HGGA, Volume *2*

Supplemental information

Biallelic variants in *MESD*, which encodes a WNT-signaling-related protein, in four new families with recessively inherited osteogenesis imperfecta

Thao T. Tran, Rachel B. Keller, Brecht Guillemyn, Melanie Pepin, Jane E. Corteville, Samir Khatib, Mohammad-Sadegh Fallah, Sirous Zeinali, Fransiska Malfait, Sofie Symoens, Paul Coucke, Peter Witters, Elena Levtchenko, Hamideh Bagherian, Deborah A. Nickerson, Michael J. Bamshad, Jessica X. Chong, University of Washington Center for Mendelian Genomics, and Peter H. Byers

SUPPLEMENTARY MATERIAL & METHODS

Informed Consent: Written informed consent was obtained for all individuals in this study.

Exome filtering: In order to identify the causative gene in Family 1, we used the same WES data set described by Pyott *et al.*[5] which included five unrelated infants with OI phenotypes and the unaffected sibling of one of them. This data set was used to identify one of the first families in which *WNT1* [MIM 164820] pathogenic variants were shown to cause OI.[5] In all five families represented, the affected infant was from a recurrent sibling set. Four of the families were thought to be consanguineous and the fifth was from a defined ethnic community. In the included sibling pair, the affected child was found to be homozygous for a missense variant in *WNT1*. [5] Separately, another infant in the group was found to have a homozygous 20 kilobase deletion that encompassed exon 4 of *TMEM38B* [MIM 611236] and flanking splice sites. Prior to this analysis, the specific *TMEM38B* deletion had been described in other infants from the same population.[6] The exome sequence data had no coverage in the region of exon 4 of *TMEM38B*, which confirmed the deletion.

We took two approaches to the data analysis. The first was to repeat the method described by Pyott *et al* [5] using one of the unsolved infants from a consanguineous family in the previously described dataset (1-II:6) (Figure 3C) as the unknown. Briefly, we filtered the WES data for variants matching the expected homozygous recessive pattern of inheritance. We used the other three infants in the dataset (unaffected, *WNT1*, and*TMEM38B*) as "unaffected" controls and assumed that none of the remaining infants shared a causative variant. The analysis identified 55,915 variants from the reference exome seen in one or more of the individuals who were sequenced. We removed variants found in 1-II:6 that were also found in the other three individuals which left 9198 variants. Removal of intergenic

variants and those that appeared more than 10 times in the Exome Sequencing Project Exome Variant Server (EVS) left 936 variants. All heterozygous variants were discarded, as the family history suggested homozygous recessive inheritance, which left 123 variants. After intronic variants and synonymous coding variants were discarded, there were 62 candidates left. Those variants with a conservation score (consScoreGERP) >2 were discarded, leaving 24 variants. Of the 24, 11 genes were identified as potential candidates based on allele frequencies in the ExAC database (Supplemental Table 2).

For the second approach, the WES data were annotated with the Variant Effect Predictor v89[7] and analyzed with GEMINI 0.20.2.[7] We first checked for predicted pathogenic variants in all genes known to underlie OI.[2] We then focused on the same highly consanguineous family (Figure 2; Family 1), and filtered the WES data for variants that matched the expected homozygous recessive pattern of inheritance. Variants unlikely to impact protein-coding sequence (for which GEMINI impact_severity = LOW), variants filter flagged by the Genome Analysis Toolkit (GATK) as low quality (quality score ≤ 30, long homopolymer run > 5, low quality by depth < 5, within a cluster of SNPs), and variants with an alternative allele frequency >0.0005 in any superpopulation in gnomAD (v2.0), the Exome Variant Server (EVS), 1000 Genomes (phase 3 release), or >0.05 in an internal database of ~6400 individuals were excluded. Individual genotypes with depth < 6 or genotype quality < 20 were treated as missing in analysis. Due to the severity of OI in this family, we excluded all variants that were found in the homozygous state in the gnomAD database as gnomAD contains only putatively healthy adults. This yielded twenty candidate genes.

