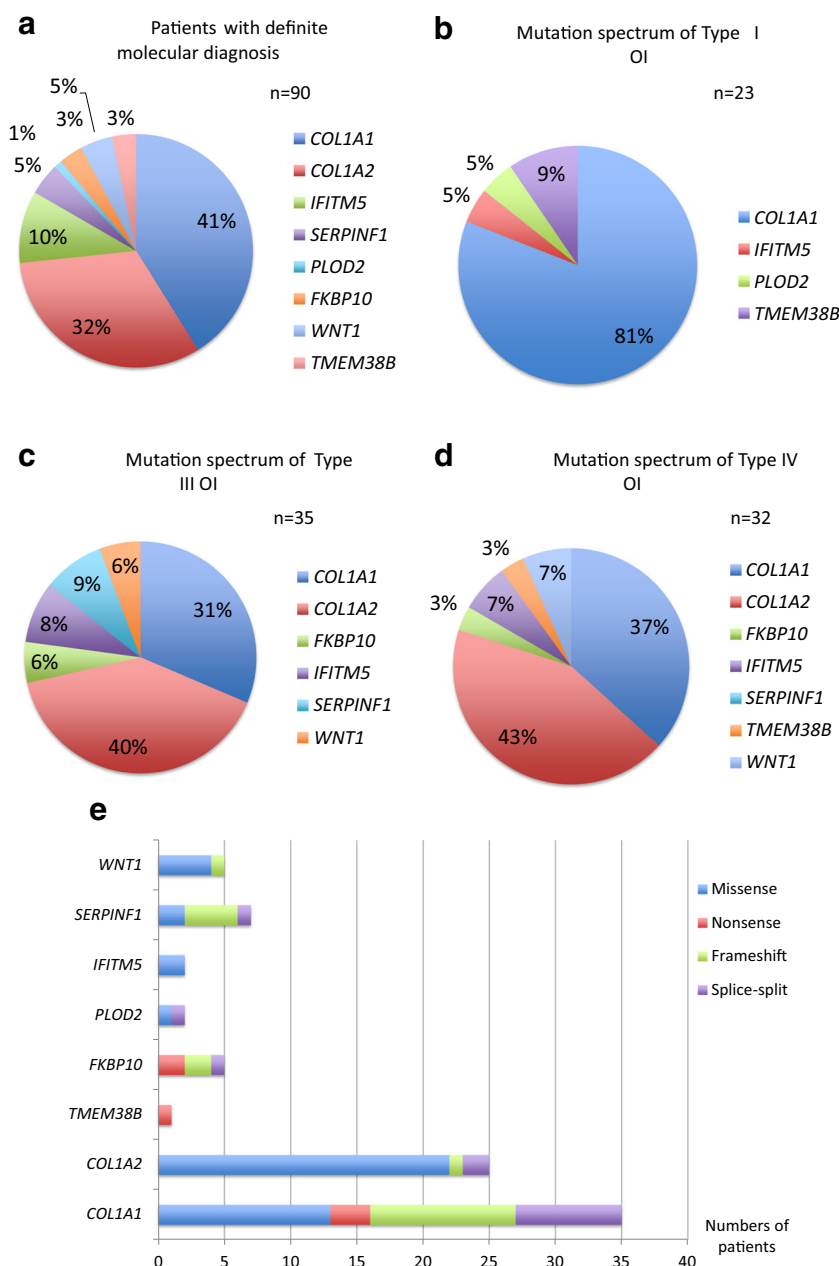
gene. All of the novel missense mutations were predicted to affect the N-terminal domain of the Wnt1 protein.

Genotype-phenotype correlation

Mutations leading to an early stop codon or frameshift in *COL1A1* were regarded as the quantitative group (haploinsufficiency). Mutations causing amino acid substitutions in the triple helical domain of *COL1A1* or *COL1A2* were classified into the collagen qualitative defect group. As the effect of splice site mutation was difficult to predict, we did not include

splice site mutations in the genotype-phenotype correlation of type I collagen.

