Citation: BoneKEy Reports **2**, Article number: 456 (2013) | doi:10.1038/bonekey.2013.190
© 2013 Macmillan Publishers Limited All rights reserved 1476-5640/13
www.nature.com/bonekey

ORIGINAL ARTICLE

Whole exome sequencing is an efficient, sensitive and specific method of mutation detection in osteogenesis imperfecta and Marfan syndrome

Aideen M McInerney-Leo[1], Mhairi S Marshall[1], Brooke Gardiner[1], Paul J Coucke[2], Lut Van Laer[3], Bart L Loeys[3,4], Kim M Summers[5], Sofie Symoens[2], Jennifer A West[6], Malcolm J West[6], B Paul Wordsworth[7], Andreas Zankl[8,9,10], Paul J Leo[1], Matthew A Brown[1,12] and Emma L Duncan[1,11,12]

[1]The University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, Brisbane, Queensland, Australia. [2]Medical Genetics, The University Hospital Ghent, Gent, Belgium. [3]University of Antwerp, Antwerp University Hospital, Antwerp, Belgium. [4]Department of Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. [5]The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, UK. [6]The University of Qld Northside Clinical School, Prince Charles Hospital, Chermside, Queensland, Australia. [7]NIHR Oxford Musculoskeletal Biomedical Research Unit, Nuffield Orthopaedic Centre, Oxford, UK. [8]The University of Queensland, UQ Centre for Clinical Research, Herston, Queensland, Australia. [9]Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia. [10]Academic Department of Medical Genetics, The Children's Hospital at Westmead, Sydney, New South Wales, Australia. [11]Department of Endocrinology, Royal Brisbane and Women's Hospital, Herston, Queensland, Australia.

Osteogenesis imperfecta (OI) and Marfan syndrome (MFS) are common Mendelian disorders. Both conditions are usually diagnosed clinically, as genetic testing is expensive due to the size and number of potentially causative genes and mutations. However, genetic testing may benefit patients, at-risk family members and individuals with borderline phenotypes, as well as improving genetic counseling and allowing critical differential diagnoses. We assessed whether whole exome sequencing (WES) is a sensitive method for mutation detection in OI and MFS. WES was performed on genomic DNA from 13 participants with OI and 10 participants with MFS who had known mutations, with exome capture followed by massive parallel sequencing of multiplexed samples. Single nucleotide polymorphisms (SNPs) and small indels were called using Genome Analysis Toolkit (GATK) and annotated with ANNOVAR. CREST, exomeCopy and exomeDepth were used for large deletion detection. Results were compared with the previous data. Specificity was calculated by screening WES data from a control population of 487 individuals for mutations in *COL1A1*, *COL1A2* and *FBN1*. The target capture of five exome capture platforms was compared. All 13 mutations in the OI cohort and 9/10 in the MFS cohort were detected (sensitivity = 95.6%) including non-synonymous SNPs, small indels (<10 bp), and a large UTR5/exon 1 deletion. One mutation was not detected by GATK due to strand bias. Specificity was 99.5%. Capture platforms and analysis programs differed considerably in their ability to detect mutations. Consumable costs for WES were low. WES is an efficient, sensitive, specific and cost-effective method for mutation detection in patients with OI and MFS. Careful selection of platform and analysis programs is necessary to maximize success.

BoneKEy Reports **2**, Article number: 456 (2013) | doi:10.1038/bonekey.2013.190

Introduction

Heritable disorders of connective tissue (CT) are among the most common Mendelian diseases. As most organs and tissues contain some CT, these are systemic diseases but will manifest predominantly in CT-rich structures such as the skin, ligaments, tendons, bone, cartilage and aorta. Anomalies of the ocular

Correspondence: Dr EL Duncan, Department of Endocrinology, Royal Brisbane and Women's Hospital, James Mayne Building, Butterfield Road, Herston, Queensland 4029, Australia.
E-mail: e.duncan@uq.edu.au
[12]These authors contributed equally to this work.

Received 7 August 2013; revised 16 October 2013; accepted 23 October 2013; published online 4 December 2013

and auditory system are also common. Both osteogenesis imperfecta (OI) and Marfan syndrome (MFS) are common Mendelian CT disorders. However, despite their prevalence, their diagnosis is not straightforward due to clinical variability and genetic heterogeneity.

Osteogenesis imperfecta. OI is a heterogeneous group of heritable disorders of bone fragility, usually inherited in an autosomal dominant (AD) fashion, with prevalence of 6–7 per 100 000.[1] In addition to fractures, clinical manifestations may include discolored sclerae, deafness, dentinogenesis imperfecta and short stature.

Mutations in *COL1A1* and *COL1A2* are the commonest cause for AD OI.[2] Both genes have unusual genomic structures: both consist of multiple exons (>50), almost all of 54 base pairs (bp), that code for a repeating amino-acid sequence Gly-X-Y. Glycine, the smallest amino acid, is necessary in this position to fit the tight helix of type 1 collagen.[3] Common OI-causing mutations of *COL1A1* and *COL1A2* include splice site mutations, causing exon skipping, abnormal splicing and/or frameshifts; and point mutations, particularly when a larger amino acid is substituted for glycine. Of note, *de novo* mutations are frequent (∼25–30%).[4]

Autosomal recessive (AR) forms of OI are much less frequent and are rarely due to mutations in *COL1A1* or *COL1A2 per se*. Rather, AR OI may arise from mutations in genes involved in the synthesis and processing of type 1 collagen. These include the 3′ prolyl hydroxylation complex and/or chaperone proteins (*CRTAP*[5] *LEPRE1*,[6] *PPIB*,[7] *SERPINH1*,[8] *FKBP10*,[9] and *SERPINF1*[10,11]). Mutations in *BMP1,* involved in procollagen processing, have also been described.[12] Mutations in *IFITM5* encoding the bone restricted IFITM-like protein, BRIL[13,14] cause OI type V; as yet there is no known functional interaction of BRIL with type 1 collagen.[15] Other OI genes not currently known to involve type 1 collagen include *SP7*,[16] *TMEM38B*[17] and *WNT1.*[18]

The diagnosis and classification of OI remains predominantly clinical.[19] In part this has been due to the expense of screening large genes with diverse mutation types (typically $A1850 for *COL1A1*/*COL1A2* alone), although for AR OI there is an increasing number of genes that need testing. Some authors have argued that the current system of clinical and radiological classification of OI should be replaced by a classification system based on the underlying genetic defect.[20] This would substantially improve the accuracy and usefulness of genetic counseling, clarifying the mode of inheritance and recurrence risk and—at least in part—allowing some prediction of disease severity according to mutation type. Potentially, treatment options could be targeted according to their appropriateness for the OI subtype.