Genomic & cDNA Sequencing: DNA was extracted from patient fibroblasts or blood using the DNeasy Blood & Tissue Kit (Qiagen). RNA was extracted from patient fibroblasts using the RNeasy Mini Kit (Qiagen) and cDNA was generated with the Superscript First Strand

Synthesis System (Thermofisher Scientific). For confirmatory sequencing of *MESD*, exon 3 was amplified from either genomic DNA or cDNA using AmpliTaq Gold Polymerase (Applied Biosystems). The cycling program was: 95°C for 12 minutes, 95 °C for 10 seconds, 61 °C for 40 seconds, 72 °C for 50 seconds for 35 cycles, then, 72 °C for 7 minutes. Amplicons were treated with ExoSAP according to a standard protocol. Sequencing reactions were assembled using Big Dye v3.1 (Applied Biosystems) with the following program: 96°C for 10 seconds, 50 °C for 5 seconds, 60°C for 4 minutes for 40 cycles. Sequencing was run on an ABI 3730. Sequences were analyzed using Mutation Surveyor v5.0 (Softgenetics). Primers used for genomic and cDNA sequencing of MESD were as follows: MESD Genomic Sense 5'-TGCTCTGACCCCTTAGCACC-3', MESD Genomic Antisense 5'- GGGCAAAGAGCTCTCCACG-3', MESD cDNA Sense 5'- GTCGGGTAAGCGCGTCTAGG-3', and MESD cDNA Antisense 5'-AAGAGCTCTCCACGTCCACC-3'.

Immunoblotting: Protein lysates were prepared from primary fibroblasts using a buffer containing 0.05M Tris-HCl at pH 8.0, 0.15M NaCl, 5mM EDTA, 1% NP-40, and a protease inhibitor cocktail at 4 °C. 50µg of lysate was loaded per lane and resolved on a 10% SDS-PAGE gel in loading buffer containing 7.7 mg/ml dithiothreitol (DTT). Proteins were transferred to an Amersham Protran Premium 0.45□m Nitrocellulose Membrane (GE Healthcare). Detection for western blot was performed with primary antibodies for MESD (sc-139397; Santa Cruz Biotechnology, Inc.) at 1:00, LRP5 (sc-390267; Santa Cruz Biotechnology, Inc.) at 1:200, LRP6 (sc-25317; Santa Cruz Biotechnology, Inc.) at 1:300, and GAPDH (ab179811: Abcam) at 1:100. Detection was achieved using Amersham ECL Western Blotting Detection Reagent (GE Healthcare).

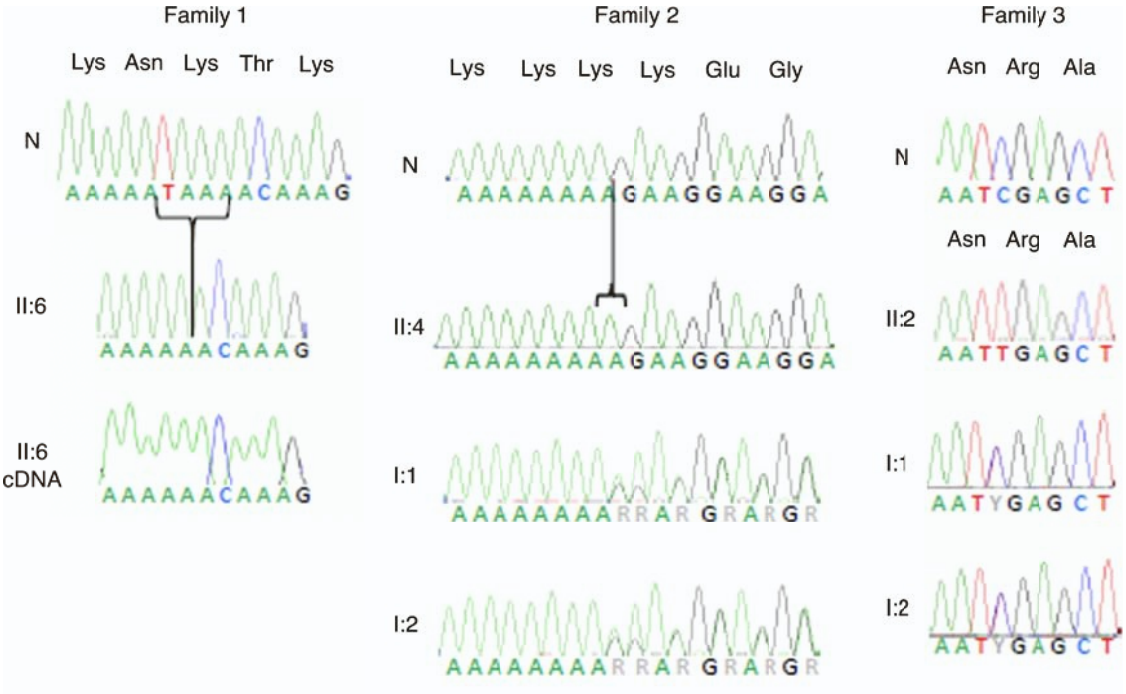
In Vitro Transcription/Translation (IVTT): In vitro transcription/translation (IVTT) was achieved using the 1-Step Human Coupled IVT Kit–DNA (ThermoFisher Scientific). Templates for IVTT were generated from an *MESD* cDNA expression plasmid (SC304244) from OriGene Technologies. Family-specific mutations were introduced into the *MESD* sequence using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies). Primers used for mutagenesis were:

QuikMESD_603- 606delTAAA_F 5'-
caaaggaggaggaagcaaagagaaaaaacaaagcaagacaa-3'; QuikMESD_603-
606delTAAA_R 5'- ttgtcttgctttgtttttttctctttgcttcctcctcctttg-3';
QuikMESD_632dupA_F 5'- caaagcaagacaagggcaaaaaaaaagaaggaaggagatct-3';
QuikMESD_632dupA_R 5'- agatctccttccttctttttttttgcccttgtcttgctttg-3';
QuikMESD_c676t_F 5'- ttattcccagctcaattttcttccttggaagaccgaga-3'; and
QuikMESD-c676t_R 5'- tctcggtcttccaaggaagaaaattgagctgggaataa-3'.

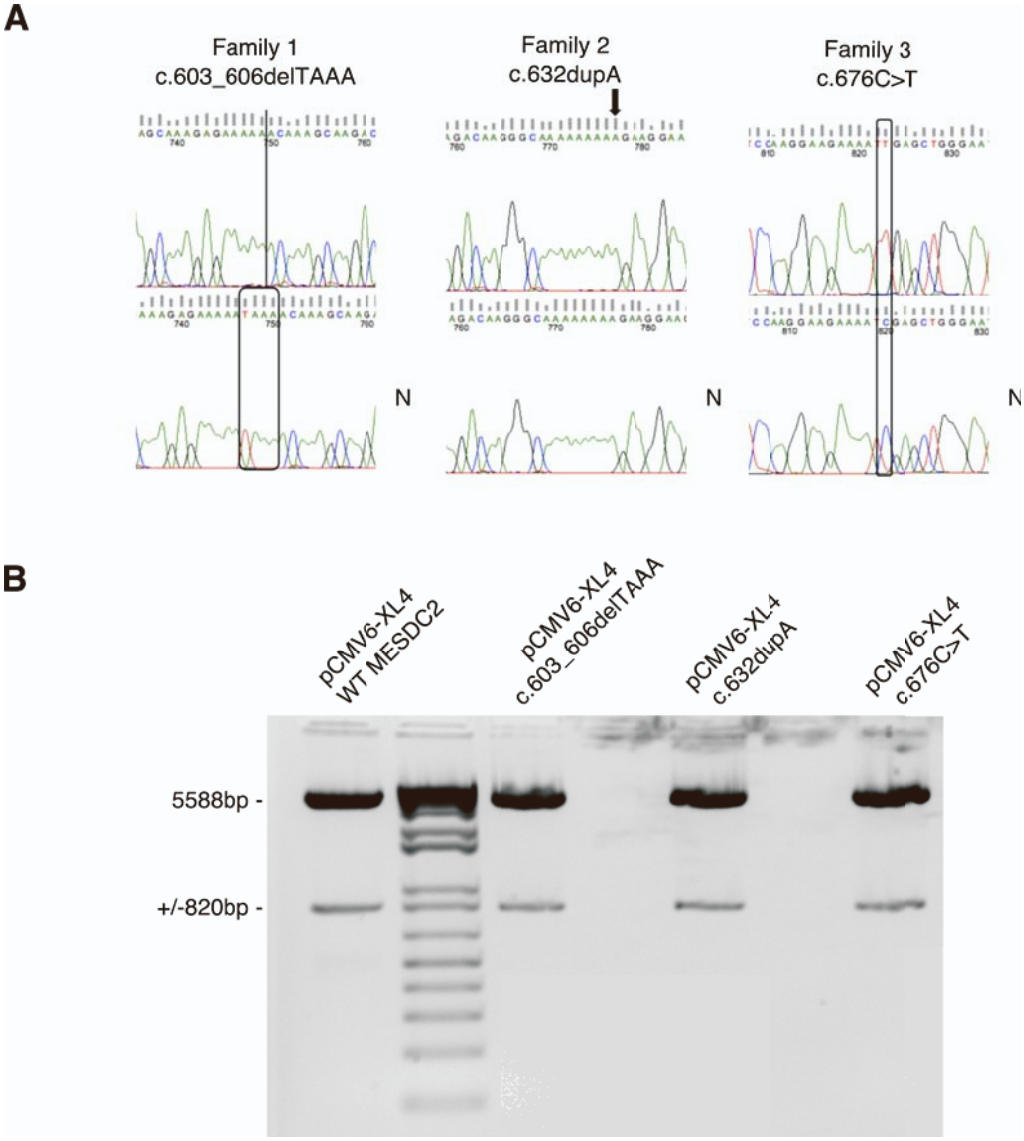
Fragments containing the cDNA sequence with upstream T7 promotors were liberated using Sac-I (New England BioLabs) and gel purified using the QIAquick Gel Extraction Kit (Qiagen). Approximately 125ng of linearized template was used in IVTT reactions that were incubated for 6 hours at 30°C. 2microliters of each IVT reaction was run on a 10% SDS-PAGE gel in loading buffer containing 7.7 mg/ml dithiothreitol (DTT). Western Blot was performed using primary antibody against MESD (sc-139397; Santa Cruz Biotechnology, Inc.) at 1:100 followed by HRP-Conjugated Goat Anti-Rabbit IgH H&L (ab6721; Abcam) secondary antibody at 1:20,000. Detection was achieved using Amersham ECL Western Blotting Detection Reagent (GE Healthcare).

Immunocytochemistry (ICC): Primary dermal fibroblasts from I-II:6 and from a control patient (A8) cultured in DMEM 10% FBS were fixed in cold methanol for 10 minutes and incubated with 2.5% normal horse serum/PBS (Vector Laboratories) for 20 minutes at room temperature to block nonspecific binding of antibodies. The cells were incubated with primary antibody to MESD (sc-139397; Santa Cruz Biotechnology, Inc.) at 1/50 at 4°C overnight. This was followed by incubation with AlexaFluor488 Goat Anti-Rabbit IgH (H&L) (A11034; ThermoFisher Scientific), 1:500, at room temperature for 1 hour. This stepwise staining process was repeated for KDEL (ab176333; Abcam) at 1:100 with AlexaFluor594 Goat Anti-Rabbit IgH (H&L) (A11037; ThermoFisher Scientific) at 1:800. The cells were counterstained with DAPI and mounted on slides. Images were acquired with a Zeiss Imager A1.