BMD

AR inherited patients were prone to have lower lumbar spine BMD (Z-score − 3.6 ± 2.1) compared with AD inherited patients (Z-score −2.6 ± 1.9, $P = 0.05$). Compared with the quantitative change in collagen, patients with triple helical mutations had significantly lower femoral BMD (−3.1 ± 2.1 for quantitative and −5.0 ± 3.2 for qualitative, $P = 0.034$, Table 3).

Table 3   Genotype-phenotype relationship in Chinese OI patients

	Number	Fracture frequency (number/year)	LS BMD Z-score (mean ± SD)	FN BMD Z-score (mean ± SD)	Ht Z-score (mean ± SD)	Blue sclera *n* positive (percentage)	DI *n* positive (percentage)
AD	76	1.48 ± 1.41	−2.6 ± 1.9	−4.3 ± 3.2	−4.5 ± 6.9	54 (71.1%)	17 (22.4%)
AR	15	2.81 ± 3.49	−3.6 ± 2.1	−4.5 ± 2.0	−3.1 ± 3.7	4 (26.7%)	1 (6.7%)
P value		0.166	**0.050**	0.839	0.465	**0.001**	0.138
Type I collagen	67	1.51 ± 1.44	−2.6 ± 1.9	−4.3 ± 2.9	−4.7 ± 7.2	54 (80.6%)	17 (25.4%)
IFITM5	9	1.14 ± 1.06	−2.3 ± 1.5	−5.1 ± 5.6	−3.1 ± 3.9	0 (0%)	0 (0%)
AR	15	2.81 ± 3.49	−3.6 ± 2.1	−4.5 ± 2.0	−3.2 ± 3.7	4 (26.7%)	1 (6.7%)
P value		0.217	0.626	0.136	0.754	**<0.0001**	0.083
Quantity	13	1.28 ± 0.91	−1.9 ± 1.7	−3.1 ± 2.1	−1.2 ± 4.8	13 (100%)	0 (0%)
Quality	40	1.57 ± 1.69	−2.9 ± 2.0	−5.0 ± 3.2	−6.5 ± 8.2	30 (75%)	11 (27.5%)
P value		0.540	0.079	**0.034**	**0.022**	0.096	**0.048**
COL1A1 quality	15	1.55 ± 1.27	−2.7 ± 2.5	−4.1 ± 2.3	−5.6 ± 6.7	13 (86.7%)	3 (20%)
COL1A2 quality	25	1.57 ± 1.93	−3.0 ± 1.6	−5.6 ± 3.6	−6.9 ± 9.1	17 (68%)	8 (32%)
P value		0.972	0.654	0.185	0.620	0.269	0.403

LS lumbar spine, *FN* femoral neck, *Ht* height, *AD* autosomal dominant, *AR* autosomal recessive Bold values indicate significance.

Height

Patients with qualitative mutations were significantly shorter in height than patients with quantitative defects (−1.2 ± 4.8 for quantitative and −6.5 ± 8.2 for qualitative, *P* = 0.022). There was no difference between quantitative and any qualitative mutation in the α1- versus α2-chain. Nor was there any difference in the inheritance pattern (Table 3).

Blue sclera

Patients with AD OI had blue sclera significantly more frequently than patients with AR OI (71.1% in AD versus 26.7% in AR, *P* = 0.001). All patients with haploinsufficiency of type I collagen had blue sclera, compared with 75% of qualitative collagen defect patients. The position of glycine within a1(1) triple helix domain was associated with the presence of blue sclera (Table 3). There is a relationship between the triple helical position of glycine mutations in COL1A1 and COL1A2 and the manifestation of blue sclera and dentinogenesis imperfecta, from the N- to C-terminal. All patients with mutations affecting the N-terminal of Gly154 (p.G332) had blue sclera (Fig. 2).

Dentinogenesis imperfecta

DI was more prevalent in patients with collagen qualitative defects than quantitative defects (0 versus 27.5%, *P* = 0.023). Concerning N- to C-terminal location, none of the individuals with helical glycine substitutions in the N-terminal of p.Gly332 in the α1-chain had DI. The presence of DI was not associated with OI-type inheritance pattern or types of defects in type I collagen. No significant difference was found

in patients who had glycine mutations in the triple helical domain between *COL1A1* and *COL1A2* (Table 3).