Lastly, children with multiple fractures of varying ages may have OI as opposed to non-accidental injury (although the two are not mutually exclusive); an accurate and fast genetic diagnosis may be of great importance for the families of such children.

Marfan syndrome. MFS is an AD CT disorder with an estimated prevalence of 2–3/10 000.[21] The cardinal clinical features of MFS are aortic root dilatation and ectopia lentis; other systemic manifestations include disproportionate limb length, pectus carinatum or excavatum, scoliosis, joint laxity, myopia, spontaneous pneumothorax and mitral valve prolapse.[22,23] The most life-threatening cardiovascular manifestation, seen in almost all individuals with MFS, is dilatation of the aortic root and ascending aorta,[24] which can result in aortic dissection and sudden death. In an old review, the average age of death was 40 years in men and 50 years in women; aortic aneurysm accounted for 80% of deaths.[25] Untreated, dilatation presents clinically in the third or fourth decade although it can present earlier, particularly in professional athletes.[26] Diagnosis of MFS is thus critical, and allows surgical and/or medical intervention to prevent the high mortality of this condition.[27] However, even with the revised Ghent nosology,[23] definitive clinical diagnosis may be difficult, delayed and/or ambiguous in some individuals, including younger persons (aged <20 years), individuals with Marfanoid syndromes (showing some features of MFS without meeting the diagnostic criteria) and individuals from families with significant phenotypic variability. Excluding the diagnosis in at-risk individuals can maximize their quality-of-life and minimize unnecessary drugs, medical intervention and restriction of life activities (including exercise and child-bearing).

Mutations in fibrillin 1 (encoded by *FBN1*) cause MFS,[28] and 90–95% of individuals fulfilling diagnostic criteria for MFS carry mutations in this gene.[29–31] Three quarters of disease-causing mutations are non-synonymous (missense and nonsense) and the majority of the remainder are splice site and small (<10 bp) deletions or insertions (indels).[32] Large whole-exon deletions may account for approximately 3–4%.[30,33] Approximately 25% appear to be *de novo* mutations,[34] a small subset of these may result from germline mosaicism.[35] Increasing phenotypic specificity improves sensitivity for *FBN1* mutation detection.[29–31]

Individuals with ascending aortic dilatation who neither meet diagnostic criteria for MFS[23] nor carry mutations in *FBN1* have also been identified. Most commonly these subjects have Loeys–Dietz syndrome, with mutations in *TGFBR1* or *TGFBR2*.[36] Of note, mutations in these genes have been reported in some individuals diagnosed with MFS syndrome.[37] Mutations in other genes (for example, *ACTA2*[38] and *SMAD3*[39]) have been found in rare families with aortic dilatation and some Marfanoid features.

Targeted massive parallel sequencing (MPS) of *FBN1, TGFBR1* and *TGFBR2* detected a mutation in *FBN1* in 75 of 87 individuals meeting the revised Ghent nosology for MFS.[23] Subsequent multiplex ligation-dependent probe amplification and Sanger sequencing detected an additional four cases, which included three with large deletions (>1 exon) and one with a duplication.[30] Conversely, 10–15% of families tested because of a Marfanoid-like phenotype and found to carry *FBN1* mutations do not fulfill diagnostic criteria for MFS.[40,41]

The importance of a genetic diagnosis in MFS has recently been emphasized.[23] However, genetic screening in individuals with a Marfanoid phenotype is challenging, time-consuming and expensive (∼$A1500) due to the size of *FBN1* (65 exons, >257 kb) and wide distribution and diversity of mutations. Thus, many individuals diagnosed with MFS clinically do not undergo confirmatory genetic testing.

Massive parallel sequencing. MPS is a recent major advance in genetic technology allowing sequencing of multiple target regions simultaneously. Although the exome comprises <1%

of the whole genome, 85% of disease-causing mutations are estimated to be exonic;[42] therefore, whole exome sequencing (WES) has the potential to be a cost-effective, rapid and sensitive technology for detecting mutations in Mendelian disorders.

In particular, WES may represent an ideal means of genetic testing for MFS and OI. Both are common disorders, usually diagnosed clinically without gene sequencing because the causative genes are large with multiple exons and diverse mutation types; further, in AR OI in particular there are increasingly large numbers of genes requiring screening. Efficient genetic testing would greatly benefit at-risk family members and individuals with borderline phenotypic features, enhance genetic counseling and allow critical differential diagnoses. In this paper, we present data demonstrating that WES is an efficient, sensitive and cost-effective method of detecting mutations for OI and MFS, and compare the available platforms for their target capture of relevant genes for these conditions.

Results

Sequencing efficiency. Mapping and coverage statistics and detailed base calling statistics are presented in Supplementary Table 1 (for OI) and 2 (for MFS).

OI. Coding variants were detected in *COL1A1* or *COL1A2* in all patients (presented in **Table 1**). Of note, one mutation (a single-base deletion) was detected by Genome Analysis Toolkit (GATK) (v2.2–3) when alignment was performed by BWA but not with Novoalign v2.07.18; this mutation was subsequently detected with Novoalign v2.08.02 but not the most recent release (Novoalign v3.00.02). One sample had two apparently rare mutations, one in *COL1A1* and one in *COL1A2* but one of these was present in multiple individuals from the same run (who did not have OI) and was therefore filtered out as a platform or alignment artefact.

MFS. The mutation detection results for MFS samples are presented in **Table 2**. A coding variant was found in 8 of the 10 Marfan patients. The two remaining samples (MFS-2 and MFS-5) were screened for large deletions. Neither CREST nor ExomeCopy detected a deletion in either. After discussion with the previous investigator (BL) about the type of mutation MFS-5 carried, a further deletion detection program was used (ExomeDepth), which detected a deletion of the 5′-UTR and exon 1.

We failed to identify a mutation in the remaining MFS sample (MFS-2). After reconciliation, we learned this was a missense mutation (*FBN1*:c.2723T>G:p.908C>F). The sequencing BAM files (alignment data in binary format) were reviewed using the Integrative Genomics Viewer (IGV).[43] The variant was present but had insufficient proportional coverage for calling by GATK v2.2–3 (18 reads for the reference base and 2 reads for the alternate). The latest version of GATK v2.4–9 also failed to identify this variant.

Accuracy. At reconciliation all identified mutations for both the OI and MFS cohorts matched those reported previously.