SUPPLEMENTARY TABLES AND FIGURES



Supplemental Figure 1. MESD confirmational sequencing. Variant sequences from available family members confirming the homozygous small deletion (c.603-606delTAAA) in Family 1 (both in proband genomic DNA and cDNA), the duplication (c.632dupA) in Family 2 (in the homozygous proband and in both heterozygous parents), and the substitution mutation (c.676C>T) in Family 3 (in the homozygous proband and in both heterozygous parents)(the proband in Family 4 carries the same homozygous mutation). The normal reference sequence sequences are labeled 'N'.



Supplemental Figure 2. IVTT DNA Templates. (A) Mutagenesis introducing family variants in MESD was confirmed by Sanger sequencing. The normal reference sequences are labeled 'N'. (B) Sac-I digestion of T7-MESD cDNA fragments away from pCMV6-XL4 expression plasmid backbone.

Supplemental Table 1: Genes associated with OI.

Gene Name	MIM
COL1A1	MIM 120150
COL1A2	MIM 120160
IFITM5	MIM 614757
BMP1	MIM 112264
CREB3L1	MIM 616215
CRTAP	MIM 605497
FKBP10	MIM 607063
MBTPS2	MIM 300294
P3H1	MIM 610339
PLOD2	MIM 301865
PPIB	MIM 123841
SEC24D	MIM 607186
SERPINF1	MIM 172860
SERPINH1	MIM 600943
SP7	MIM 606633
SPARC	MIM 182120
TENT5A	MIM 611357
TMEM38B	MIM 611236
WNT1	MIM 164820
MESD	MIM 618644
CCDC134	?
KDELR2	MIM619131

Supplemental Table 2. Candidate Genes From Exome Analysis. N/A = not available.

Chromosomal Location	Gene	Protein	Coding Change	Protein Change	Alteration	Clinical Associations
Chromosome 1: 66,533,383-66,748,299	*SGIP1* MIM 611540	SH3-Domain GRB2-Like (Endophilin)-Interacting Protein 1 (SGIP1)	c.1199C>A	p.Pro400His	missense	N/A
Chromosome 1: 205,042,937-205,078,499	*CNTN2* MIM 190197	Contactin 2 (CNTN2)	c.2377C>T	p.Arg793Cys	missense	epilepsy [28]
Chromosome 3: 126,006,355-126,101,561	*SLC41A3* MIM 610803	Solute Carrier Family 41, Member 3 (SLC41A3)	c.533G>A	p.Ile178Thr	missense	N/A

Chromosome 12: 6,693,792-6,700,800	*PIANP* MIM 616065	PILR-Alpha-Associated Neural Protein (PANP)	c.214G>A	p.Arg72Trp	missense	immune regulation (*m. musculus)* [29]
Chromosome 13: 96,090,839-96,839,562	*HS6ST3* MIM 609401	Heparin Sulfate 6-O-Sulfotransferase 3 (HS6ST3)	c.688C>T	p.Arg230Cys	missense	N/A
Chromosome 14: 57,200,507-57,269,008	*EXOC5* MIM 604469	SEC10-Like 1 (SEC10L1)	c.2074C>T	p.Val692Ile	missense	N/A
Chromosome 15: 80,946,289-	*MESD* MIM 607783	Mesoderm Development Gene (MESD)	c.603-606delTAA	p.Asn201Lysfs *15	frameshift	failure to form mesoderm, patterning defects, embryonic lethality (*m. musculus*) [9; 14]

80,989,878

Chromosome 17: 7,839,904-7,854,796	*KDM6B* MIM 611577	Jumonji Domain-Containing Protein 3 (JMJD3)	c.1825G>A	p.Ala609Thr	missense	Mental Retardation, Autosomal Recessive 33 [MIM 614341] (not confirmed) [30; 31]
Chromosome 19: 11,834,341-11,419,342	*RGL3* MIM 616473	Ral Guanine Nucleotide Dissociation Stimulator-Like 3 (RGL3)	c.1046G>T	p.Ala349Asp	missense	N/A
Chromosome 19: 49,766,968-49,807,113	*AP2A1* MIM 601026	Clathrin Adaptor Complex AP2, Alpha Subunit (AP2-Alpha)	c.2063G>T	p.Gly688Val	missense	N/A