Discussion

To the best of our knowledge, our study was the largest OI cohort in Asian patients. The results of this study outlined a significant mutation spectrum in Chinese patients. By using a self-designed panel with 722 candidate genes, we identified 81 variants in 90 probands which included 45 novel variants.

The accurate and reliable molecular diagnosis of OI is important, because it enables appropriate therapeutic interventions and facilitates a more precise description of prognosis. Our panel sequenced 722 genes spanning 9712 exons simultaneously. The depth of sequencing coverage was high for over 200× of the amplicons, indicating that the majority of amplicons were covered sufficiently. In total, the panel was able to identify 90 probands with gene mutations in the patients analyzed (detection rate 87.4%). While further validation in a larger population is needed, our panel in itself proves the versatility of NGS for variant detection in OI patients.

Our medical center is a tertiary referral center in China, and therefore, our study includes more patients with moderate and severe types of OI, especially in adult patients. Compared with children and adolescent OI patients, adult patients were found to be shorter than people of the same age. This was similar to what was reported by Lindahl et al., and may be related to the existence of multiple fractures, leading to severe skeletal deformities and a shorter statue [22]. All of the clinical information was collected before the initiation of treatment with bisphosphonates. Compared with children and adolescents,

Fig. 2 Relationship between the triple helical position of glycine mutations in COL1A1 and COL1A2 and the manifestation of blue sclera and dentinogenesis imperfecta, from the N- to C-terminal. The *colored boxes* suggest that all patients with mutations affecting N-terminal of Gly154(p.G332) had blue sclera with an absence of dentinogenesis imperfecta
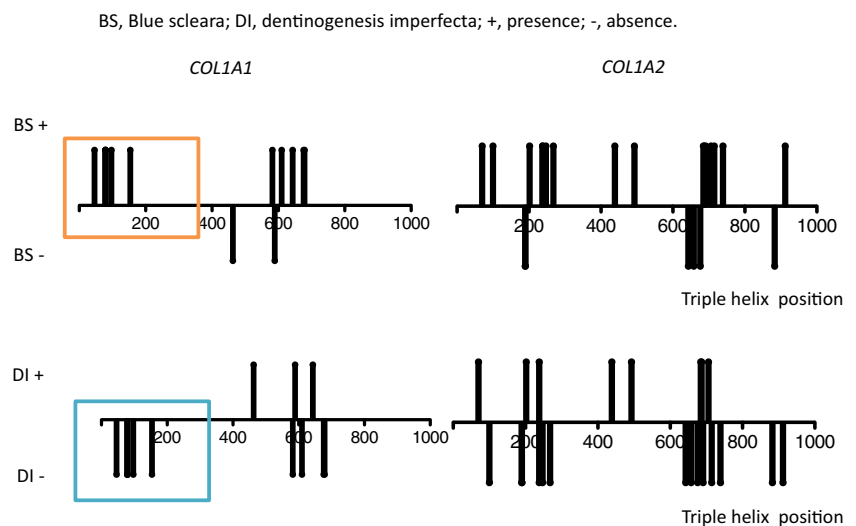
the average interval of diagnosis for adults was significantly longer, and therefore likely to result in more severe phenotypes to be present in adults and suggesting that early diagnosis is of great importance in these patients [22]. Moreover, type III and type IV OI are more prevalent in adult patients. To limit this bias, we further conducted analyses on different types of OI.

