SUPPLEMENTAL MATERIAL

Functional variants of *POC5* identified in patients with idiopathic scoliosis

Shunmoogum A. Patten, Patricia Margaritte-Jeannin, Jean-Claude Bernard, Eudeline Alix, Audrey Labalme, Alicia Besson, Simon L. Girard, Khaled Fendri, Nicolas Fraisse, Bernard Biot, Coline Poizat, Amandine Campan-Fournier, Kariman Abelin-Genevois, Vincent Cunin, Charlotte Zaouter, Meijiang Liao, Raphaelle Lamy, Gaetan Lesca, Rita Menassa, Charles Marcaillou, Melanie Letexier, Damien Sanlaville, Jerome Berard, Guy A Rouleau, Françoise Clerget-Darpoux, Pierre Drapeau, Florina Moldovan*, Patrick Edery*

* F.M. and P.E. jointly directed this work.

Supplemental Results

High-Throughput Sequencing of *POC5* in IS Families and Cases

Three additional *POC5* SNVs, including 2 rare missense SNVs and a novel 5'UTR SNV were identified (Supplemental Table 6). The pathogenicity and possible role of these mutations in IS will be determined by further studies.

Supplemental Methods

Patients

The 41 multiplex idiopathic scoliosis (IS) families (F1-F41) include 135 affected individuals, 45 individuals of uncertain status and 150 unaffected individuals. 8/41 multiplex families included 5-11 affected individuals and in the remaining 33/41 multiplex families, 2-4 individuals were affected (Supplemental Figure 11 A-C). Participants were seen at the Massues Center and at the Hôpital Femme Mère Enfant, Lyon, France. Recruitment for this study started on March 1, 2000 and ended on July 31, 2012. The collection of 41 multiplex IS families was established between March 1, 2000 and January 2, 2012. Multiplex IS families in which disease transmission appeared to be consistent with an autosomal dominant trait were selected. Families where IS was diagnosed in both parental branches were excluded from the study. An additional collection of 150 IS cases for whom familial information was not initially recorded was established between January 2, 2012 and July 31, 2012. The control population was of similar ancestry (French, French Canadian or European) and consisted of 1268 individuals. This control population was not screened for the presence of IS.

Phenotypic Characterization of Idiopathic Scoliosis

IS was diagnosed by combining clinical examination of the spine, including the forward bending test (Adams test), with measurement of Cobb's angle on X-ray images. Both a Cobb's angle greater than or equal to 15° (to minimize the risk of phenocopies) and vertebral rotation were required for positive diagnosis of IS. Patients presenting spine curvature but no rotational component were classed as "unknown status", as were those with any other associated anomaly of the spine or, more generally, with atypical IS. Individuals with normal clinical examination and a strictly normal spine on radiograph were considered to be "unaffected". All individuals were clinically examined by at least one clinical geneticist and one orthopedist. Each radiograph was carefully, and often repeatedly, checked by the geneticist (PE) and an orthopedist (BB, JCB, NF, KAG, VC or JB). Medical records and spine radiographs from IS multiplex families and cases harboring any of the c.G1336A (p.A446T), c.G1363C (p.A455P) or c.C1286T (pA429V) *POC5* functional SNVs (where "functional SNVs" indicates SNVs which, when over-expressed in zebrafish, produce scoliosis-like traits) are presented in Supplemental Fig. 3A-E and Supplemental Table 5.

Sample Collection and Legal Issues

EDTA blood samples were obtained from each participant, i.e., 150 IS cases and affected members of families F1-F41 and their first-degree relatives (either affected or unaffected), as often as possible. Similar blood samples were subsequently obtained also from relatives of IS cases harboring the c.G1336A (p.A446T) *POC5* SNV, i.e., one parent and daughter of case C39 and both parents of case C83. Blood samples were not available from relatives of case C58. DNA was extracted from peripheral blood using a QIAmp DNA Blood Midi Kit (Qiagen), according to the manufacturer's instructions. Lithium-heparinate blood samples were collected from at least one proband in each multiplex IS family. Lymphoblast cell lines were established and standard

blood karyotyping was performed. The protocol for this study was approved by the local ethics committee and was sponsored by the Hospices Civils de Lyon, France.

Genetic Refinement of the 5q13.3 Idiopathic Scoliosis Critical Region

A minimum common haplotype, shared by all affected members from family F2 was determined. Towards this aim, DNA samples (500 ng) from all available individuals belonging to family F2 were hybridized on 700k Illumina HumanOmniExpress SNP arrays (Illumina) as described in the manufacturer's protocol. Genotypes were analyzed and the locations of the recombination events were refined using Merlin° software (1).

Whole-Exome Sequencing, Read Mapping and Variant Calling

After unsuccessful direct sequencing of a number of candidate genes from the refined 5q13.3 and 3q12.3 IS critical intervals, whole-exome sequencing was performed in 3 patients from family F2, as previously described (2, 3). Exome capture and high-throughput sequencing were performed at McGill University and Génome Québec Innovation Center (Montréal, Canada). Exomes were enriched using an Agilent SureSelect all-exome kit (V4 optimized for Illumina HiSEQ sequencing), with 2 µg of subjects' genomic DNA. This enrichment is designed to cover approximately 50 Mb of genomic sequences, mainly protein coding sequences. Exon-enriched DNA libraries were sequenced (paired-end, 2x100 bp) using an Illumina HiSEQ 2000 platform in accordance with the manufacturer’s instructions.

The total number of reads was on average 6,845,663,492 per patient. The Burrows-Wheeler Aligner (BWA, (4) was used as the main aligner for mapping against the human genome (hg19), which was indexed using the bwtsw algorithm included with BWA. Alignment was performed using a maximum mismatch penalty of three. All other parameters from BWA were left at their

default values. The alignment was generated in paired-end mode, and SAMTOOLS (5) was used to store the alignment. Duplicate reads were marked using Picard (http://picard.sourceforge.net) and were excluded. Average coverage of consensus coding sequence (CCDS) was calculated for each sample using GATK (6). After duplicate read removal, mean coverage was 134X, with a range of 130X to 141X for the different DNA samples (Supplemental Table 1, Supplemental Figure 2). The average transition to transversion ratio was 2.81 after applying the PASS only filter value. The "Best Practice Variant Detection with GATK v2" was used to generate SNP and call indels (6, 7). SNP and indels in samples sequenced on the Illumina HiSEQ system were called using a set of 88 samples sequenced in the same conditions. Variant frequencies were established by comparison to the 1000 Genomes database. Sequences were annotated using ANNOVAR (8) and the RefSeq and dbSNP132 databases. To identify putative IS mutations, a filter was applied to retain novel variants or variants with very low-frequency missense alleles (minor allele frequency (MAF) < 5%) in the databases (1000 Genomes Project, dbSNP and our in-house control exomes (n=1165)).

Sanger Sequencing and Statistical Analyses

Sanger sequencing

A DNA fragment containing the c.G1336A (p.A446T) *POC5* rare SNV identified by whole-exome sequencing was PCR-amplified for classic Sanger sequencing in all individuals from IS family F2, one proband from each of the remaining 40 IS families (F1 and F3-F41) and 150 IS cases (C1-C150). This amplicon was then sequenced in all available individuals from IS families F19, F31, F35 and F41, in which the c.G1336A (p.A446T) or the c.G1363C (p.A455P) *POC5* SNVs were detected. The PCR primers used were as follows: Forward 5'CTTTTCATAAGGTGGGACCT3'; Reverse 5'TCCGATGCCCTTACCAG3'. PCR was

performed on a FlexCycler (AnalytikJena). PCR products were purified using commonly applied methods before analysis of amplicons on an ABI 3730xl DNA Analyzer (Applied Biosystems).

Statistical analysis

The allelic frequencies of the c.G1336A (p.A446T) and c.C1286T (p.A429V) *POC5* SNVs were each compared in 191 IS cases (41 familial cases and 150 isolated cases from this study) to the control population (1268 individuals). A one-tailed Fischer's exact test was performed to test the hypothesis that the *POC5* variant are more frequent in the IS population than in controls.

Founder Effect Studies

DNA samples from all available individuals of IS families F2, F19, F35 and F41, case C39, her husband and daughter, case C83 and her parents, and case C58 (no samples were available from relatives of case C58) were genotyped using 700k Illumina HumanOmniExpress SNP arrays (Illumina), as described above. Haplotypes were reconstructed when possible and the length of the shared haplotype was determined (Supplemental Figure 4). Haplotypes were reconstructed when possible, using Merlin software and figures were drawn using Haplopainter° software (9).

High-Throughput Sequencing of *POC5* in IS Families and Cases

Whole exonic, flanking intronic and regulatory *POC5* sequences were studied using a Fluidigm Access Array° device (IntegraGen). A proband was analyzed from each of 40 multiplex IS families (F1/ F3-F41) and 150 IS cases (C1-150). Primer pairs were designed using an in-house pipeline based on Primer3° (Supplemental Table 7). 5.5 Kb of *POC5* sequences were covered by 42 overlapping amplicons, average length 277 bp. Each sample was quantified with Picogreen°, and 50 ng of DNA were used to prepare the library, according to the Fluidigm recommendations. Universal tags, Rd1 and Rd2, were added to the 5' end of the forward and reverse primers,

respectively, for the first round of PCR, which was performed on the Access Array°. Illumina adapters, P5 and P7, as well as barcodes were added to the pooled PCR products for the second round of PCR, performed in microplates. Paired-end sequencing was performed on the Illumina° MiSeq system after quality control (Fragment Analyser, AATI°) and quantification.

Zebrafish Maintenance

Zebrafish (*Danio rerio;* wild type AB strain) embryos were raised at 28.5 °C, collected and staged using standard methods. All procedures described here were carried out in accordance with the guidelines set out by the Canadian Council for Animal Care (CCAC), the CHU Sainte-Justine Research Center, and the Comité de Déontologie de l'Expérimentation sur les Animaux (CDEA), which is the local animal care committee at the University of Montreal. This study was approved by the ethics committee for CHU Sainte-Justine Research Center, University of Montreal (ZF-09-60/Category B). Fish were anaesthetized in 0.02% tricaine (MS-222; Sigma Chemical, St. Louis, MO) in phosphate-buffered saline (PBS) prior to all procedures.