Specificity. Six variants in *COL1A1* or *COL1A2* were seen in the high bone mass control population. As all of the variants in the patient cohorts were predicted by SIFT to be either deleterious or of no known consequence (due to frameshift or stop mutations), we filtered the control group similarly. Only one variant in *COL1A2* remained. This variant has not been previously reported in the OI database LOVD but has been reported in another control population, does not involve a glycine residue, and is located in the first exon where mutations have rarely been demonstrated in patients with OI. Repeating this analysis for *FBN1* identified eight variants in the high bone mass control population; of these, one was predicted to be

Table 1 Summary of identified variants for OI samples. Exons for *COL1A1* and *COL1A2* numbered according to transcripts NM_000088 and NM_000089 respectively

	No. of variants identified after GATK VQSR	*No. of good quality variants in* COL1A1 *and* COL1A2	*No. of remaining variants passing platform and sample quality filters*	*No. of remaining non-synonymous variants in coding or splice site regions*	*No. of remaining variants with minor allele frequency <0.001*	*Identified mutation*	*Concurs with previous reported mutation*
OI-1	21 763	7	3	1	1	*COL1A2* Exon 25 c.1459 G>A p.(G487R)	Yes
OI-2	12 283	8	2	1	1	*COL1A1* Exon 11 c.697-1 G>C	Yes
OI-3	31 234	13	2	1	1	*COL1A2* Exon 19 c.964 G>A p.(G322S)	Yes
OI-4	44 895	18	3	2	1	*COL1A2* Exon 19 c.1009 G>A p.(G337S)	Yes
OI-5	44 817	19	4	2	1	*COL1A1* Exon 42 c.3079delG p.(D1027fs)	Yes
OI-6	43 794	20	4	2	1	*COL1A1* Exon 9 c.658C>T p.(R220*)	Yes
OI-7	44 657	17	3	2	1	*COL1A1* Exon 44 c.3258delC p.(P1086fs)	Yes
OI-8	44 457	19	4	2	1	*COL1A1* Exon 11 c.769 G>A p.(G257R)	Yes
OI-9	44 849	16	4	2	1	*COL1A1* Exon 48 c.3638delG p.(G1213fs)	Yes
OI-10	44 641	19	2	1	1[a]	*COL1A1* Exon 49 c.3765delCA p.(I1256fs)[a]	Yes
OI-11	44 745	20	3	2	1	*COL1A1* Exon 26 c.1792C>T p.(R598*)	Yes
OI-12	44 966	18	4	2	1	*COL1A1* Exon 17 c.1128 delT p.(P377fs)	Yes
OI-13	45 639	16	3	2	1	*COL1A1* Exon 47 c.3495 delT p.(P1165fs)	Yes

Abbreviations: GATK, Genome Analysis Toolkit; OI, osteogenesis imperfecta; VQSR, variant quality score recalibration. Good quality variants: variants passing GATK VQSR.
[a]Missed by Novoalign v 2.07.18 and v 3.00.02 but detected by Novoalign v 2.08.02 and by Burrows–Wheeler Aligner BWA-SW (v-0.7.3).

Table 2 Summary of identified variants for MFS samples

	No. of variants identified after GATK VQSR	*No. of good quality variants in* FBN1	*No. of remaining variants passing platform and sample quality filters*	*No. of remaining non-synonymous variants in coding or splice site regions*	*No. of remaining variants with minor allele frequency <0.001.*	*Identified mutation*	*Concurs with previous reported mutation*
MFS-1	41 330	6	1	1	1	*FBN1* Exon 40 c.4864T>C p.(C1622R)	Yes
MFS-2	44 825	4	0	0	0	[a]Not detected	N/A
MFS-3	44 027	6	2	1	1	*FBN1* Exon 16 c.1904A>G p.(Y635C)	Yes
MFS-4	44 568	8	3	1	1	*FBN1* Exon 35 c.4259G>T p.(C1420F)	Yes
MFS-5	45 060	10	1	0	0	*FBN1* Deletion 5′-UTR and Exon 1	Yes
MFS-6	44 262	13	4	1	1	*FBN1* Exon 9 c.960_961 insA p.(Y320_ T321 delins*)	Yes
MFS-7	47 686	13	8	4	2[b]	*FBN1* Exon 1 c.1481 G>A p.(C494Y)	Yes
MFS-8	41 800	5	1	1	1	*FBN1* Exon 26 c.3131G>A p.(C1044Y)	Yes
MFS-9	44 122	8	1	1	1	*FBN1* Exon 14 c.1622G>T p.(C541F)	Yes
MFS-10	45 329	7	3	1	1	*FBN1* Exon 40 c.4930C>T p.(R1644*)	Not available[c]

Abbreviations: GATK, Genome Analysis Toolkit; MFS, Marfan syndrome; VQSR, variant quality score recalibration. Good quality variants: variants passing GATK VQSR. Exons are numbered according to Ensembl (*FBN1* ENST00000316623).
[a]Not called by GATK v2.2–3 or by GATK v 2.4–9 as insufficient calls for the alternate allele. [b]Only one variant predicted to be deleterious. [c]This variant has been reported four times in UMD associated with patients with Marfan syndrome. It has also been reported in the HapMap HCB (Asian) population ($n = 43$ individuals) and the HapMap JPT (Asian) population ($n = 85$ individuals) (rs140630). Exons are numbered according to transcript NM_000138.

deleterious by SIFT. This mutation has not been previously reported in the UMD *FBN1* database nor did it involve a change to or from cysteine. Thus, out of 487 individuals with high bone mass there were two variants of uncertain significance in *COL1A1, COL1A2* or *FBN1* resulting in a specificity of 99.5%.

Capture efficiency. There were marked differences between the theoretical target capture of known OI and MFS genes comparing the currently available exome capture platforms. Results are presented in **Table 3**. Exact target capture percentages for each exon and its genomic location are presented in Supplementary Table 3.

For OI, we first considered theoretical target capture for the two primary OI genes *COL1A1* and *COL1A2*. Illumina-Rapid targeted 100% of all exons of both genes. Illumina-TruSeq and Illumina-Rapid Exp targeted 96% and 98% of the coding region of *COL1A1* and *COL1A2*, respectively, with no target capture of exons 9 and 47 of *COL1A1* and exon 35 of *COL1A2*. Agilent and NimbleGen platforms targeted less overall, with target capture <90% for many exons.

When considering target capture for all remaining OI genes, Illumina-Rapid theoretically targeted all exons of known genes. However, it does not target the 5′-UTR region of *IFITM5* where the causative mutation for OI type V is located. NimbleGen targeted part of every exon of every gene but targeted capture was <90% for many exons. Similarly, the Agilent platform also targeted <90% for many exons. The remaining two Illumina platforms (Illumina-Rapid Exp and Illumina-TruSeq) failed to target seven exons of the remaining OI genes overall with an additional five exons having target capture <90%.