Similar to previous studies, *COL1A1* and *COL1A2* mutations were dominant in our cohort [23]. However, the percent of *COL1A1* and *COL1A2* mutations was lower than expected, accounting for only 73.3% of all variations. We further analyzed the gene mutations that were observed in different types of OI and found that *COL1A1* and *COL1A2* explained 81% of type I OI and 71% of moderate to severe OI. The detection rate in moderate and severe OI is similar to previous studies, whereas the detection rate in type I OI is lower. The reason for this may stem from the definition of type I OI, as Bardai et al. used the definition of typical type OI, which was defined as having both extra-skeletal characteristics and bone fragility [24]. In contrast, we used only bone fragility to differentiate OI types and included patients with *TMEM38B*, *PLOD2*, and *IFITM5* in our type I OI analysis. This is similar to a recent study, which found that 12 out of 28 individuals with mild bone fragility and an absence of extraskeletal manifestation, who were identified to have gene mutations, carried mutations in *COL1A1* and *COL1A2* [25].

As for genotype-phenotype analysis, we confirmed that patients with collagen qualitative defects had lower BMDs than patients with collagen quantitative defects, especially at the femoral neck. As previous studies mainly focused on the axial skeleton and BMD is known to vary by location in OI patients, our results provide valuable information about the appendicular skeleton of OI patients [26–28]. Consistent with previous studies, we found that patients with triple helical mutations were taller than patients with haploinsufficiency of type I collagen. Similarly, previous studies have reported that the triple helical

position of glycine in *COL1A1* was associated with the presence of blue sclera [22, 28], which we also observed, as all patients with mutations affecting the first 152 amino acids in *COL1A1* presented blue sclera with the absence of DI.

As the third most common mutation in our cohort, mutations in *IFITM5* were detected in nine patients, eight with c.-14C>T and one with c.119C>T. The patient with c.-14C>T was quite similar to what was reported previously in large phenotype variability, typical calcification, and hypercallus formation. Similar to previous studies, patients with *IFITM5* c.119C>T had severe skeletal deformities [29]. However, we observed no significant difference in BMD, height, sclera hue, and dentinogenesis imperfecta between type V and other types of OI. The heterogeneity of *IFITM5*, which has been shown to be related to OI, may explain the similarity in these clinical phenotypes.

To date, there have been limited studies focusing on AR OI in Chinese patients. We successfully detect 15 patients with AR OI. As such, here, we further extend the known mutation spectrum to include *WNT1*, *SERPINF1*, *PLOD2*, *FKBP10*, and *TMEM38B*, thus contributing to a better understanding of the genotype that is present in Chinese patients. In our cohort, *SERPIFN1* and *WNT1* mutations were the most common causes of AR OI. This is different from what was reported by Bardai et al., who found *SERPINF1* and *CRTAP* to be more prevalent in Canadian OI patients [24]. As our study focused on a Chinese population, we attributed this discrepancy to a difference in ethnicity. In addition, we further detected three patients with *FKBP10* and one patient with *PLOD2* mutation. In contrast to previous studies, which reported that *FKBP10* and *PLOD2* were responsible for Bruck syndrome, we found that joint contracture was absent in patients with a *FKBP10* mutation but was present in a patients with a *PLOD2* mutation [30].

Compared with AD OI patients, AR OI tends to have more severe skeletal deformities and were less likely to have blue sclera. This finding is similar to what was reported previously.

The reason for this phenomenon has yet to be defined; however, we hypothesize that this may be due to the fact that AD OI is usually associated with collagen defects, whereas AR OI tends to affect the post-translational process of collagen or bone mineralization.

Despite successful detection of most mutations in our patients, we failed to identify any related gene mutation in 12 patients. This may have occurred because probands carry mutations in genes not targeted by this study given that we chose to examine the most highly implicated genes in skeletal dysplasia causation. For instance, *PLS3*, a gene that encodes a protein in the formation F-actin and associated with juvenile osteoporosis, was not included at the time we designed our panel [31]. One of the test-negative patients was later found to have a large deletion involving intron 9 to the 3′-UTR (E10-E16 del) in *PLS3*. For those patients without a detected mutation, further whole exome sequencing may be able to resolve some of these possibilities [32]. Other possible reasons for the detection of false negatives may be poor coverage due to high GC content and the difficulties associated with the detection of large deletions [33]. Alternative methods like comparative genome hybridization or multiplex ligation-dependent probe amplification would be needed to investigate further [34].