***Poc5* Knockdown in Zebrafish**

poc5 expression was knocked down using a morpholino antisense oligomer targeting the ATG of the zebrafish *poc5* ortholog (XM_685988). The translation-blocking morpholino, 5′-GTTCATTTGAAGGTCTATTACATCT-3′, was supplied by Gene Tools (Philomath, OR). The morpholino was injected into single-cell stage zebrafish embryos at doses of 2 ng/embryo, 4 ng/embryo and 6 ng/embryo.

A splice-blocking morpholino (5′-ACCGCAAGTGCAATACAAACCTTAA-3′) was also used to knocdown poc5 expression in zebrafish. The splice blocking morpholino was designed to bind *poc5* mRNA at the junction across the 3′ end of exon 5 and the 5′ end of intron 5. To test for loss of proper *poc5* mRNA processing, PCR primers (forward- 5′-CATGTCAGCCAGGTCTGTGT -

3′, reverse- 5′-TCCATCTCAGCATTCACAGC-3′) were designed to bind cDNA at sites corresponding to the 5′ end of exon 5 and the 3′ end of exon 6 of the poc5 mRNA. Amplified cDNA was visualized using gel electrophoresis.

Expression of Wild-Type and Mutated Human *POC5* SNVs in Zebrafish

In vitro mRNA synthesis and microinjection into embryos

Wild-type and mutated versions of human *POC5* were produced from a myc-tagged ORF clone of human *POC5* (Origene) and injected into zebrafish embryos. Site-directed mutagenesis was performed on this vector using a QuikChange® XL Site-Directed Mutagenesis Kit (Agilent). The sequences of primers used for this assay are listed in Supplemental Table 8. Messenger RNAs were obtained from linearized constructs, using the T7 RNA polymerase and the mMESSAGE mMACHINE kit (Ambion). Transcription products were extracted by phenol:chloroform, precipitated in isopropanol, and diluted in nuclease-free water (Ambion) with 0.05% Fast Green vital dye (Sigma-Aldrich). mRNAs were injected into one- or two-cell stage embryos using a Picospritzer III pressure ejector. The final injection volume was ~1.5 nl, at a concentrations of 25 ng/µl, 50 ng/µl, 100 ng/µl and 150 ng/µl mRNA. Injected and non-injected embryos were then incubated in appropriate media at 28.5 °C for 24 h, and assessed for viability. Morphological differences between mutant injected, wild-type injected and non-injected embryos were assessed under an Olympus SZX12 stereoscope.

Western blot

Total protein extracts (40 µg) were obtained from 3 dpf wt-*POC5* and mut-*POC5* zebrafish. Proteins were resolved on a 12% polyacrylamide gel and transferred onto a PVDF membrane. Membranes were blocked in 0.1% PBS-Tween, 5% Skim Milk for one hour followed by overnight incubation at 4 °C with primary antibodies: rabbit anti-myc (rabbit polyclonal; at a

dilution of 1:2000; Sigma; catalogue# AV38156) or mouse anti-γ tubulin (Sigma; catalogue# T6557; 1:5000) in 0.1% PBS-Tween with 5% BSA. After washing, membranes were incubated for 1 h at room temperature with secondary antibodies: donkey anti-rabbit-HRP or donkey anti-mouse HRP, as appropriate (both polyclonal antibodies from Jackson ImmunoResearch; catalogue# 715-035-151 and 711-036-152 respectively; dilution 1:10,000) in 0.1% PBS-Tween with 5% milk. Blots were revealed by ECL after a 10-second exposure.

Statistical analysis

Statistical analyses were performed and data were plotted using SigmaPlot 11.0 (Systat Software Inc., CA). A Chi-squared ($\chi 2$) test was used to analyze the statistical significance of differences in the zebrafish phenotype distributions between experimental groups.

Three-dimensional Imaging and Reconstruction of Zebrafish Bone

Juvenile zebrafish underwent a micro-CT scan (SkyScan 1072 High Resolution Desktop Micro-CT System, Microtomograph, SkyScan) for three-dimensional (3D) visualization of the skeleton after 3D imaging and subsequent reconstruction. Acquisition parameters for the scan were as follows: 35 kV, 215 μA, step rotation of 0.9°, pixel size 4-7 microns; images were reconstructed using NRecon (Version: 1.6.1.3).

Poc5 *in situ* Hybridization in Zebrafish

Total RNA was extracted from 48 hours post-fertilization zebrafish embryos using Trizol (Invitrogen). This RNA was reverse transcribed using a Reverse Transcription Kit (Qiagen, Valencia, CA). The cDNA produced was PCR-amplified using Poc5 primers, and the PCR

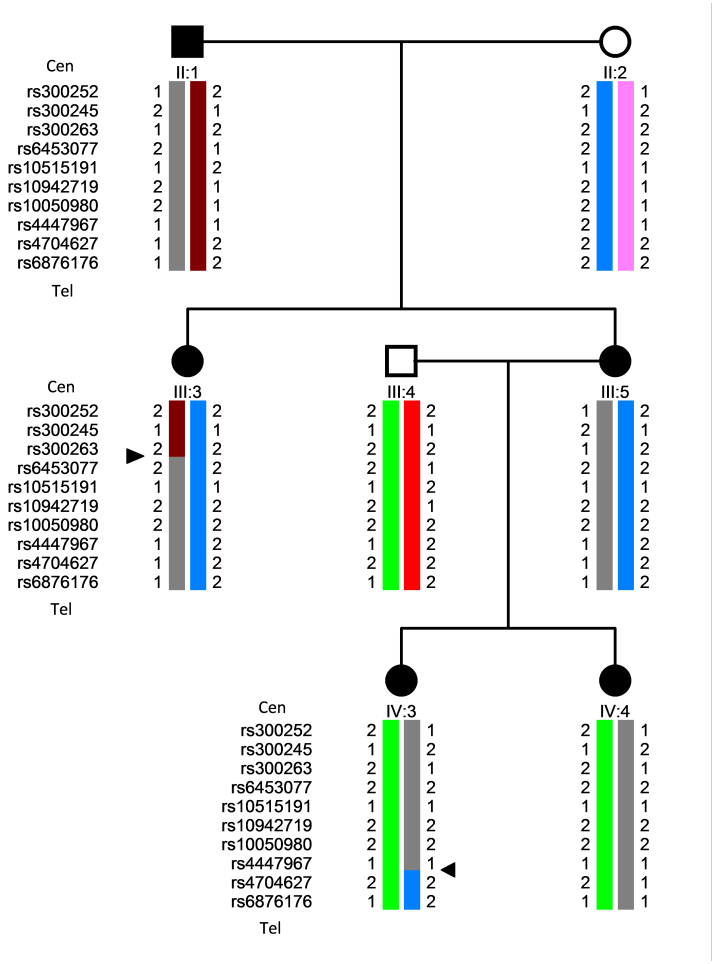
products served as templates for *in vitro* transcription to produce Poc5 RNA probes, as previously described (10, 11).

Poc5 primers were as follows: Poc5F primer: CAGATCTCTAACCAGAGGAAAGATG and T7_Poc5R primer TAATACGACTCACTATAGGGAGAGTATTGGACTCTCCATGACTATTGG (T7 promoter sequence is underlined). These primers were used to generate an antisense probe. T7_Poc5F: TAATACGACTCACTATAGGGAGACAGATCTCTAACCAGAGGAAAGATG (T7 promoter sequence is underlined) and Poc5R: GTATTGGACTCTCCATGACTATTGG were used to generate the sense (control) probe. PCR products were then transcribed *in vitro*, using T7 RNA polymerase, to produce RNA probes. RNA probes were labeled with DIG using a DIG RNA labeling kit (Roche). DIG-labeled RNA probes were precipitated in 0.2 M EDTA, 4 M LiCl, and 100% ethanol overnight at −20 °C, and suspended in DEPC-treated water. The purified probe was visualized on Agarose gel. Probes were stored at −80 °C.

Whole-mount *in situ* hybridizations were performed on staged zebrafish embryos using both sense and antisense *poc5* riboprobes. Briefly, staged embryos (15-72 hpf) were fixed overnight in 4% paraformaldehyde before dehydrating in methanol. For use, embryos were rehydrated in phosphate-buffered saline with 0.1% Tween-20 (PBSt). Embryos were permeabilized with proteinase K and hybridized with riboprobes overnight at 70 °C. The next day, embryos were prehybridized in graded solutions of 75%, 50%, and 25% 2X saline-sodium citrate (SSC) solutions, then washed in 0.2X SSC for 30 minutes at 68 °C. Embryos were placed in blocking solution for several hours, before incubating with α-DIG antibody overnight. Finally, embryos were washed again and incubated in NBT/BCIP staining solution in the dark, until staining on the embryos was sufficiently visible. Younger embryos (15-24 hpf) were automatically processed for hybridization, SSC washes and incubation with α-DIG antibody in the *in situ* hybridization

system, Flogentec (www.flogentec.com), prior to staining as previously described (12). Embryos were stored in glycerol and visualized using an Olympus Stereomicroscope.

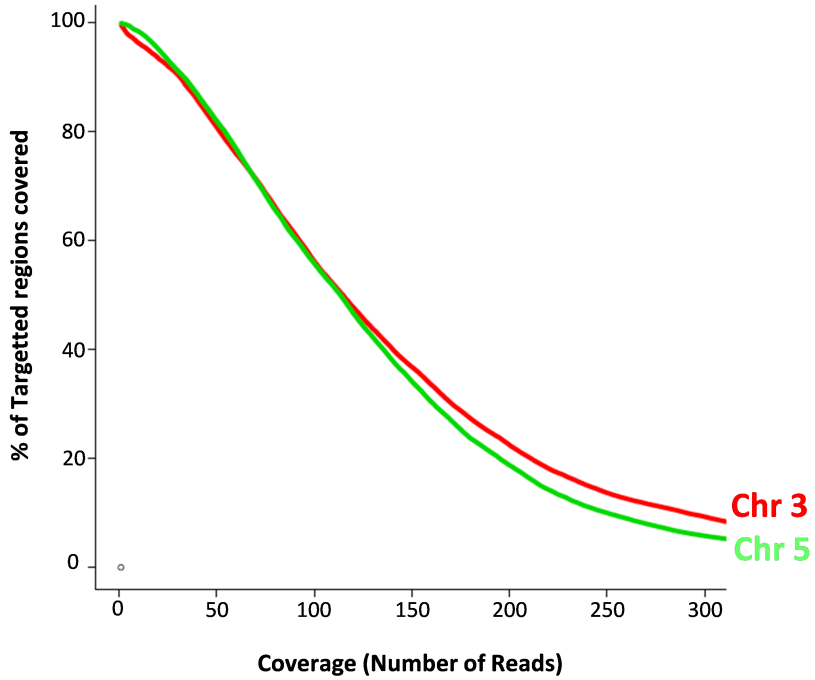
Supplemental Figure 1



Supplemental Figure 1. Genetic Refinement of the 5q13.3 Idiopathic Scoliosis Interval in Family F2

Whole-genome genotyping was performed in IS family F2 using an IlluminaOmniExpress chip, revealing a minimum IS critical interval of 5.582 Mb on chromosome 5q. Haplotypes are illustrated for some family members and with only some SNPs for clarity. The grey haplotype harbors the IS-causing gene. Arrowheads show centromeric and telomeric recombination events. The refined 5q13.3 critical IS interval is proximally bounded by rs300263 and distally bounded by rs4704627.