When considering the genes associated with Marfanoid phenotypes, all three Illumina platforms theoretically targeted 100% of every exon of *FBN1*. Agilent and NimbleGen targeted some portion of every exon of *FBN1* but this was <90% for six and eleven exons, respectively. For *TGFBR1* and *TGFBR2,* Illumina-Rapid failed to target exon 1 of *TGFBR1*. The two remaining Illumina platforms also failed to target this exon but also missed exon 2 of *TGFBR2*. Both Agilent and NimbleGen platforms captured 95% of the coding region of these genes; however, both targeted <90% in exon 1 of *TGFBR1* and exon 4 of *TGFBR2*.

In-house sequencing data from unrelated individuals was used to assess the ‘real-world’ experience of Illumina-TruSeq and NimbleGen capture technologies. The results are presented in Supplementary Table 4.

Costs. The reagent costs for these experiments (<A$1000 per individual) were less than current commercially available sequencing costs for *COL1A1*/*COL1A2* for OI and *FBN1* for MFS. Analysis time was not considered in the costings of this proof-of-concept study.

Discussion

This study demonstrates that WES is an efficient and sensitive method for detecting coding variants, indels and large deletions in *COL1A1, COL1A2* and *FBN1* in patients with OI and MFS. However, both capture technology and analysis programs affect the sensitivity of mutation detection.

Detection of three mutations proved challenging. The first problematic mutation was a single-base deletion (*COL1A1*: c.3765delC:p.1256A>I), missed initially due to alignment difficulties. It was detected when alignment was performed by BWA[44] but with only one of three versions of Novoalign (interestingly, not the most recent release).[45] This mutation highlights the importance of using more than one analysis method for MPS data.

The second challenging mutation was a large deletion of much of the 5′-UTR region (including all of exon 1) of *FBN1*. Analysis approaches to capture small and large indels in MPS data vary according to the size of the indel. For small indels (<15 bp) GATK uses an integrated program called DINDEL that detects indels using alignment discrepancies.[46] For larger indels, one needs to use programs that detect break points and/or relative sequencing depth. CREST uses a process called ‘soft-clipping’ to detect deletions from 40 to several hundred bp in length[47] through the detection of break points and will

Table 3 Comparison of five different capture platforms (using the manufacturer's browser extensible data files) for targeting of exons in known OI genes, *FBN1* and genes associated with a Marfanoid phenotype

Known OI genes	*Number of exons*	*Illumina Nextera Rapid Capture Exome*	*Illumina Nextera Rapid Capture Expanded Exome*	*Agilent SureSelect All Exon V5 + UTRs*	*NimbleGen SeqCap EZ V 3.0*	*Illumina TruSeq Exome Enrichment Kit v2.0*
COL1A1	51	100%	96% No target capture of exons 9 or 47	90% No target capture of exon 42 Targeted <90% of a further 10 exons	85.4% Targeted <90% of 18 exons	96% No target capture of exons 9 or 47
COL1A2	52	100%	98% No target capture of exon 35	94% Targeted <90% 9 exons	91% Targeted <90% 12 exons	98% No target capture of exon 35
CRTAP	7	100%	86% No target capture of exon 2	93% Targeted <90% 2 exons	86% Targeted <90% 2 exons	86% No target capture of exon 2
LEPRE1	15	100%	94% Targeted <90% 2 exons	96% Targeted <90% 3 exons	86% Targeted <90% 7 exons	94% Targeted <90% 2 exons
FKBP10	11	100%	91% No target capture of exon 7	86% Targeted <90% 7 exons	91% Targeted <90% 3 exons	91% No target capture of exon 7
SP7	2	100%	100%	100%	88.78% Targeted <90% exon 1	100%
SERPINF1	7	100%	100%	91% Targeted <90% 3 exons	100%	100%
PPIB	5	100%	54% No target capture of exons 1 or 2 Targeted <90% of exon 3	92% Targeted <90% 2 exons	87% Targeted <90% 2 exons	54% No target capture of exons 1 or 2 Targeted <90% of exon 3
PLOD2	20	100%	97% Targeted <90% exon 10	97% Targeted <90% 2 exons	94% Targeted <90% 4 exons	97% Targeted <90% exon 10
SERPINH1	4	100%	98%	79% Targeted <90% 3 exons	94% Targeted <90% exon 3	98%
BMP1	21	100%	100%	95% Targeted <90% 4 exons	94% Targeted <90% 5 exons	100%
LRP5	24	100%	92% No target capture of exons 1 or 9	89% Targeted <90% 12 exons	93% Targeted <90% 5 exons	92% No target capture of exons 1 or 9
IFITM5	2	100%	100%	84% Targeted <90% exon 1	100%	100%
TMEM38B	6	100%	100%	97% <90% of exon 6	98% <90% of exon 6	100%
WNT1	4	100%	75% No target capture of exon 3	96% <90% of exon 2	87% <90% of two exons	75% No target capture of exon 3

Known genes associated with Marfan syndrome or Marfanoid phenotype	*Number of exons*	*Illumina Nextera Rapid Capture Exome*	*Illumina Nextera Rapid Capture Expanded Exome*	*Agilent SureSelect All Exon V5 + UTRs*	*NimbleGen SeqCap EZ V 3.0*	*Illumina TruSeq Exome Enrichment Kit v2.0*
FBN1	66	100%	100%	97% Targeted <90% 8 exons	95% Targeted <90% 11 exons	100%
TGFBR1	9	89% No target capture of exon 1	88% No target capture of exon 1	92% Targeted <90% of two exons	95% Targeted <90% exon 1	88% No target capture of exon 1
TGFBR2	8	100%	87% No target capture of exon 2	97% Targeted <90% exon 4	92% Targeted <90% of two exons	87% No target capture of exon 2

Abbreviations: OI, Osteogenesis imperfecta; UTRs, untranslated regions.

therefore only detect deletions in WES data if the break points occur within captured regions. In our sample, the break points occurred outside the exon, hence CREST was unable to detect the deletion. Both ExomeCopy and ExomeDepth use read depth data to detect copy number variants and thus identify very large deletions (100s–1000s bases); in our case the mutation was only detected using ExomeDepth.

The final problematic mutation was a coding variant mutation (*FBN1*:c.2723G > T:p.908C > F) not detected by GATK. Although the overall depth of coverage at this location was 20-fold, there were only two reads for the alternate allele, which was insufficient for GATK identification. It is hard to understand the origin of this strand bias. Greater coverage depth would have resulted in a higher absolute number of alternate allele reads, but the ratio of reference to alternate alleles may have remained below GATK threshold cutoff for detection. Alternatively, altering the sensitivity thresholds within IGV would have allowed visual scanning of the data and detection in this manner, although the results would have required further validation.