Conclusions

The achievement of more accurate diagnosis would greatly benefit the management of patients with congenital skeletal dysplasia. In this study, we present the largest OI sample in China as screened by NGS. In particular, we successfully identified 81 variants, which included 45 novel variants. The novel variants that were identified in these Chinese OI patients enrich the known pathogenic spectrum of OI. Finally, our analysis of genotype-phenotype correlation helps make a differential diagnosis and helps predict the prognosis of OI.

References

1. Forlino A, Marini JC (2016) Osteogenesis imperfecta. Lancet 387:1657–1671
2. Sillence DO, Senn A, Danks DM (1979) Genetic heterogeneity in osteogenesis imperfecta. J Med Genet 16:101–116
3. Van Dijk FS, Sillence DO (2014) Osteogenesis imperfecta: clinical diagnosis, nomenclature and severity assessment. Am J Med Genet A 164A:1470–1481
4. Semler O, Netzer C, Hoyer-Kuhn H, Becker J, Eysel P, Schoenau E (2012) First use of the RANKL antibody denosumab in osteogenesis imperfecta type VI. J Musculoskelet Neuronal Interact 12:183–188
5. Semler O, Garbes L, Keupp K et al (2012) 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 91:349–357
6. Homan EP, Rauch F, Grafe I et al (2011) Mutations in SERPINF1 cause osteogenesis imperfecta type VI. J Bone Miner Res Off J Am Soc Bone Miner Res 26:2798–2803
7. Baldridge D, Schwarze U, Morello R et al (2008) CRTAP and LEPRE1 mutations in recessive osteogenesis imperfecta. Hum Mutat 29:1435–1442
8. Barnes AM, Carter EM, Cabral WA et al (2010) Lack of cyclophilin B in osteogenesis imperfecta with normal collagen folding. N Engl J Med 362:521–528
9. Christiansen HE, Schwarze U, Pyott SM, AlSwaid A, Al Balwi M, Alrasheed S, Pepin MG, Weis MA, Eyre DR, Byers PH (2010) Homozygosity for a missense mutation in SERPINH1, which encodes the collagen chaperone protein HSP47, results in severe recessive osteogenesis imperfecta. Am J Hum Genet 86:389–398
10. Kelley BP, Malfait F, Bonafe L et al (2011) Mutations in FKBP10 cause recessive osteogenesis imperfecta and Bruck syndrome. J Bone Miner Res Off J Am Soc Bone Miner Res 26:666–672
11. Laine CM, Joeng KS, Campeau PM et al (2013) WNT1 mutations in early-onset osteoporosis and osteogenesis imperfecta. N Engl J Med 368:1809–1816
12. Lapunzina P, Aglan M, Temtamy S et al (2010) Identification of a frameshift mutation in Osterix in a patient with recessive osteogenesis imperfecta. Am J Hum Genet 87:110–114
13. Symoens S, Malfait F, D'Hondt S, Callewaert B, Dheedene A, Steyaert W, Bachinger HP, De Paepe A, Kayserili H, Coucke PJ (2013) Deficiency for the ER-stress transducer OASIS causes severe recessive osteogenesis imperfecta in humans. Orphanet J Rare Dis 8:154
14. Lindert U, Cabral WA, Ausavarat S et al (2016) MBTPS2 mutations cause defective regulated intramembrane proteolysis in X-linked osteogenesis imperfecta. Nat Commun 7:11920
15. Donahue LR, Chang B, Mohan S et al (2003) A missense mutation in the mouse Col2a1 gene causes spondyloepiphyseal dysplasia congenita, hearing loss, and retinoschisis. J Bone Miner Res Off J Am Soc Bone Miner Res 18:1612–1621
16. Ku CS, Cooper DN, Polychronakos C, Naidoo N, Wu M, Soong R (2012) Exome sequencing: dual role as a discovery and diagnostic tool. Ann Neurol 71:5–14
17. Glorieux FH, Rauch F, Plotkin H, Ward L, Travers R, Roughley P, Lalic L, Glorieux DF, Fassier F, Bishop NJ (2000) Type V osteogenesis imperfecta: a new form of brittle bone disease. J Bone Miner Res Off J Am Soc Bone Miner Res 15:1650–1658
18. Ruppe MD, Brosnan PG, Au KS, Tran PX, Dominguez BW, Northrup H (2011) Mutational analysis of PHEX, FGF23 and DMP1 in a cohort of patients with hypophosphatemic rickets. Clin Endocrinol 74:312–318
19. Li H, Ji CY, Zong XN, Zhang YQ (2009) Height and weight standardized growth charts for Chinese children and adolescents aged 0 to 18 years. Chin J Pediatr 47:487–492
20. Xu H, Zhao Z, Wang H, Ding M, Zhou A, Wang X, Zhang P, Duggan C, Hu FB (2013) Bone mineral density of the spine in 11,898 chinese infants and young children: a cross-sectional study. PLoS One 8:e82098
21. Khadilkar AV, Sanwalka NJ, Chiplonkar SA, Khadilkar VV, Mughal MZ (2011) Normative data and percentile curves for dual