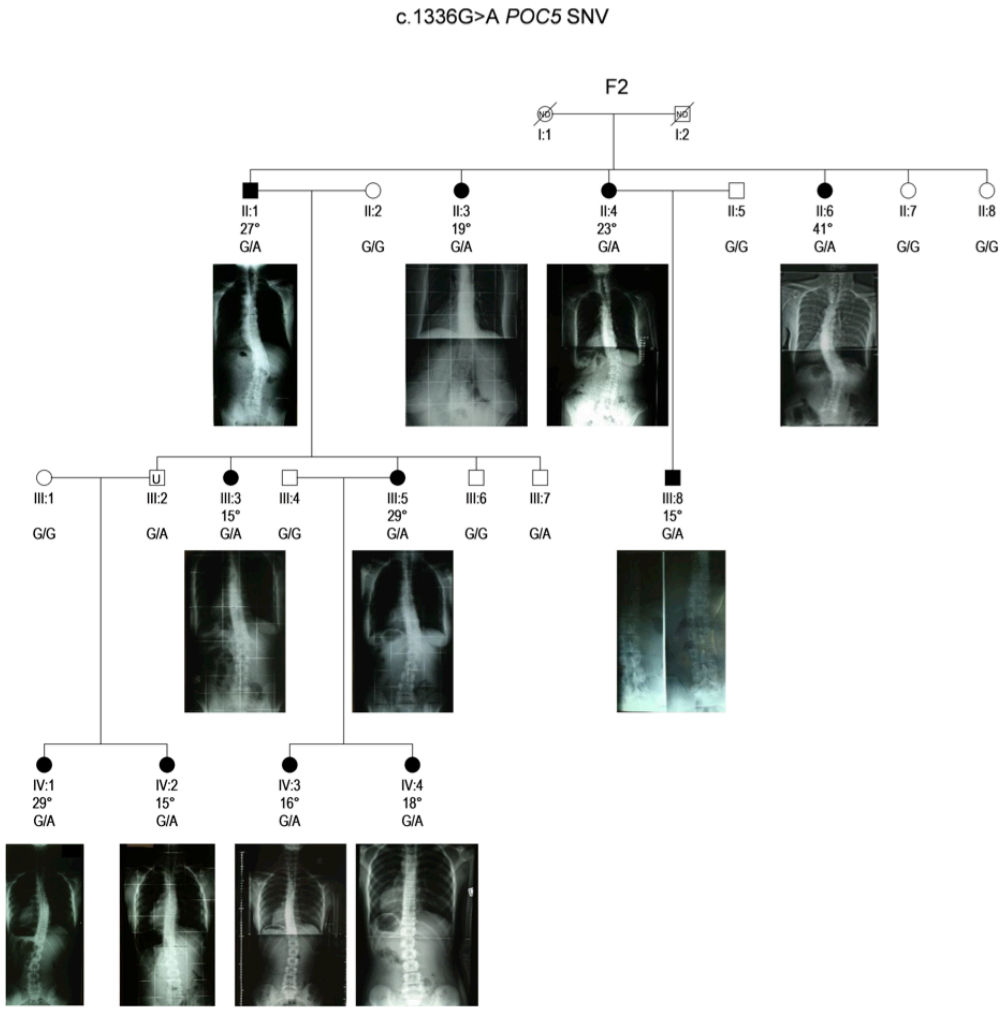
Supplemental Figure 2



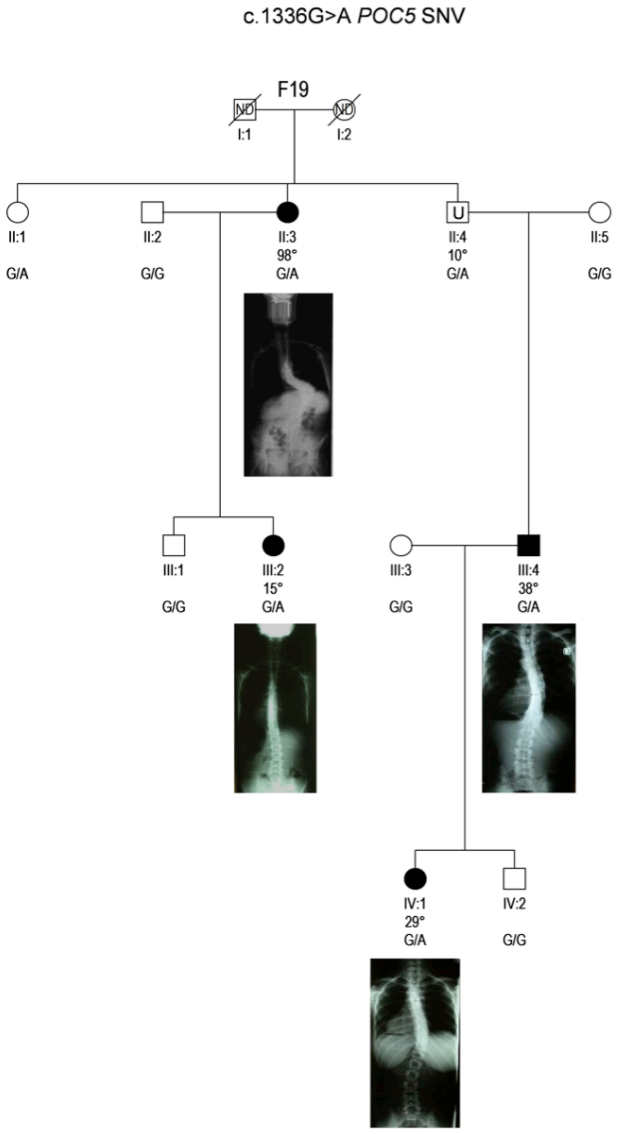
Supplemental Figure 2. Exome Capture Efficiency

Exome capture efficiency is shown for each individual sequenced in this study. The x-axis presents the coverage in total number of reads, while the y-axis shows the percentage of the total targeted region, on a per-base calculation.

