The American Journal of Human Genetics

Supplemental Data

Homozygosity for Frameshift Mutations in *XYLT2* Result in a Spondylo-Ocular Syndrome with Bone Fragility, Cataracts, and Hearing Defects

Craig F. Munns, Somayyeh Fahiminiya, Nabin Poudel, Maria Cristina Munteanu, Jacek Majewski, David O. Sillence, Jordan P. Metcalf, Andrew Biggin, Francis Glorieux, François Fassier, Frank Rauch, and Myron E. Hinsdale

Supplemental Data and Methods

Table S1				
Comparison between patients Ind. 1, Ind. 2 and Spondylo-Ocular syndrome (Schmidt *et al.,* 2001)[1]				
	Ind. 1	Ind. 2	Ind. 3	SOS
Normal height	+	+	+	+
Vertebral compression fractures	+	+	+	+
Long-bone fractures	+	+	+	-
Flat feet	+	+	+	+
Cataract	+	+	+	+
Retinal detachment	+	-	+	+
Heart defect	+	+	n.r.	+/-
Hearing loss	+	+	+	-
Ureter dilatation	+	+	n.r.	n.r.

n.r., not reported. Ind.1, individual 1; Ind. 2, individual 2

Table S2			
RT-PCR primers and accession numbers for gene targets			
Gene	Forward Primer	Reverse primer	NCBI Accession number
XYLT1	ACCGAGATATGAATTTCTTGAAGTCA	AGGCCCTGCTTCCGAATG	NM_022166
XYLT2	GGGTGAGACCCGCTTCCT	GCATCATCTTTCCTGAGAGGTAGTT	NM_022167
GALT7	TGACAAGACCGCCACACC	GTCCTGAGCCTGAGCAATATG	NM_007255
GALT6	CCTAGGTCAGGCCGTTGAGTT	GCGGTCAGTCCTGGATTCA	NM_080605
GAT3A	GCCAACTGCACTCGGGTACT	CCTGCTTCATCTTGGGCTTCT	NM_012200

Supplemental Table S2. Many primer sets are available from Origene. Comparative gene expression was determined using $2^{-\Delta\Delta Ct}$ method [2].