These challenges highlight the importance of strong bioinformatics involvement in analyzing the sequencing data, including the use of different programs (and even different versions of programs) and algorithms to maximize detection. The repetitive and homologous nature of exons of *COL1A1* and *COL1A2* makes them particularly susceptible to multi-mapping artefacts, and a robust bioinformatics analysis is required to identify these false positives. Diseases characterized by a high frequency of deletions will require particularly careful analysis.

Files with proof-of-concept cases like these are also very useful for assessing the quality of updated software programs.

We evaluated specificity in an unrelated control population with high bone mineral density, identifying two variants of unknown significance when analyzing the major genes for these disorders, giving a specificity of $>99\%$. As discussed in the results, neither variant had been previously reported in association with these disorders and, if we were using this approach prospectively, we would have reported them as having unknown significance.

Our data also highlight the importance of clear communication between the sequencing laboratory and the clinical team requesting genetic testing, as the optimal exome capture platform varied according to the clinical question under consideration. Our data illustrate that for screening of known genes for diagnostic purposes it is imperative to review the target capture of the available platforms to determine which best targets the exons of interest. This can be done visually, using IGV; more formally, the overlap of the exonic regions of each gene under consideration and the target capture of the platform of choice can be compared and quantified bioinformatically, as we present here (Supplementary Table 4). For AD OI, robust coverage of *COL1A1* and *COL1A2* is most critical; theoretically the Illumina-Rapid targets 100% of the coding region of these genes, as well as all exons of all other known OI genes. However, the reported target region for this recently-released platform is strictly the exome, and surrounding UTR mutations will not be captured—of relevance, for example, for OI type V (caused by a 5′UTR mutation), which would be targeted by Agilent, Illumina-Rapid Exp and Illumina-TruSeq. For MFS, all *FBN1* exons were targeted completely by all three Illumina capture platforms. With respect to other Marfanoid genes, none of the Illumina platforms target the first exon in *TGFBR1*; the second exon of *TGFBR2* is also not targeted for two additional platforms. From available data, Loeys–Dietz syndrome patients usually carry mutations in the later exons of these genes;[48] thus, these missing exons may be of lower clinical relevance. We have highlighted the differences between the remaining platforms in **Table 3**.

In considering target capture for known genes, although in absolutely percentage terms the overall region of these genes captured by each platform might seem similar, Illumina platforms generally either completely target or completely miss exons; in contrast, both NimbleGen and Agilent platforms provide some capture of most exons but often only partially (see **Table 3** and Supplementary Table 3). Knowing *a priori* which exons are not comprehensively captured by WES should prompt Sanger sequencing of these missing exons should a mutation not be identified. Thus for clinical utility, use of less comprehensive platforms would necessitate additional Sanger sequencing of a greater number of exons before a mutation could be confidently excluded. When embarking on gene discovery, however, targeting a larger percentage of both the exome and surrounding UTRs may be of greater relevance. Again, the extent to which the exome and surrounding UTRs are captured by any particular platform can be assessed and quantified bioinformatically.

We would highlight that target capture represents an ideal that may not be attained with 'real-life' experience. For comparison, we have presented our experience with exomes captured and sequenced in-house from individuals with unrelated disorders (Supplementary Table 4). These data should not be over-interpreted—for example, protocols were updated and fine-tuned over time to improve performance—but we include the data to emphasize the importance of per-experiment and per-individual assessment of capture (including depth of coverage) in mutation screening. This obviously pertains not just to OI and MFS but to any disease for which these methods might be applied.

To our knowledge, this is the first report testing the sensitivity of WES in detecting previously reported mutations for both OI and MFS. A recent study used custom targeted capture for MFS, targeting only *FBN1*, *TGFBR1* and *TGFBR2*, in a cohort of individuals with MFS or Loeys–Dietz syndrome. In the previously untested MFS cohort, a mutation in *FBN1* was ultimately found in 92%, although only 86% had a mutation found by MPS; thus, the sensitivity of MPS in this study is calculated as 93%.[30] Other studies using targeted capture and MPS for diagnostic purposes have included adult polycystic kidney disease,[49] cardiomyopathy,[50] mitochondrial disorders[51] and retinitis pigmentosa.[52] Targeted capture limited to known genes is more economical with greater depth of coverage over the targeted regions for the same number of total reads as WES.[50] The disadvantages of this approach include the greater cost of custom-designed capture and the lack of flexibility in the event of future gene discovery. Further, new genes cannot be identified in a family negative for known genes without further sequencing. Of note, WES has been used to identify several new genes for OI.[10,13,14]

Although targeted MPS performed comparably with Sanger sequencing in previous studies, some authors have argued that WES is too inconsistent in its exome coverage for diagnostic purposes.[50] Our data suggest otherwise. Despite the unusual repetitious gene structure of both *COL1A1* and *COL1A2,* with potentially difficult alignment, we were able to detect all mutations in the OI cohort. Although our study does not involve large numbers, the detection rate of 9/10 *FBN1* mutations is similar to that reported by targeted capture in MFS,[30] although we would acknowledge that this prior study was performed on previously untested samples. As discussed above, for clinical utility per-experiment QC should include target exon capture and depth of coverage to enable assessment of the probability that a mutation may have been missed. Our data also demonstrate the importance of careful capture platform selection and analysis strategy to maximize the sensitivity of WES in mutation detection. To this end, discussion between the caring clinicians and the sequencing and bioinformatics teams is imperative.

In Australia, mutation screening of *COL1A1*/*COL1A2* costs ∼$A1850, and testing for *FBN1* is $A1500. Even at this early stage in MPS development, WES is a less expensive and more time-efficient means of mutation detection for these diseases than conventional commercial sequencing services, and further provides a comprehensive screen of all known genes for these phenotypes.

Although not currently the standard of care, genetic testing can guide clinical management and provide psychological benefits for both OI and MFS populations, particularly in those cases where the clinical phenotype is ambiguous. Testing at-risk family members can ensure appropriate screening is provided to mutation carriers and, conversely, eliminate unnecessary screening and distress in non-carriers. There is

general concordance among MFS experts[23] and some agreement in OI communities[20] that affordable, sensitive and rapid genetic testing will provide clinical utility and enhance psychosocial adjustment for these families. Further, mutation identification may influence medical management. For example, given the mineralization defect seen on biopsy in OI type VI, a confirmed *SERPINF1* mutation in a child with OI may suggest that bisphosphonates (the most commonly used medication in individuals with OI) might not be appropriate management.