energy X-ray absorptiometry in healthy Indian girls and boys aged 5-17 years. Bone 48:810–819

22. Lindahl K, Astrom E, Rubin CJ, Grigelioniene G, Malmgren B, Ljunggren O, Kindmark A (2015) Genetic epidemiology, prevalence, and genotype-phenotype correlations in the Swedish population with osteogenesis imperfecta. Eur J Human Genet: EJHG 23: 1112

23. Marini JC, Forlino A, Cabral WA et al (2007) 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 28:209–221

24. Bardai G, Moffatt P, Glorieux FH, Rauch F (2016) DNA sequence analysis in 598 individuals with a clinical diagnosis of osteogenesis imperfecta: diagnostic yield and mutation spectrum. Osteoporos Int 27:3607–3613

25. Bardai G, Ward LM, Trejo P, Moffatt P, Glorieux FH, Rauch F (2017) Molecular diagnosis in children with fractures but no extraskeletal signs of osteogenesis imperfecta. **Osteoporos Int**

26. Rauch F, Lalic L, Roughley P, Glorieux FH (2010) Genotype-phenotype correlations in nonlethal osteogenesis imperfecta caused by mutations in the helical domain of collagen type I. Eur J Human Genet: EJHG 18:642–647

27. Rauch F, Land C, Cornibert S, Schoenau E, Glorieux FH (2005) High and low density in the same bone: a study on children and adolescents with mild osteogenesis imperfecta. Bone 37:634–641

28. Rauch F, Lalic L, Roughley P, Glorieux FH (2010) Relationship between genotype and skeletal phenotype in children and adolescents with osteogenesis imperfecta. J Bone Miner Res Off J Am Soc Bone Miner Res 25:1367–1374

29. Hoyer-Kuhn H, Semler O, Garbes L, Zimmermann K, Becker J, Wollnik B, Schoenau E, Netzer C (2014) A nonclassical IFITM5 mutation located in the coding region causes severe osteogenesis imperfecta with prenatal onset. J Bone Miner Res Off J Am Soc Bone Miner Res 29:1387–1391

30. Zhou P, Liu Y, Lv F, Nie M, Jiang Y, Wang O, Xia W, Xing X, Li M (2014) Novel mutations in FKBP10 and PLOD2 cause rare Bruck syndrome in Chinese patients. PLoS One 9:e107594

31. van Dijk FS, Zillikens MC, Micha D et al (2013) PLS3 mutations in X-linked osteoporosis with fractures. N Engl J Med 369:1529–1536

32. Rauch F, Lalic L, Glorieux FH, Moffatt P, Roughley P (2014) Targeted sequencing of a pediatric metabolic bone gene panel using a desktop semiconductor next-generation sequencer. Calcif Tissue Int 95:323–331

33. Benjamini Y, Speed TP (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res 40: e72

34. van Dijk FS, Byers PH, Dalgleish R, Malfait F, Maugeri A, Rohrbach M, Symoens S, Sistermans EA, Pals G (2012) EMQN best practice guidelines for the laboratory diagnosis of osteogenesis imperfecta. Eur J Human Genet: EJHG 20:11–19