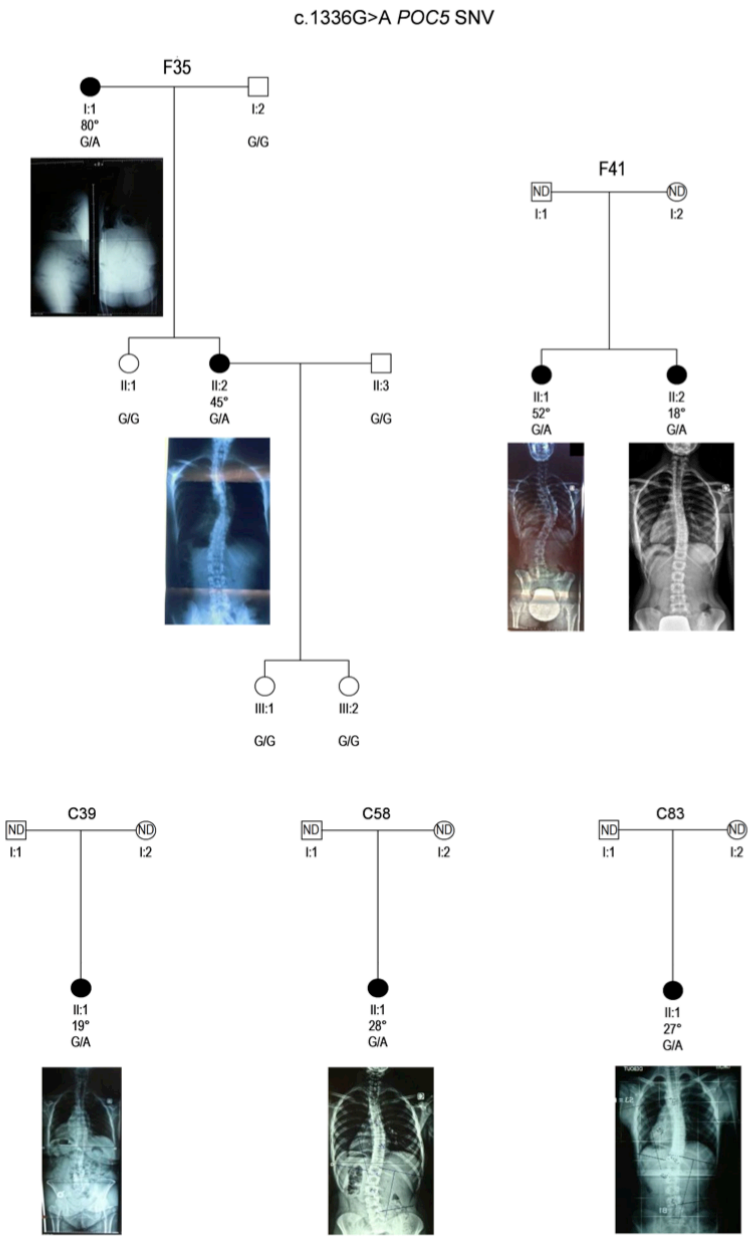
Supplemental Figure 3A



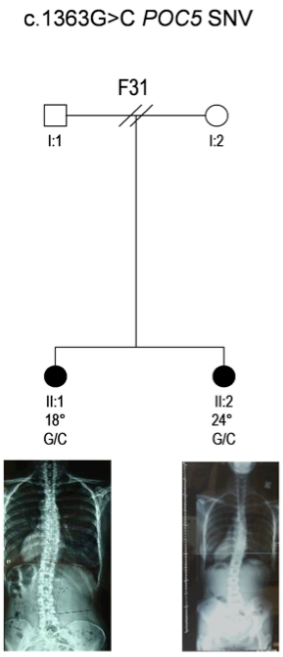
Supplemental Figure 3B



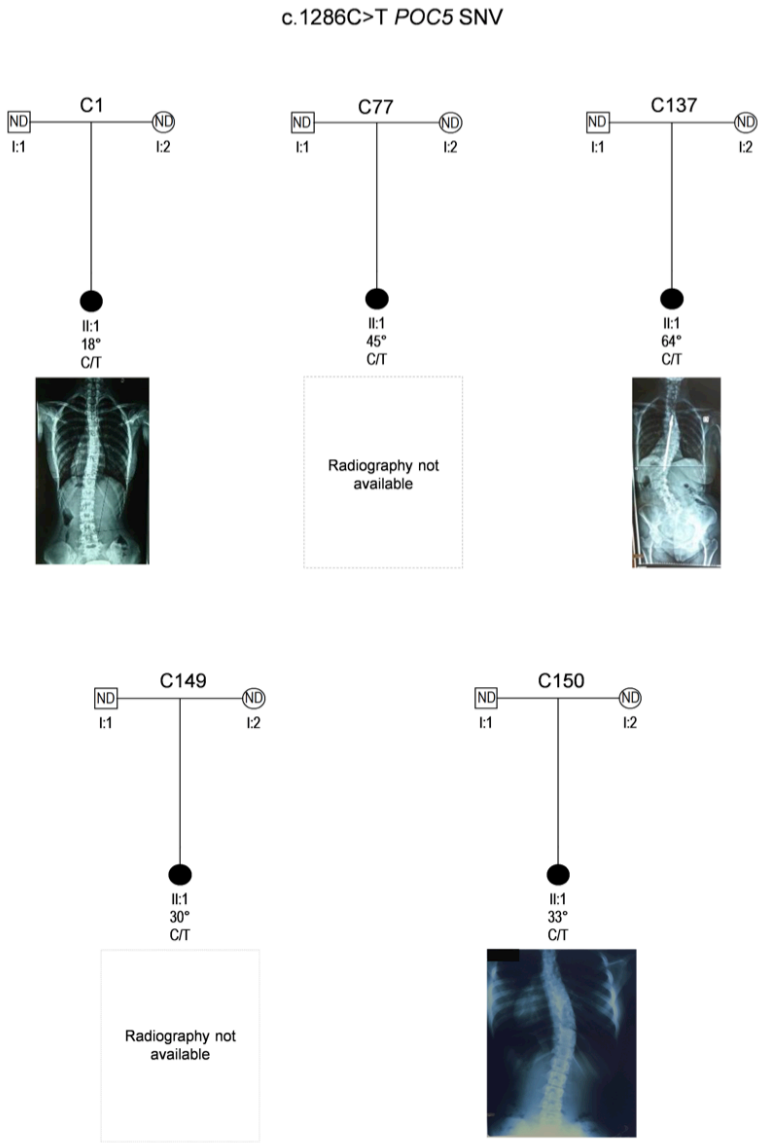
Supplemental Figure 3C



Supplemental Figure 3D

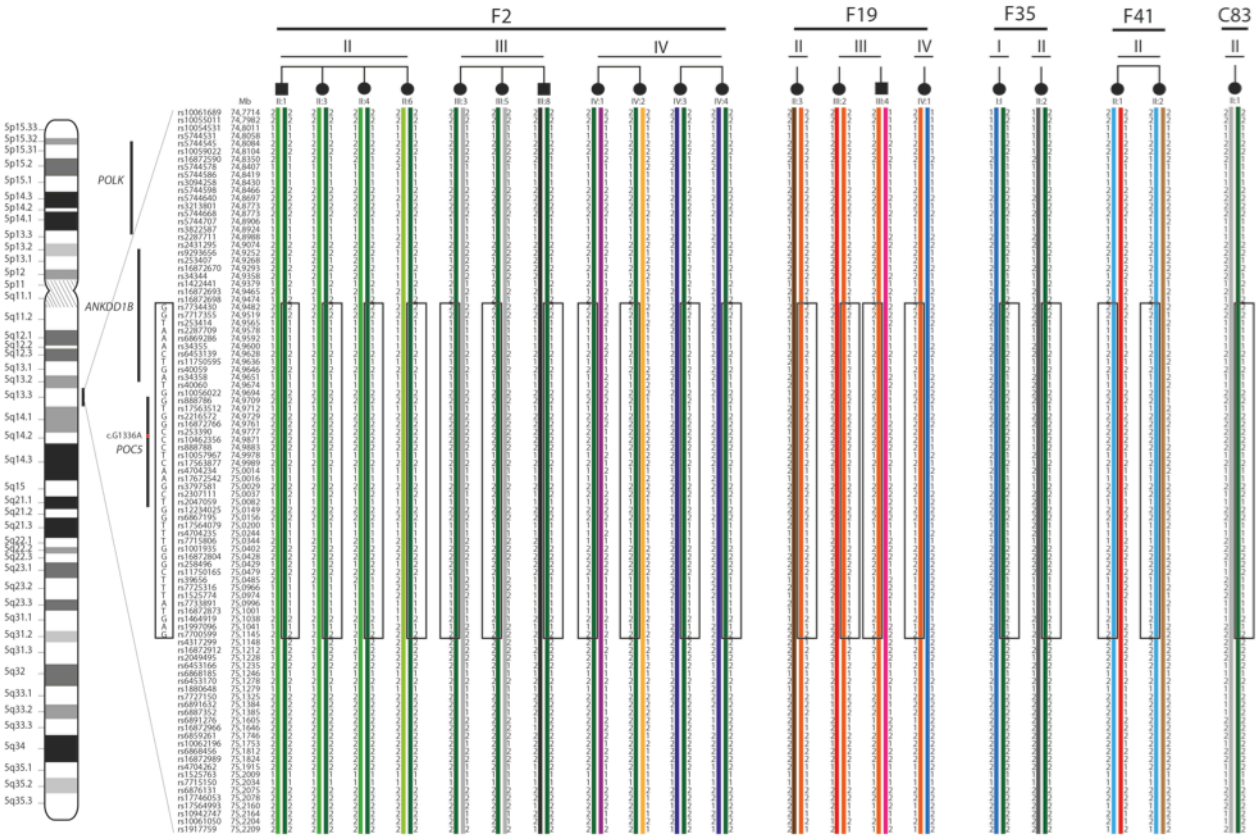


Supplemental Figure 3E



Supplemental Figures 3A-E. Pedigrees and Spine Radiographs of Idiopathic Scoliosis Families and Cases Carrying either c.G1336A, c.G1363C or c.C1286T *POC5* SNVs.

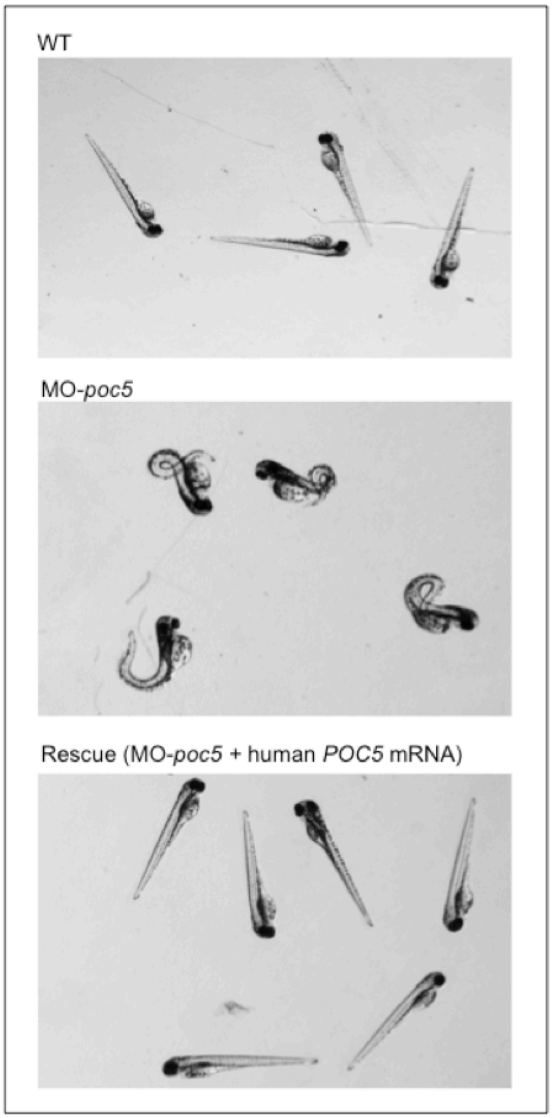
Supplemental Figure 4



Supplemental Figure 4. Haplotype Analysis of Families and Cases harboring the c.G1336A *POC5* SNV

Haplotypes at the *POC5* locus were reconstructed. The most likely haplotype was determined using Merlin° software. When different haplotypes had a similar likelihood, reconstruction was considered impossible (IS patients C39 and C58). All IS patients in whom haplotypes could be reconstructed carried the c.G1336A *POC5* SNV on the same ancestral haplotype, denoted H (boxed). Genotypes of C39 and C58 (not shown) were compatible with haplotype H.