In conclusion, using WES we identified previously reported mutations in 13/13 individuals with OI and 9/10 individuals with MFS and showed that this approach is sensitive, specific, efficient and comparatively cost-effective for detecting mutations in these patient cohorts and allows for further gene screening and/or identification. However, successful mutation identification requires careful consideration of platform selection and a variety of bioinformatics approaches to maximize sensitivity.

Materials and Methods

We performed exome capture and MPS on genomic DNA from 13 participants with OI and 10 participants with MFS in whom mutations had been previously identified by conventional means such as Sanger sequencing.

Patient cohorts. De-identified, previously tested, samples from individuals meeting diagnostic criteria for OI or MFS were provided by authors with clinical expertise in these conditions (PC, AZ for OI; BL, MW, PW for MFS). Ethics permission was obtained for performing WES and analyzing target genes for both disorders (UQ #201100876).

Laboratory techniques. Sequencing libraries were constructed using a modification of the Illumina TruSeqDNA sample preparation kit. Briefly, 1.6 μg of genomic DNA was sheared to average fragment size of 200 bp using the Covaris E220 (Covaris, Woburn, MA, USA). Fragments were purified using AmpPureXP beads (Beckman Coulter, Brae, CA, USA) to remove small products ($<$100 bp), yielding 1 μg of material that was end-polished, A-tailed and adapter ligated according to the manufacturer's protocol. Libraries were subjected to minimal PCR cycling and quantified using the Agilent High Sensitivity DNA assay (Agilent, Santa Clara, CA, USA). Libraries were combined into pools of six for solution phase hybridization using the Illumina (Illumina, San Diego, CA, USA) TruSeq Exome Enrichment Kit. Captured libraries were assessed for both quality and yield using the Agilent High Sensitivity DNA assay and KAPA Library Quantification Kit. MPS was performed with six samples per flow cell lane using the Illumina HiSeq2000 platform and version 3 sequencing-by-synthesis reagents to generate 100 bp paired-end reads (2×100PE).

Illumina Data Analysis Pipeline software (CASAVA 1.8.2) was used for de-multiplexing and initial base calling. Sequence data were aligned to the current build of the human genome (hg19, released in February 2009) using the Novoalign alignment tool (V 2.07.18 and V 2.08.02)[45] and the Burrows–Wheeler Aligner (BWA v 0.7.3 SW).[44] Sequence alignment files were converted using SAMtools (v0.1.16) and Picard tools (v1.42), which includes the MarkDuplicates tool. Single nucleotide polymorphisms (SNPs) and indels were called using GATK (v2.2–3 and v2.4–9)[53] and annotated using ANNOVAR.

Screening for coding mutations and indels. Further analysis of sequence data was performed using custom scripts employing R and Bioconductor. Variants with sufficient coverage (minimum depth of coverage for variant calling: $>$5-fold for homozygous variants, $>$10-fold for heterozygous variants) were retained and platform and alignment artifacts were excluded. Further, retained variants all passed GATK Variant Quality Score Recalibration (incorporating quality parameters of sequencing depth and quality scores at the variant position, maximal length of the homopolymer run and strand bias).[53]

Filtering initially selected only variants present in the target genes (*COL1A1* and *COL1A2* in OI samples and *FBN1* for MFS samples). Data were then filtered to view good quality variants predicted to be of potentially damaging consequence ('nonsynonymous SNV', 'splicing', 'frameshift substitution', 'stopgain SNV', 'stoploss SNV') using RefSeq, Ensembl and UCSC transcripts. Further filtering excluded variants with a minor allele frequency (MAF) $>$0.001 (a liberal threshold chosen based on population frequencies for both diseases) observed in NCBI dbSNP (release 135), 1000 Genomes,[54] 1000 Genomes small indels (called using the DINDEL program,[46] the SNPs of 46 Genomes, and other whole exomes from over 1200 control samples run internally using similar capture technology.

Deletion screening was performed on all samples using three programs: CREST (Clipping Reveals Structure),[47] ExomeCopy[55] and ExomeDepth.[56] ExomeDepth was performed using 10 individuals from the same sequencing run using the program's default settings.

We compared our sequencing results with the previous mutation screening of these samples.

Control population. To determine the specificity of WES we evaluated the data on 487 individuals with high bone mass (BMD *z*-scores between $+1.5$ to $+4$) who have undergone exome sequencing in-house with similar (or identical) protocols. Data were analyzed looking for rare (MAF$<$0.001) variants of potentially damaging consequence as defined above in *FBN1, COL1A1* and *COL1A2*. Deletion analysis was not performed on control samples.

Comparison of capture technologies. The theoretical target capture of *FBN1*, *TGFBR1, TGFBR2, COL1A1, COL1A2* and other known OI genes was analyzed for five exome capture platforms: Agilent SureSelect (Agilent), NimbleGen SeqCap EZ v3.0 (Roche NimbleGen, Madison, WI, USA), Illumina TruSeq (Illumina-TruSeq), Illumina Nextera Rapid Capture Expanded Exome (Illumina-Rapid Exp) and Illumina Nextera Rapid Capture Exome (Illumina-Rapid) by downloading the target capture regions or Browser Extensible Data files from the websites of each manufacturer. A custom R script compared the percentage overlap of the gene exon coding regions (CDS) using the Bioconductor package TxDb.Hsapiens.UCSC.hg19.knownGene with the target capture regions of each platform.

Using data from this study and from 209 individuals sequenced for unrelated conditions using the Illumina-TruSeq

and Nimblegen platforms, we assessed the capture achieved (at $>10\times$ depth) for these candidate genes by calculating the percentage coverage of the Exon CDS regions of each gene obtained from the Bioconductor package TxDb.Hsapiens.UCSC.hg19.knownGene, using Novoalign (V 2.08.02).

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to thank Sharon Song, Marina Donskoi and Lisa Anderson for technical support. AML is supported by a University of Queensland postgraduate scholarship. AZ is funded by the Royal Children's Hospital Foundation, ANZ Trustee's Children's Medical Research Establishment Grant. MAB is supported by an NHMRC senior fellowship.

