Identification of variants by genome sequencing

For genome sequencing, library preparation was conducted using the TruSeq PCR-free DNA Library Prep Kit with sequencing on the Illumina NovaSeq 6000 following the manufacturer’s instructions. The Care4Rare in-house pipeline was used to align the raw sequencing reads, and call and annotate variants (https://github.com/ccmbioinfo/crg2). Briefly, reads were aligned to the GRCh37 reference genome with decoy sequence using BWA [e1], and duplicates marked with Picard (https://broadinstitute.github.io/picard/). Single nucleotide variants and indels were called and jointly genotyped using GATK HaplotypeCaller v4 [e2]. Single nucleotide variants and indels were annotated using VEP [e3] and vcfanno [e4]. Structural variants (SVs) were called using Manta [e5], Wham [e6], and Smoove (https://github.com/brentp/smoove). SVs were merged using metaSV [e7] and annotated using snpeff [e8]. CNVs were called by CNVnator and ERDS using the TCAG’s CNV pipeline [e9]. Short tandem repeats in known disease loci were genotyped using ExpansionHunter [e10], and short tandem repeats in non-reference loci were genotyped using ExpansionHunter Denovo [e11].

Variants were disregarded if they were present at >1% in gnomAD or seen in more than 5 samples from our in-house database (~2000 exomes previously sequenced at the McGill University and Genome Quebec Innovation Centre). PCR and Sanger sequencing were used to validate the variants identified by genome-wide sequencing.

Genomic imprinted expression

RT-PCR was conducted on cells from fibroblasts and lymphoblasts on unrelated patients for whom exome data was available and showed a SNP in SOX8 with corresponding parental data for segregation. Total RNA was extracted from affected and control fibroblast cell lines with the RNeasy Mini Kit (QIAGEN) and reverse transcribed into cDNA using the iScript cDNA Synthesis Kit (Bio-Rad Laboratories) as per the manufacturer’s instructions. RT-qPCR experiments were conducted from cDNA using the iQ SYBR Green Supermix (Bio-Rad Laboratories) and using the following conditions: 95°C for 10 s, 55°C for 20 s, 72°C for 30 s, repeated 35 times. Negative controls without reverse transcriptase
were prepared in parallel and included for all RT-qPCR experiments. cDNA fragments were cloned into a TOPO-TA cloning vector (Thermofisher), transformed and colonies chosen for colony PCR reactions and subsequent Sanger sequencing. Data was obtained from a minimum of 20 clones per sample.

**Real-time PCR analysis**

Fibroblasts cells from the affected individual were derived and established from a skin biopsy by The Centre for Applied Genomics (Toronto, Canada). Total RNA was extracted from affected and control fibroblast cell lines with the RNeasy Mini Kit (QIAGEN) and reverse transcribed into cDNA using the iScript cDNA Synthesis Kit (Bio-Rad Laboratories) as per the manufacturer’s instructions. RT-qPCR experiments were conducted from cDNA using the iQ SYBR Green Supermix (Bio-Rad Laboratories) and using the following conditions: 95°C for 10 s, 55°C for 20 s, 72°C for 30 s, repeated 35 times. Negative controls without reverse transcriptase were prepared in parallel and included for all RT-qPCR experiments. Quantification of gene expression was performed using CFX Manager Software (Bio-Rad Laboratories) and all data normalized against GAPDH as an internal control. Primer sequences are available upon request.

**Minigene assay**

DNA from fibroblast cells of the affected individual was extracted using the QIAamp DNA blood kit (QIAGEN) and was amplified using primers surrounding exons 1 and 2 of SOX8 (primers were located outside of exons in intergenic and intron regions; sequences available upon request) and high fidelity Taq DNA polymerase. PCR products were run in an agarose gel and purified using a QIAquick gel extraction kit (QIAGEN) before being used for pGEM-T Easy Vector cloning (Promega). Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (QIAGEN), which was then used for Sanger sequencing to determine the inserted DNA sequence. Plasmids containing the c.422+5C>G variant from the affected individual, as well as the wild type sequence as control, were individually digested and gel extracted to isolate the gDNA sequences, followed by ligation into the pcDNA3.1 vector and transformation into JM109 competent cells. Plasmid DNA for the control and the affected c.422+5C>G variant were individually isolated as above and transfected into HeLa cells using TurboFect transfection reagent (ThermoFisher...
RNA extraction from transfected cells and cDNA synthesis were performed as above, followed by PCR amplification of cDNA, gel purification and Sanger sequencing to determine splicing of the plasmid insert.

**Western blot analysis**

Fibroblast cells from the affected individual and controls were lysed using radioimmunoprecipitation assay (RIPA) buffer supplemented with 10 mg/ml each of aprotinin, phenylmethanesulfonyl fluoride and leupeptin (Sigma-Aldrich) and were incubated on ice for 20 min. Protein samples were centrifuged at 13,000xg for 15 min and the supernatant was collected. Protein determination was achieved using a Bradford assay (Bio-Rad Laboratories). Protein samples were resolved by SDS-PAGE, transferred onto nitrocellulose membranes and processed for western blotting. Briefly, membranes were incubated in 5% non-fat milk diluted in Tris-buffered saline (TBS) containing 0.5% Tween-20, for at least 1 h at room temperature, followed by an overnight incubation with the appropriate primary antibody at 4°C. Membranes were washed three times with TBS containing 0.5% Tween-20 and incubated with the secondary antibody for 1 h at room temperature. Following three washes, membranes were visualized after incubation with the Clarity Western ECL Substrate (Bio-Rad Laboratories). The primary antibodies used for this study included SOX8 – N-terminal region (Abcam; ab104245), SOX8 – C-terminal region (Abgent; AI16236), SOX9 (Abcam; ab185230), SOX10 (Abcam; ab155279), non-phospho (Active) β-Catenin (Ser33/37/Thr41) (Cell Signaling; 8814), β-catenin (Abcam; ab16051), axin (Santa Cruz Biotechnology, Inc.; sc-14029) and β-tubulin (Abcam, ab6046).

**Co-immunoprecipitation**

Protein samples were extracted as described above. For co-immunoprecipitation experiments, 200 µg of protein was measured and diluted in phosphate-buffered saline (PBS) containing 0.02% Tween-20 to a final volume of 500 µL. Protein lysates were incubated with the SOX8 N-terminal antibody (Abcam; ab104245) overnight at 4°C on a rotator. Normal rabbit IgG (Santa Cruz Biotechnology, Inc.) was used as a negative control. Samples were then incubated with Dynabeads protein G (Invitrogen) for
1 h at 4°C, and then were washed three times with PBS containing 0.04% Tween-20 and resolved by SDS-PAGE as described earlier.

**Chromatin immunoprecipitation**

Chromatin immunoprecipitation was performed with the EZ-ChIP kit (Upstate) according to the manufacturer’s instructions. Briefly, affected and control fibroblast cells were crosslinked in 1% formaldehyde, lysed in sodium dodecyl sulfate (SDS) buffer and sonicated. Sheared, crosslinked chromatin were incubated with a SOX8 N-terminal antibody (Abcam; ab104245) overnight at 4°C. Normal rabbit IgG (Santa Cruz Biotechnology, Inc.) was used as a negative control. Input samples were saved, which represented 1/25 of the total chromatin input for each IP. Amplification conditions for subsequent qPCRs with ChIP DNA were 95°C for 5 min, followed by 30 cycles of 95°C for 10 s, 55.5°C for 20 s and 72°C for 30 s. Primer sequences are available upon request. Quantification was conducted as previously described [e12].

**eReferences**