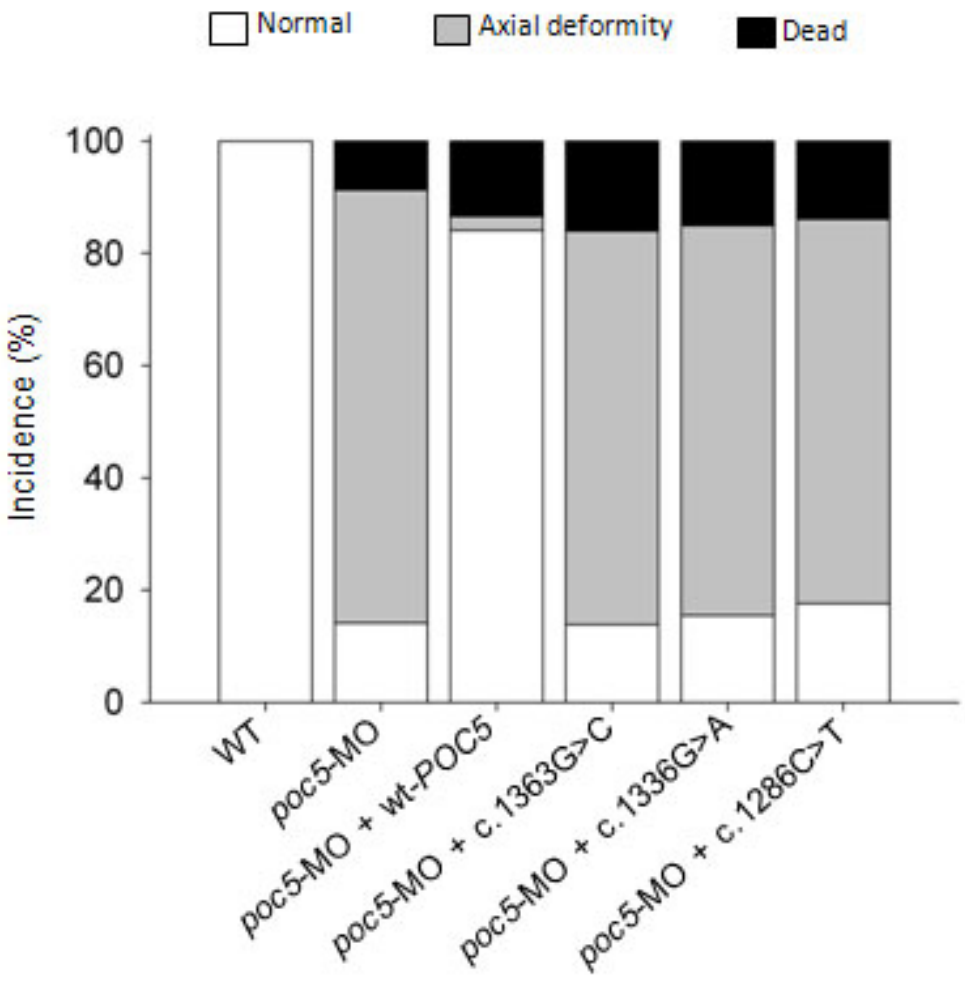
Supplemental Figure 5



Supplemental Figure 5. Morphological phenotype of *Poc5* knockdown zebrafish

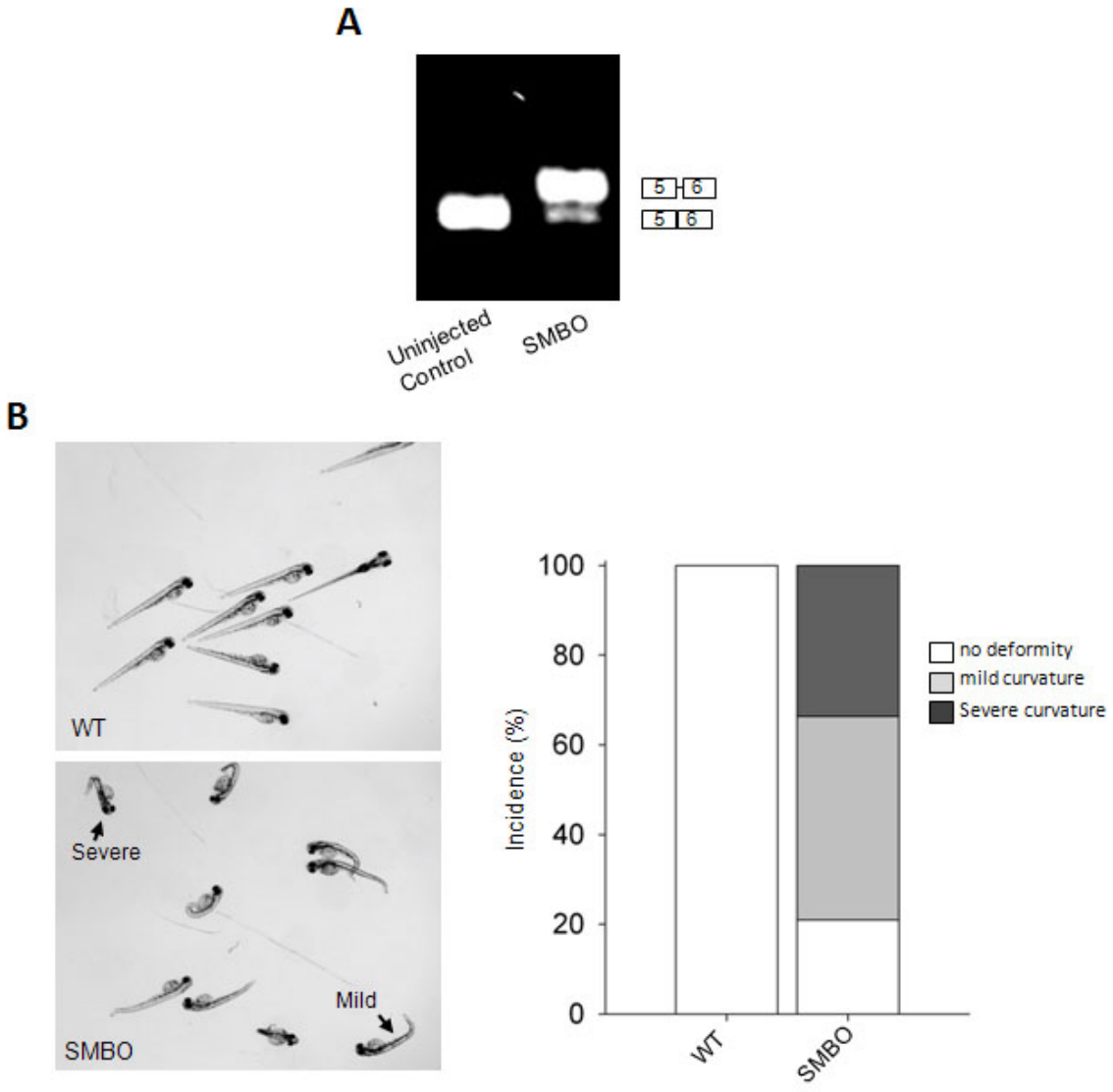
Knockdown embryos (MO-*poc5*) show abnormal axial phenotypes compared to non-injected wild-type (WT) embryos. The morphological phenotype of *poc5* knockdown zebrafish can be rescued by over-expression of human *POC5*.

Supplemental Figure 6



Supplemental Figure 6. Co-injection of *poc5*-MO with wt or mutated *POC5*-mRNA. Knockdown embryos with *poc5*-MO show abnormal axial phenotypes that were rescued by co-injection with wt-POC5 mRNA but unaffected upon by co-injection with mutated versions of *POC5*-mRNA.

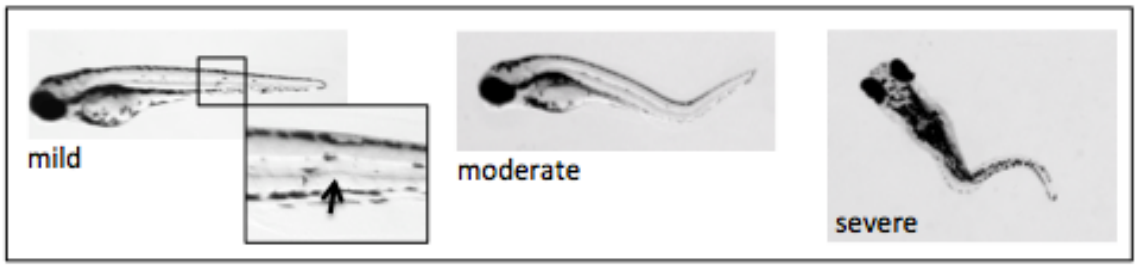
Supplemental Figure 7



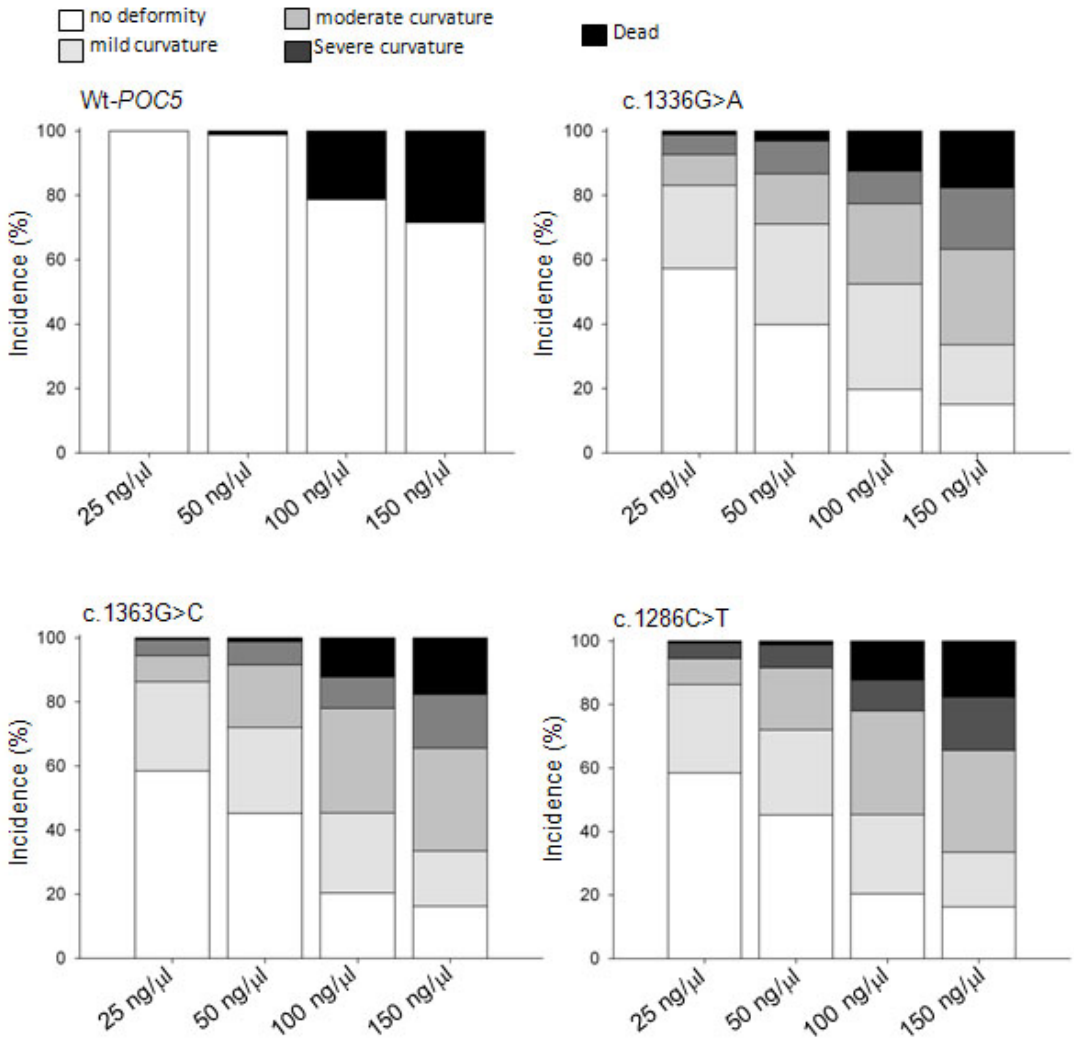
Supplemental Figure 7. *Poc5* splice-blocking morpholino (SBMO) injection results in axial deformities in zebrafish. Reverse transcriptase-polymerase chain reaction (RT-PCR) exhibiting the loss of proper splicing of *poc5* mRNA in *poc5*-splice blocking morphants (A). The increased band size is an indicator of the retention of poc5 intron 5 following mRNA processing. Knockdown embryos (48 hpf) with *poc5*-SMBO show abnormal axial phenotypes (mild to severe) compared to non-injected wild-type (WT) embryos (B). SMBO, splice blocking morpholino; WT, wild-type.