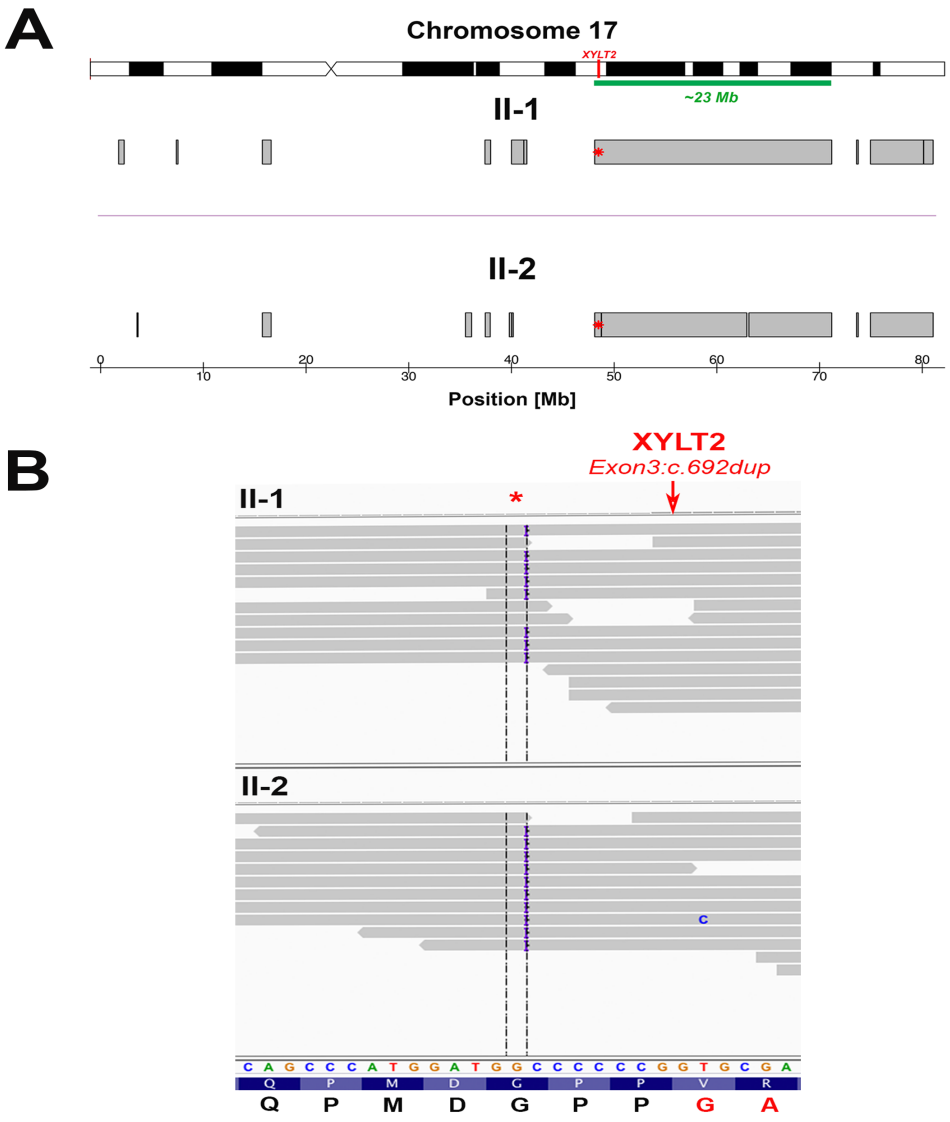

Figure S1: A) The gray bars show the location of the region of homozygosity on chromosome 17 for individuals 1 (II-1) and 2 (II-2). The ~23Mb shared region of homozygosity surrounding *XYLT2* is indicated with the green bar. *XYLT2* is located on chromosome 17 and in both individuals harbour a homozygous mutation at genomic position 4,843,186 (highlighted by a red asterisk). B) The IGV snapshot shows the position of the homozygous single-nucleotide duplication in *XYLT2* (NM_022167: c.692dup)(red arrow) that is shared by the two patients. The gray bars are 100 bp paired-end reads that mapped to the human reference genome (hg19) using BWA aligner. The insertion introduces a stop codon 54 residues downstream (p.Val232Glyfs*54) as shown in supplemental figure S2 below. In the lower part of figure, the reference DNA sequence, the reference protein sequence (two shades of blue) and protein predicted from variant coding sequence are illustrated where the variant sequence is in red.