Web resources:

1000 Genomes: http://www.1000genomes.org/
46 Genomes from Complete Genomics: http://www.completegenomics.com/sequence-data/download-data/
Agilent SureSelect: (https://earray.chem.agilent.com/suredesign/index.htm)
ANNOVAR: http://www.openbioinformatics.org/annovar/
BEAGLE Utilities program Cluster2haps: http://faculty.washington.edu/browning/beagle/beagle_3.3.2_31Oct11.pdf
Bioconductor package TxDb.Hsapiens.UCSC.hg19.knownGene: (http://www.bioconductor.org/packages//2.10/data/annotation/html/TxDb.Hsapiens.UCSC.hg19.knownGene.html)
CASAVA: http://www.illumina.com/software/genome_analyzer_software.ilmnCONDEL: http://bg.upf.edu/condel/home
Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine (dbSNP Build ID: dbSNP137): http://www.ncbi.nlm.nih.gov/projects/SNP/
Exome Variants Analysis and Reporting (EVAR): www.exome.info
GATK: http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit
Illumina TruSeq: http://support.illumina.com/sequencing/sequencing_kits/truseq_exome_enrichment_kit/downloads.ilmn
MutationTaster: www.mutationtaster.org
Nextera Rapid Capture Exome Targeted Regions: http://support.illumina.com/downloads.ilmn
Nextera Rapid Capture Expanded Exome Targeted Regions: (http://support.illumina.com/downloads/nextera_rapid_capture_expanded_exome_targeted_regions_manifest_bed.ilmn)
NHLBI Exome Sequencing Project: http//:evs.gs.washington.edu/EVS
NimbleGen: (http://www.nimblegen.com/products/seqcap/ez/v3/index.html)
Novalign alignment tool: www.Novocraft.comOnline Mendelian Inheritance in Man (OMIM): http://www.omim.org
Package hapFabia—Bioconductor: http://www.bioconductor.org/packages/2.11/bioc/html/hapFabia.htmlPicard tools (v1.42): http://picard.sourceforge.net
PolyPhen: http://genetics.bwh.harvard.edu/pph/SAMtools: http://samtools.sourceforge.net/
SIFT: http://sift.jcvi.org/
UCSC genome browser: http://genome.ucsc.edu/
UK10K project: http://www.uk10k.orgUniProt: http://www.uniprot.org/uniprot/

References

1. Van Dijk FS, Pals G, Van Rijn RR, Nikkels PG, Cobben JM. Classification of Osteogenesis Imperfecta revisited. *Eur J Med Genet* 2010;**53**:1–5.
2. Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S *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.
3. Engel J, Prockop DJ. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu Rev Biophys Biophys Chem* 1991;**20**:137–152.
4. Kuurila K, Grenman R, Johansson R, Kaitila I. Hearing loss in children with osteogenesis imperfecta. *Eur J Pediatr* 2000;**159**:515–519.
5. Morello R, Bertin TK, Chen Y, Hicks J, Tonachini L, Monticone M *et al.* CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell* 2006;**127**:291–304.
6. Cabral WA, Barnes AM, Adeyemo A, Cushing K, Chitayat D, Porter FD *et al.* A founder mutation in LEPRE1 carried by 1.5% of West Africans and 0.4% of African Americans causes lethal recessive osteogenesis imperfecta. *Genet Med* 2012;**14**:543–551.
7. van Dijk FS, Nesbitt IM, Zwikstra EH, Nikkels PG, Piersma SR, Fratantoni SA *et al.* PPIB mutations cause severe osteogenesis imperfecta. *Am J Hum Genet* 2009;**85**:521–527.
8. Christiansen HE, Schwarze U, Pyott SM, AlSwaid A, Al Balwi M, Alrasheed S *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.
9. Alanay Y, Avaygan H, Camacho N, Utine GE, Boduroglu K, Aktas D *et al.* Mutations in the gene encoding the RER protein FKBP65 cause autosomal-recessive osteogenesis imperfecta. *Am J Hum Genet* 2010;**86**:551–559.
10. Homan EP, Rauch F, Grafe I, Lietman C, Doll JA, Dawson B *et al.* Mutations in SERPINF1 cause osteogenesis imperfecta type VI. *J Bone Miner Res* 2011;**26**:2798–2803.
11. Becker J, Semler O, Gilissen C, Li Y, Bolz HJ, Giunta C *et al.* Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. *Am J Hum Genet* 2011;**88**:362–371.
12. Martinez-Glez V, Valencia M, Caparros-Martin JA, Aglan M, Temtamy S, Tenorio J *et al.* Identification of a mutation causing deficient BMP1/mTLD proteolytic activity in autosomal recessive osteogenesis imperfecta. *Hum Mutat* 2012;**33**:343–350.
13. Semler O, Garbes L, Keupp K, Swan D, Zimmermann K, Becker J *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.
14. Cho TJ, Lee KE, Lee SK, Song SJ, Kim KJ, Jeon D *et al.* A single recurrent mutation in the 5'-UTR of IFITM5 causes osteogenesis imperfecta type V. *Am J Hum Genet* 2012;**91**:343–348.
15. Moffatt P, Gaumond MH, Salois P, Sellin K, Bessette MC, Godin E *et al.* Bril: a novel bone-specific modulator of mineralization. *J Bone Miner Res* 2008;**23**:1497–1508.
16. Lapunzina P, Aglan M, Temtamy S, Caparros-Martin JA, Valencia M, Leton R *et al.* Identification of a frameshift mutation in Osterix in a patient with recessive osteogenesis imperfecta. *Am J Hum Genet* 2010;**87**:110–114.
17. Shaheen R, Alazami AM, Alshammari MJ, Faqeih E, Alhashmi N, Mousa N *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.
18. Keupp K, Beleggia F, Kayserili H, Barnes AM, Steiner M, Semler O *et al.* Mutations in WNT1 cause different forms of bone fragility. *Am J Hum Genet* 2013;**92**:565–574.
19. Warman ML, Cormier-Daire V, Hall C, Krakow D, Lachman R, LeMerrer M *et al.* Nosology and classification of genetic skeletal disorders: 2010 revision. *Am J Med Genet A* 2011;**155A**:943–968.
20. Forlino A, Cabral WA, Barnes AM, Marini JC. New perspectives on osteogenesis imperfecta. *Nat Rev Endocrinol* 2011;**7**:540–557.
21. Judge DP, Dietz HC. Marfan's syndrome. *Lancet* 2005;**366**:1965–1976.
22. Marfan A. Un cas de déformation congénitale des quartre membres, plus prononcée aux extrémitiés, caractérisée par l'allongement des os avec un certain degré d'amincissement [A case of congenital deformation of the four limbs, more pronounced at the extremities, characterized by elongation of the bones with some degree of thinning]. *Bull memoires Soc med hôsp Paris* 1896;**13**:220–226.
23. Loeys BL, Dietz HC, Braverman AC, Callewaert BL, De Backer J, Devereux RB *et al.* The revised Ghent nosology for the Marfan syndrome. *J Med Genet* 2010;**47**:476–485.
24. Porciani MC, Attanasio M, Lepri V, Lapini I, Demarchi G, Padeletti L *et al.* Prevalence of cardiovascular manifestations in Marfan syndrome. *Ital Heart J* 2004;**5**:647–652.
25. Murdoch JL, Walker BA, Halpern BL, Kuzma JW, McKusick VA. Life expectancy and causes of death in the Marfan syndrome. *New Engl J Med* 1972;**286**:804–808.
26. Silverman DI, Burton KJ, Gray J, Bosner MS, Kouchoukos NT, Roman MJ *et al.* Life expectancy in the Marfan syndrome. *Am J Cardiol* 1995;**75**:157–160.