Supplemental Figure 8

A

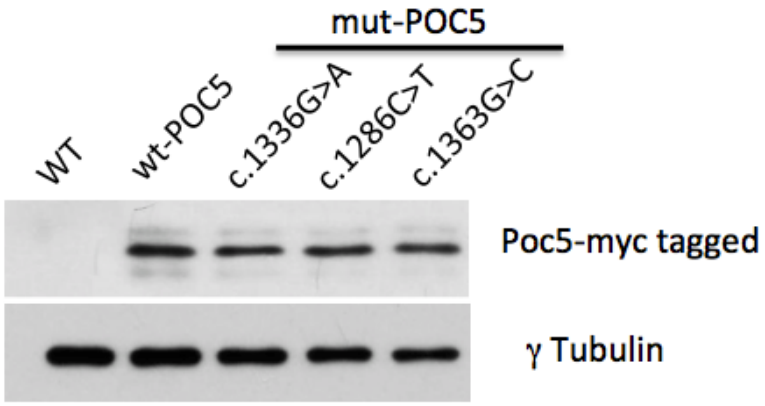


B



Supplemental Figure 8. Dose-reponse of *POC5* mRNA overexpression in Zebrafish. (**A**) POC5 mRNAs overexpression led to mild to severe axial phenotypes. Mild axial phenotype is highlighted by a black arrow. (**B**) Wild-type *POC5* (wt-*POC5*) or mutated *POC5* (mut-*POC5*) mRNAs were injected at concentrations of 25 ng/µl, 50 ng/µl, 100 ng/µl and 150 ng/µl.

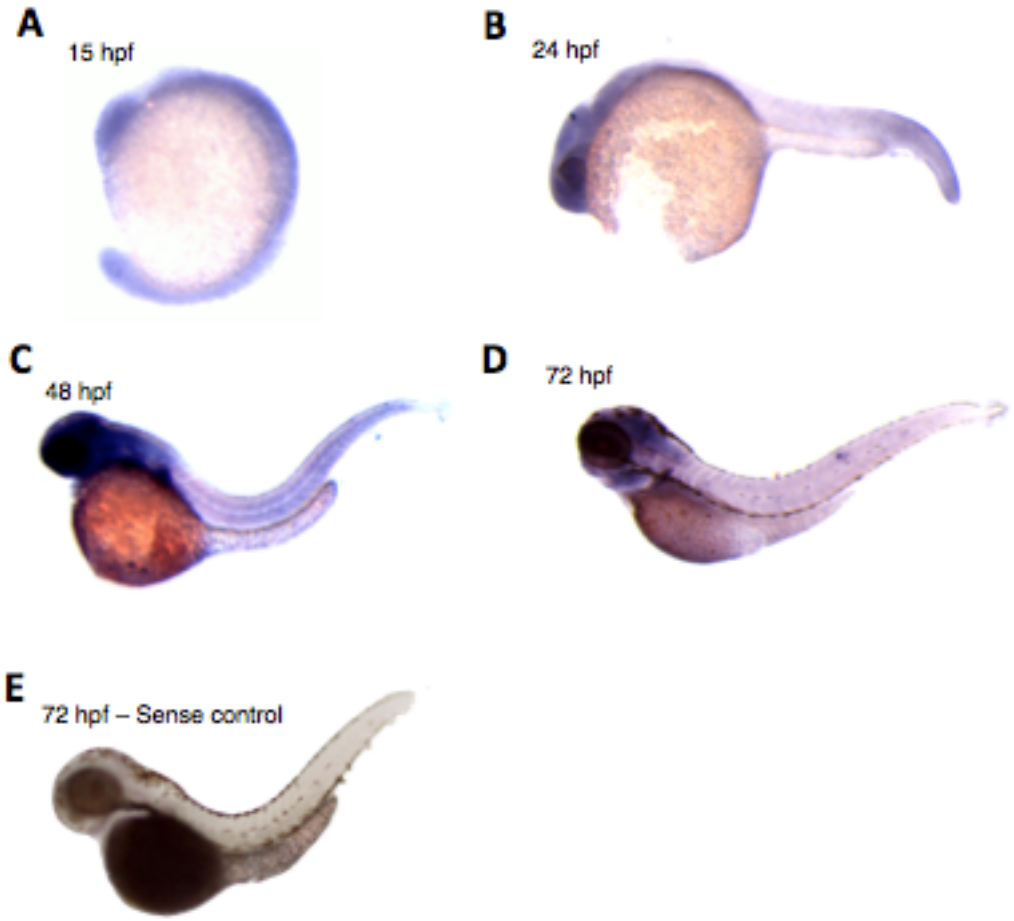
Supplemental Figure 9



Supplemental Figure 9. Expression of Human Poc5 in Zebrafish

Human *POC5* is expressed in zebrafish injected with myc-tagged wild-type *POC5* (wt-*POC5*) or mutated *POC5* (mut-*POC5*) mRNAs, but not in non-injected wild-type (WT) fish.