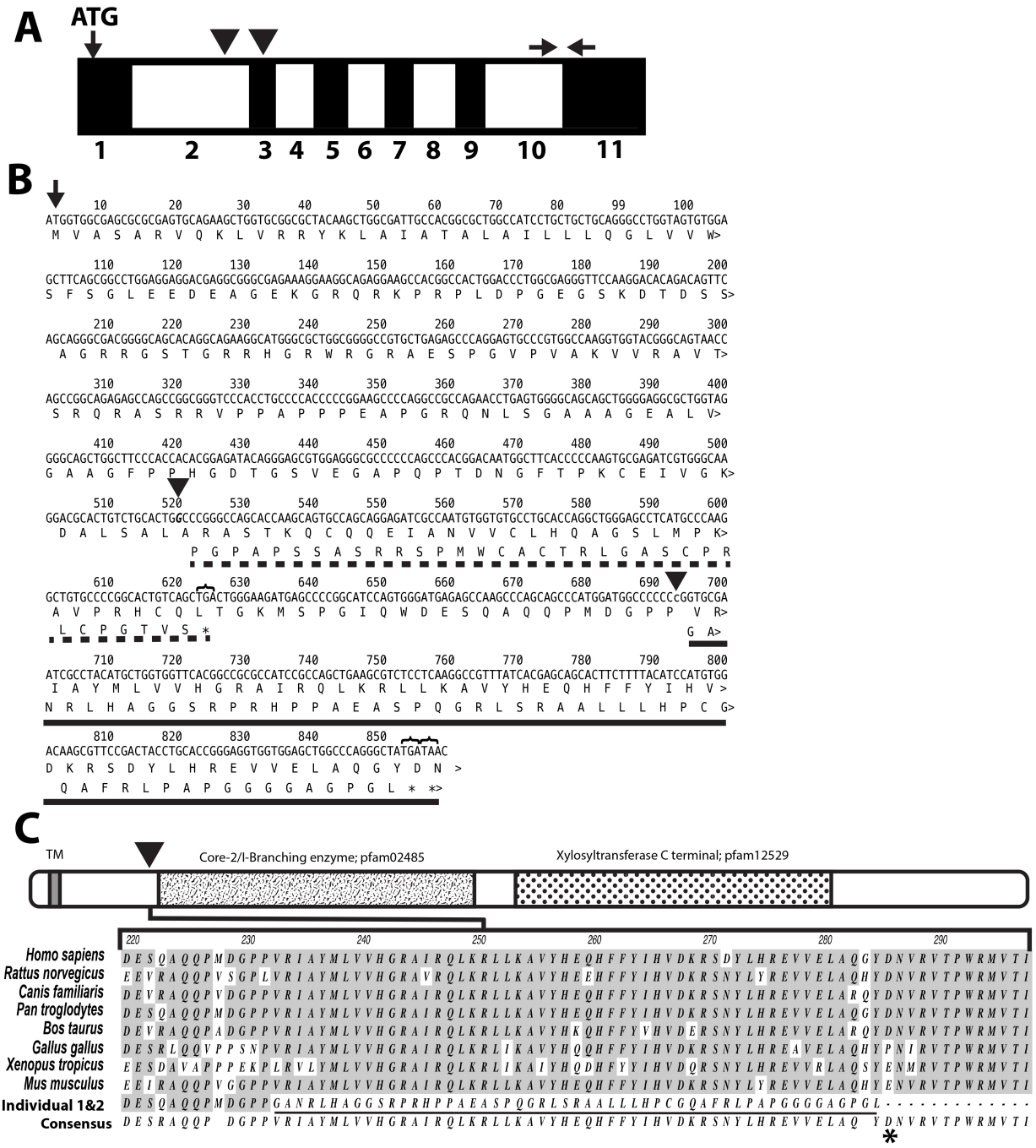

Figure S2. A) Alternating black and white boxes with indicated numbers indicate exons within the cDNA with exon 1 being black. Translational initiation codon ATG is indicated. First black arrowhead indicates site of individual 3 mutation. Second arrowhead is shared mutation in individuals I and 2. Black horizontal arrows are real-time PCR primers using for expression analyses B) Normal cDNA sequence with translation. Translational initiation codon ATG indicated by the black arrow. First black arrowhead indicates guanidine deletion with bolded base in individual 3. Top amino acid sequence is wildtype. Dotted line underlines amino acid sequence of individual 3 with new amino acids due to the frame shift and asterisk/bracket is premature translational termination. Second arrowhead is cytosine duplication in individuals 1 and 2. Black bar indicates start of aberrant protein sequence ending in two stop codons indicated by asterisk/brackets. C) Diagram shows known domain structure of XylT2. Transmembrane domain, TM, is shown in N-terminal region. Arrowhead shows location of insertion and frame shift and dotted domain, pfam 12529, shows predicted catalytic domain. Underlined amino acids are novel non-conserved 53 residues due to frame shift mutation and asterisk is premature translational termination in individuals 1 and 2.