27. Brooke BS, Habashi JP, Judge DP, Patel N, Loeys B, Dietz 3rd HC. Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. *New Engl J Med* 2008;**358**:2787–2795.
28. Lee B, Godfrey M, Vitale E, Hori H, Mattei MG, Sarfarazi M *et al.* Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes. *Nature* 1991;**352**:330–334.
29. Loeys B, De Backer J, Van Acker P, Wettinck K, Pals G, Nuytinck L *et al.* Comprehensive molecular screening of the FBN1 gene favors locus homogeneity of classical Marfan syndrome. *Hum Mutat* 2004;**24**:140–146.
30. Baetens M, Van Laer L, De Leeneer K, Hellemans J, De Schrijver J, Van De Voorde H *et al.* Applying massive parallel sequencing to molecular diagnosis of Marfan and Loeys-Dietz syndromes. *Hum Mutat* 2011;**32**:1–10.
31. Robinson PN, Arteaga-Solis E, Baldock C, Collod-Beroud G, Booms P, De Paepe A *et al.* The molecular genetics of Marfan syndrome and related disorders. *J Med Genet* 2006; **43**:769–787.
32. Collod-Beroud G, Beroud C, Ades L, Black C, Boxer M, Brock DJ *et al.* Marfan Database (third edition): new mutations and new routines for the software. *Nucleic Acids Res* 1998;**26**:229–3.
33. Hilhorst-Hofstee Y, Hamel BC, Verheij JB, Rijlaarsdam ME, Mancini GM, Cobben JM *et al.* The clinical spectrum of complete FBN1 allele deletions. *Eur J Hum Genet* 2011;**19**:247–252.
34. Pyeritz R. The Marfan Syndrome. In: Royce P, Steinmann B (eds) *Connective tissue and its heritable disorders* New York: Wiley-Liss, 1993, pp437–468.
35. Rantamaki T, Kaitila I, Syvanen AC, Lukka M, Peltonen L. Recurrence of Marfan syndrome as a result of parental germ-line mosaicism for an FBN1 mutation. *Am J Hum Genet* 1999; **64**:993–1001.
36. Loeys BL, Chen J, Neptune ER, Judge DP, Podowski M, Holm T *et al.* A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. *Nat Genet* 2005;**37**:275–281.
37. Stheneur C, Collod-Beroud G, Faivre L, Gouya L, Sultan G, Le Parc JM *et al.* Identification of 23 TGFBR2 and 6 TGFBR1 gene mutations and genotype-phenotype investigations in 457 patients with Marfan syndrome type I and II, Loeys-Dietz syndrome and related disorders. *Hum Mutat* 2008;**29**:E284–E295.
38. Akutsu K, Morisaki H, Okajima T, Yoshimuta T, Tsutsumi Y, Takeshita S *et al.* Genetic analysis of young adult patients with aortic disease not fulfilling the diagnostic criteria for Marfan syndrome. *Circ J* 2010;**74**:990–997.
39. van de Laar IM, Oldenburg RA, Pals G, Roos-Hesselink JW, de Graaf BM, Verhagen JM *et al.* Mutations in SMAD3 cause a syndromic form of aortic aneurysms and dissections with early-onset osteoarthritis. *Nat Genet* 2011;**43**:121–126.
40. Hung CC, Lin SY, Lee CN, Cheng HY, Lin SP, Chen MR *et al.* Mutation spectrum of the fibrillin-1 (FBN1) gene in Taiwanese patients with Marfan syndrome. *Ann Hum Genet* 2009;**73**:559–567.
41. Loeys B, Nuytinck L, Delvaux I, De Bie S, De Paepe A. Genotype and phenotype analysis of 171 patients referred for molecular study of the fibrillin-1 gene FBN1 because of suspected Marfan syndrome. *Arch Intern Med* 2001;**161**:2447–2454.
42. Cooper DN, Krawczak M, Antonorakis SE. The nature and mechanisms of human gene mutation. In: Scriver C, Beaudet AL, Sly WS, Valle D) (eds) *The Metabolic and Molecular Bases of Inherited Disease*. 7th edn (McGraw-Hill: New York, 1995, pp259–291.
43. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative GENOMICS Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;**14**:178–192.
44. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;**26**:589–595.
45. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform* 2010;**11**:473–483.
46. Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. Dindel: accurate indel calls from short-read data. *Genome Res* 2011;**21**:961–973.
47. Wang J, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J *et al.* CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat Methods* 2011;**8**:652–654.
48. Beroud C, Collod-Beroud G, Boileau C, Soussi T, Junien C. UMD (Universal mutation database): a generic software to build and analyze locus-specific databases. *Hum Mutat* 2000;**15**:86–94.
49. Qi XP, Du ZF, Ma JM, Chen XL, Zhang Q, Fei J *et al.* Genetic diagnosis of autosomal dominant polycystic kidney disease by targeted capture and next-generation sequencing: utility and limitations. *Gene* 2013;**516**:93–100.
50. Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, van den Berg MP *et al.* Targeted next-generation sequencing can replace sanger sequencing in clinical diagnostics. *Hum Mutat* 2013;**34**:1035–1042.
51. Dinwiddie DL, Smith LD, Miller NA, Atherton AM, Farrow EG, Strenk ME *et al.* Diagnosis of mitochondrial disorders by concomitant next-generation sequencing of the exome and mitochondrial genome. *Genomics* 2013;**102**:148–156.
52. Fu Q, Wang F, Wang H, Xu F, Zaneveld JE, Ren H *et al.* Next generation sequencing based molecular diagnosis of a Chinese patient cohort with autosomal recessive Retinitis Pigmentosa. *Invest Ophthalmol Vis Sci* 2013;**54**:4158–4166.
53. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;**20**:1297–1303.
54. Genomes Project CAbecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010;**467**: 1061–1073.
55. Love MI, Mysickova A, Sun R, Kalscheuer V, Vingron M, Haas SA. Modeling read counts for CNV detection in exome sequencing data. *Stat Appl Genet Mol Biol* 2011;**10**:1.
56. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S *et al.* A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* 2012;**28**:2747–2754.

Supplementary Information accompanies the paper on the BoneKEy website (http://www.nature.com/bonekey)