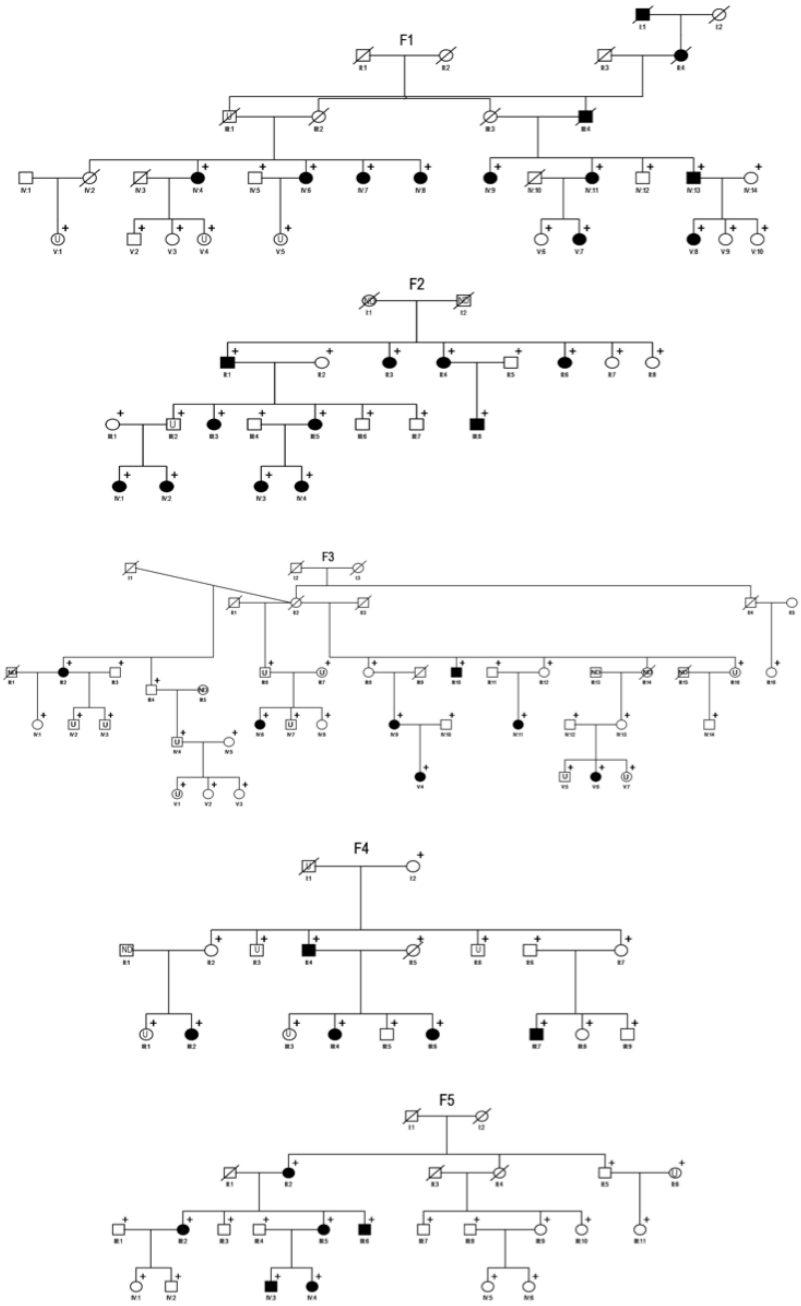
Supplemental Figure 10



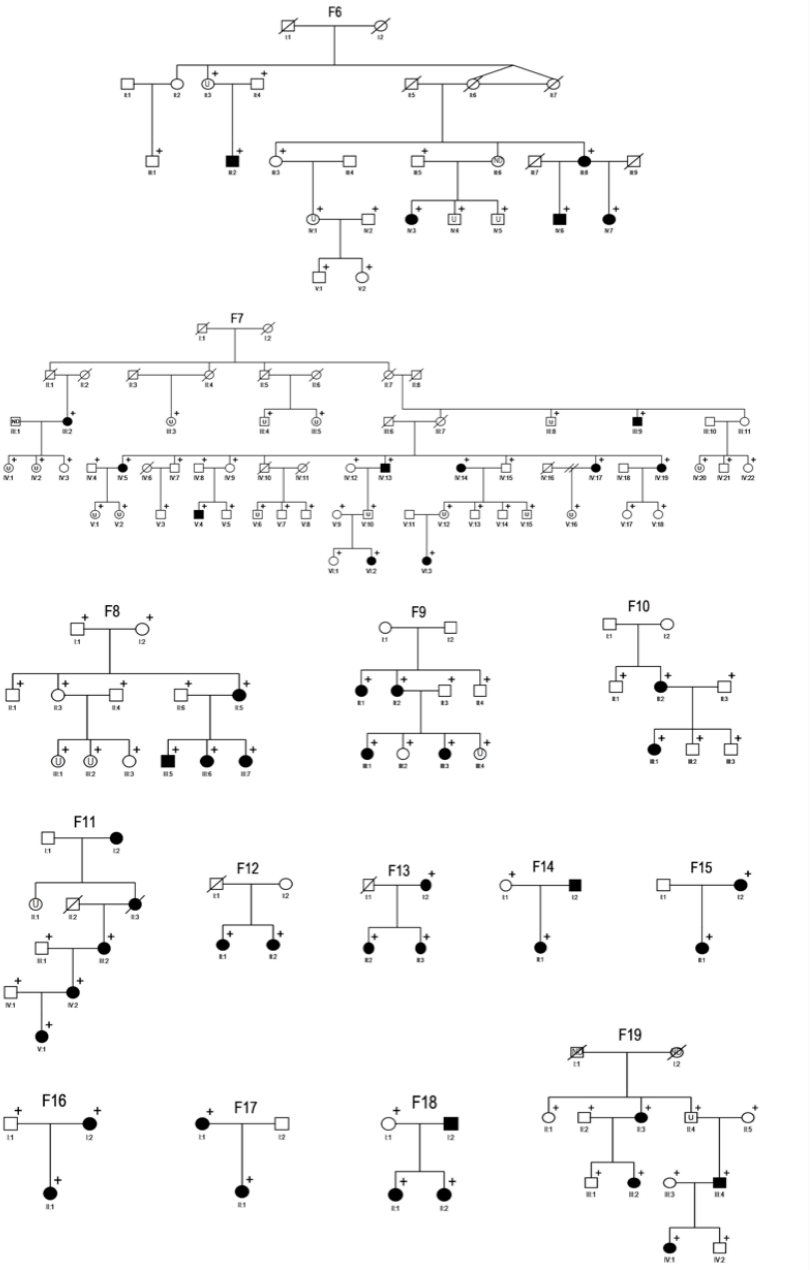
Supplemental Figure 10. *In situ* Poc5 Expression Pattern in Zebrafish

In situ hybridization with specific zebrafish *poc5* antisense at 15 hpf (**A**), 24 hpf (**B**), 48 hpf (**C**), 72 hpf (**D**) and sense probes at 3dpf (**E**). Poc5 was expressed ubiquitously during early somitogenesis. Its expression became restricted to the head and bud region by 24 hpf. By 48 hpf and 72 hpf, its expression became even more confined to the brain.

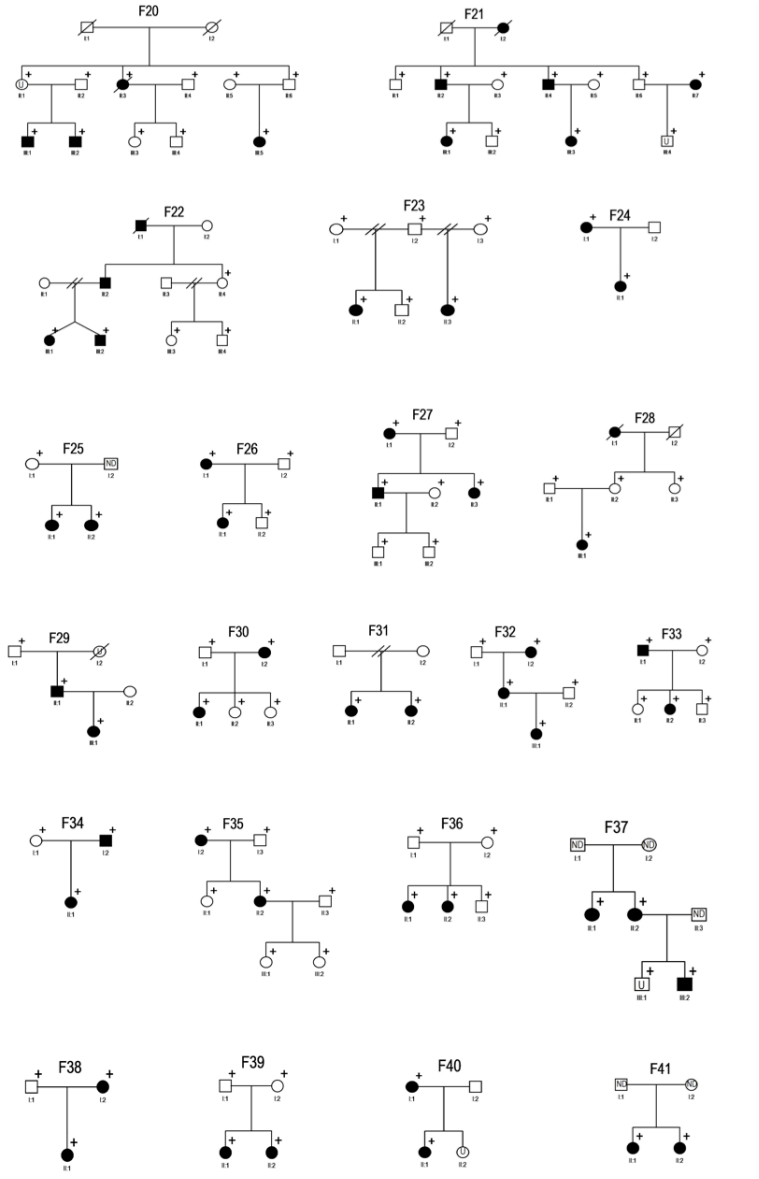
Supplemental Figure 11A



Supplemental Figure 11B



Supplemental Figure 11C



Supplemental Figures 11A-C. Pedigrees of Idiopathic Scoliosis Families F1-F41

Filled symbols indicate affected individuals (i.e., idiopathic scoliosis with Cobb's angle of at least 15° and rotation of vertebrae). U: uncertain status (e.g. idiopathic scoliosis with Cobb's angle below 15° or no rotation). +: DNA sample or blood lymphocytes available. ND: Status not determined.

Supplemental Table 1- Whole-Exome Coverage

	Patient 1 (II:6)	Patient 2 (III:8)	Patient 3 (IV:4)
Mean depth coverage	131X	130X	141X
Coverage > 10X	98.5%	98.5%	98.8%
Coverage > 20X	89.9%	87.6%	90.0%

Supplemental Table 2- Complete List of the 172 Candidate Variants (SNVs (A) +Indels (B))

(A) Single Nucleotide Variants (SNVs)

Chromosome 3			
Variant (NCBI:hg 19)	**RS_ID**	**Variant (NCBI:hg 19)**	**RS_ID**
g.96069538T>A	rs13096522	g.98252027G>A	rs1529047
g.97517118G>C	rs4857294	g.98281078C>T	rs6797035
g.97541018C>T	rs974572	g.98281349G>T	rs9850648
g.97591153C>T	rs17301717	g.98299365T>G	rs1051712
g.97594261G>A	rs6782766	g.98307630C>T	rs75450904
g.97660106A>C	rs4857302	g.98312581G>C	rs4857406
g.97664725C>T	rs2172257	g.98512825T>A	rs14310
g.97726747T>A	rs832032	g.98518072A>G	rs17270986
g.97805954T>C	rs13082722	g.99643176C>T	rs793440
g.97806616G>A	rs4518168	g.99886662G>A	rs11537816
g.97806944T>C	rs80220955	g.100354524A>G	rs1144122
g.97806999T>C	rs6439602	g.100368546A>G	rs61730367
g.97851998A>C	rs79920650	g.100374740T>C	rs9866111
g.97852083C>T	rs75045884	g.100712249T>C	rs3732895
g.97852229T>A	rs9849637	g.100944932A>G	rs75852013
g.97868795A>G	rs4857076	g.100963154G>A	rs571391
g.97887865G>A	rs4133320	g.101066717T>A	rs2433031
g.97887985T>A	rs4133321	g.101232048A>G	
g.97888042A>T	rs4133322	g.101232093C>A	rs55749605
g.97926625A>G	rs9837684	g.101283792C>G	rs3762735
g.97927329C>T	rs28411367	g.101370529T>A	
g.97958054T>C	rs9851509	g.101383562G>A	rs11712748
g.97958253A>G	rs9847708	g.101443461T>C	rs994573
g.97958280C>T	rs9828347	g.101445570G>A	rs111912421
g.97983561G>C	rs9289564	g.105588069G>A	rs11711088
g.97983942A>G	rs9853906	g.107096547G>A	rs709564
g.97983981A>G	rs9871143		
g.97984280C>T	rs17195192		
g.98001777G>C	rs72487753		
g.98002419A>G	rs16839214		
g.98002587A>G	rs16839611		
g.98217178T>A	rs55639376		
g.98220243C>A	rs73140298		
g.98241847G>C	rs6807441		
g.98250862C>A	rs3749260		
g.98250986C>T	rs2230344		