Table S3	Genomic Position hg19				
	chr17:48431826	chr17:61902289	chr17:79225239	chr17:80136979	chrX:19725093
Consequence	frameshift insertion	Missense variant	Missense variant	Missense variant	Missense variant
Reference allele	GCCCCCC	C	G	C	C
Alternative allele	GCCCCCCC	A	A	T	T
Individual-1	Hom	Hom	Hom	Hom	Hom
Individual-2	Hom	Hom	Hom	Hom	Hom
BioType of transcript	Nonsense mediated_decay	Protein coding	Protein coding	Protein coding	Protein coding
Locus Accession Number	NM_022167	NM_017647	NM_138570	NM_198082	NM_001024666
cDNA Change	c.692_693insC	c.712G>T	c.2119C>T	c.1298G>A	c.185G>A
Protein Change	p.P231fs	p.A238S	p.L707F	p.R433H	p.R62Q
Gene name	*XYLT2*	*FTSJ3*	*SLC38A10*	*CCDC57*	*SH3KBP1*
#prev seen in-house database	0	0	1	3	0
dbSNP: rsID	-	0	rs200011449	rs200450321	-
MAF from 1000 genomes	0	0	0	0	0
EVS MAF	0	0	0.0012	0.001107	0
ExAC MAF	0	0	0.001504	0.0007	0.000008
SIFT score		0.02	0.13	0	0.03
Polyphen2 score		0.928	0	0.715	0.105
Mutation Taster prediction		Disease Causing	Polymorphism	Disease causing	Disease causing

Table S3. Frameshift and missense variants of candidate loci. Whole exome sequencing was performed on 3 µg of genomic DNA of each individual at Genome Quebec Innovation Center, Montreal, Canada. Briefly, exonic sequences were captured using the SureSelect Human Exome Kit V.4 (Agilent Technologies, Inc., Santa Clara, CA). Enriched libraries were then sequenced on Illumina Hiseq 2000 sequencer, generating 100 base pair paired-end reads for each sample. The bioinformatic analysis of exome sequencing data was carried out as previously described [3; 4]. In brief, mapping reads against the human reference genome (hg19), local realignment around small insertions or deletions (indels), depth of coverage calculation and the reads duplication removal were performed using the BWA (v. 0.5.9) [5], Genome Analysis Toolkit [6] and Picard tools (http://picard.sourceforge.net), respectively. The mean read depth for the consensus-coding sequence was 99X (II-1) and 122X (II-2), and 95% of bases were covered by ≥10 reads. The genetic variations (SNVs and indels) were detected using Samtools (v. 0.1.17) [7] and mpileup and were annotated with an in-house annotation pipeline that uses ANNOVAR and custom scripts. The potential damaging effect of variants was predicted using SIFT [18], PolyPhen-2 [19], MutationTaster [20]. To differentiate novel variants from common polymorphisms and sequencing artifacts, exonic (frameshift indels, nonsense, missense) and canonical splice site variants with minor allele frequency of more than 1% in public databases (1000Genomes or in the Exome Variant Server [EVS]) or seen in >10 individuals in our in-house exome database were filtered out. Candidate genes having shared homozygous mutations between the patients were selected and were manually examined using the Integrative Genomics Viewer (IGV) [8]. All loci from index patients and parents were confirmed to be in the *XYLT2* gene by Sanger sequencing of *XYLT2* exon 3 amplification by polymerase chain reaction followed by direct sequencing using an Applied Biosystems 3730xl sequencer (Applied Biosystems, Foster City, CA, USA). Sequence traces were aligned with GenBank reference sequence NM_022167.2.

References:

1. Schmidt, H., Rudolph, G., Hergersberg, M., Schneider, K., Moradi, S., and Meitinger, T. (2001). Retinal detachment and cataract, facial dysmorphism, generalized osteoporosis, immobile spine and platyspondyly in a consanguinous kindred--a possible new syndrome. Clin Genet 59, 99-105.
2. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402-408.
3. Fahiminiya, S., Majewski, J., Mort, J., Moffatt, P., Glorieux, F.H., and Rauch, F. (2013). Mutations in WNT1 are a cause of osteogenesis imperfecta. J Med Genet.
4. Fahiminiya, S., Al-Jallad, H., Majewski, J., Palomo, T., Moffatt, P., Roschger, P., Klaushofer, K., Glorieux, F.H., and Rauch, F. (2014). A Polyadenylation Site Variant Causes Transcript-Specific BMP1 Deficiency and Frequent Fractures in Children. Hum Mol Genet.
5. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.
6. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303.
7. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.
8. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol 29, 24-26.