Chromosome 5			
Variant (NCBI:hg 19)	**RS_ID**	**Variant (NCBI:hg 19)**	**RS_ID**
g.73932315T>C	rs9176	g.77656300G>C	rs4072852
g.73980960C>T	rs71627068	g.77784542C>T	rs11740697
g.73981270T>C		g.77784643C>T	
g.74324437G>A	rs3811986	g.78076160C>T	rs2173012
g.74324548G>A	rs3811987	g.78111674A>G	rs34152768
g.74324902C>T	rs4704166	g.78135241C>T	rs25414
g.74364300G>A	rs10942729	g.78181423C>T	rs17220759
g.74400386G>C	rs961098	g.78181477C>T	rs1065757
g.74443132C>T	rs1422698	g.78324352A>G	rs1805074
g.74921686G>A	rs9332464	g.78326750G>C	rs1805073
g.74962768C>T	rs6453139	g.78340286A>G	rs532964
g.74981103C>T	rs34678567	g.78379537T>G	
g.75001582A>G	rs17672542	g.78421959G>A	rs3733890
g.75003678T>C	rs2307111	g.78532658C>T	rs3733893
g.75427518C>T	rs1423099	g.78573790A>T	rs13182512
g.75858215C>T	rs58087114	g.78671747G>T	rs80274918
g.75913301A>G	rs2069702	g.79024734A>G	rs1541813
g.75913305T>C	rs2069685	g.79028327G>A	
g.75923294T>G	rs2431352	g.79028472C>T	rs4704585
g.75923307A>G	rs2909888	g.79028726A>G	rs13158477
g.75932965G>C	rs2455230	g.79029594T>C	rs1019762
g.75948650A>G	rs2431363	g.79086883G>A	rs1129770
g.76003254A>T	rs463188	g.79095417C>T	rs10043986
g.76003258C>T	rs464494	g.79172136A>G	rs265005
g.76114859C>G	rs2242991	g.79172189C>G	rs74916729
g.76114963C>T	rs1529505	g.79282798G>C	rs9293796
g.76115069C>T	rs2243072	g.79331434A>C	
g.76128521G>A	rs616235	g.79331450T>G	
g.76359024C>A	rs34400049	g.79351859G>A	rs405482
g.76373240A>G	rs2303713	g.79351860G>A	rs447875
g.76373241G>C	rs2303714	g.79361265G>C	rs1866389
g.76722443G>A	rs40594	g.79375724G>C	rs2288395
g.76728837T>C	rs335631		
g.76734084C>T	rs33204		
g.76878139T>C	rs13176191		
g.77298619A>T	rs11552314		
g.77425028A>T	rs6453373		

(B) Indels

Chromosome 3				
Variant (NCBI:hg19)	**Reference Allele**	**Mutant Allele**	**Variant class**	**RS_ID**
96069450	CAT	C	deletion	rs34039875
96152499	AT	A	deletion	rs112236687
97367230	C	CA	insertion	rs3214668
97926079	AT	A	deletion	rs5851109
97984691	G	GAA	insertion	rs34155016
98073591	TA	T	deletion	rs11288615
98110406	G	GA	insertion	
98220492	AAG	A	deletion	
98225846	TAGA	T	deletion	rs10603022
98518160	T	TAA	insertion	rs113737993
98518160	T	TA	insertion	rs113737993
99833338	CTG	C	deletion	
100170600	A	ATCCTAGAAGGCATTCTCATGAGGACCAGG AATTCCGATGCCGATCGTC	insertion	
100175184	TC	T	deletion	rs11338136
100295909	A	ATTGTCT	insertion	rs5851214
100570787	TA	TAAA	insertion	
100570787	TA	T	deletion	
100945069	$T(TA)_{20}$	TTA	deletion	
101177901	GA	G	deletion	
101232055	AGG	AG	deletion	
101232056	GGAA	G	deletion	
101370529	TAA	TAAA	insertion	
101370529	TAA	T	deletion	
101370529	TAA	TA	deletion	
101399910	GA	G	deletion	
101443145	AT	A	deletion	rs75043935
101576029	T	TACTTTTAGAAAGCTTTAATAACC	insertion	rs3217713
105377236	CT	C	deletion	rs55698856

Chromosome 5				
Variant (NCBI:hg19)	**Reference Allele**	**Mutant Allele**	**Variant class**	**RS_ID**
73980963	GC	G	deletion	rs70976124
74491715	TTCA	T	deletion	rs10563854
75648938	TA	T	deletion	
75648940	AT	A	deletion	rs112425421
76011613	A	ACGGCCGCGGGAAG	insertion	
76359090	G	GA	insertion	rs34239222
76916335	G	GC	insertion	rs5868876
77524068	T	TA	insertion	rs5868908
77745853	C	CA	insertion	rs113934564
78671727	A	ATT	insertion	
78981369	TAACTG	T	deletion	
78981381	TAAAA	T	deletion	
79279310	T	TTGA	insertion	rs3841613

Supplemental Table 3- Data on the c.G1336A (p.A446T), c.C1286T (p.A429V) and c.G1363C (p.A455P) *POC5* SNVs

Gene	Full name of protein	Chromosome	Rs number	Genomic	Coding DNA Sequence	Protein
POC5	POC5 centriolar protein homolog	5	rs34678567	g.74981103	c.G1336A	p.A446T
			rs146984380	g.74981153	c.C1286T	p.A429V
			-	g.74981076	c.G1363C	p.A455P

Supplemental Table 4– Summary of POC5 Sequencing Data in 41 IS families, 150 Cases and 1268 Controls

Data	Families (n=41, including 330 individuals and 135 patients)	Cases with unknown pedigree data (n=150)	Controls matched for ethnicity with families and cases (n=1268)		Comparison of allelic frequency of the rare variant in IS cases vs controls (Fischer's exact test)
Sequencing Method	Exome + Sanger	Sanger	Sanger (n=103)	Exome (n=1165)	
c.G1336A (p.A446T)	**4/41 (9.8%)** MAF= 4.88%	3/150 (2%) MAF= 1.00%	0/103	19/1165 (1.6%) MAF=0.82%	p=0.0445
c.G1363C (p.A455P)	1/41	0/150 (0%)	0/103	0/1165 (0%)	N/A
c.C1286T (p.A429V)	0/41	5/150 (3.3%) MAF= 1.67%	0/103	9/1165 (0.8%) MAF= 0.39%	p=0.0273

N/A- Not applicable (novel mutation)

Supplemental Table 5- Clinical Data for Idiopathic Scoliosis Patients with c.G1336A, c.G1363C or c.C1286T *POC5* SNV

*NA : Not available

IS Family	*Patient*	*Age at diagnosis (years old)*	*Cobb's angle on radiograph (age: years old)*	*Spine deformity*	*Apical vertebrae*	*Therapy*
F2	II-1	Fortuitous	27° (73)	Right thoracolumbar	D12-L1	None
	II-3	Fortuitous	19° (74)	Right lumbar	L1	None
	II-4	Fortuitous	23° (71)	Right lumbar	L2	None
	II-6	Adolescence	41° (67)	Right thoracolumbar	L1	None
	III-3	Adolescence	15° (49)	Right thoracolumbar	L1	Bracing
	III-5	Fortuitous	29° (46)	Right thoracolumbar	D12	None
	III-8	Fortuitous	15° (42)	Right thoracolumbar	L1	None
	IV-1	Adolescence	29°/29° (20)	Right thoracic/Left lumbar	D8/L2	Physiotherapy
	IV-2	Fortuitous	15° (16)	Right thoracic	D7	None
	IV-3	Fortuitous	14°/16° (19)	Right thoracic/Left lumbar	D8/L1	None
	IV-4	12	18° (15)	Left thoracic/Left lumbar	L1	Physiotherapy
F19	II-3	<18	98° (77)	Left lumbar	L1	None
	III-2	Fortuitous	15° (47)	Right lumbar	L3	None
	III-4	15	38° (41)	Right thoracic	T9	None
	IV-1	10	29°/21° (13)	Right thoracic/Left lumbar	T10/L2	Bracing
F35	I-1	12	76°/80° (55)	Right thoracic/Left lumbar	T9/L2	Bracing
	II-2	14	45° (43)	Right thoracic	T7	Bracing
F41	II-1	12	52°/39° (14)	Right thoracic/Left lumbar	T9/L3	Bracing
	II-2	12	15°/18°/17° (14)	Left thoracic/Right thoracic/Left lumbar	T3/T9/L2	Physiotherapy
F31	II-1	Fortuitous	18°/15° (44)	Right thoracic/Left lumbar	T9/L2	None
	II-2	10	24° (11)	Right thoracic	T9	Bracing
C150	II-1	10	33° (13)	Right thoracolumbar	T11	Bracing
C39	II-1	66	19° (69)	Left lumbar	L3	Bracing
C58	II-1	10	28°/25° (13)	Right thoracic/Left lumbar	T9/L2	Physiotherapy+bracing
C83	II-1	11	27°/25° (12)	Right thoracic/Left lumbar	T9/L3	Bracing
C1	II-1	15	18°	Right thoracic/Left lumbar	T8/L2	None
C77	II-1	12	37°/45° (12)	Left thoracic/Right thoracic	T3/T9	Bracing
C137	II-1	10	64°/60° (13)	Right thoracic/Left lumbar	T7/L2	Surgery
C149	II-1	13	30° (13)	NA*	NA*	Physiotherapy

Supplemental Table 6- Additional *POC5* SNVs Identified Using High-Throughput *POC5* Sequencing

IS Patients	Position (Mb) GRCh37/ hg19	dbSNP138	DNA Change	Mutation	AA change	Transcript ID Ensembl	Protein ID Ensembl	AA Position	Allele frequency (dbSNP, 1000 Genomes)	Effect Prediction (SNPnexus)
C10	74990497	rs190991771	T>C	Missense	I/V	ENST00000428202	ENSP00000410216	225	No allele frequency	Benign
						ENST00000380475	ENSP00000369842	108		Benign
						ENST00000446329	ENSP00000399481	200		Benign
C39, F41 (II.1, II.2)	74998501	rs200926172	C>T	Missense	D/N	ENST00000428202	ENSP00000410216	148	No allele frequency	Benign
						ENST00000380475	ENSP00000369842	31		Possibly damaging
						ENST00000446329	ENSP00000399481	123		Benign
C128	75013289	Novel	C>A	5'UTR					-	CpG:41 Island Change

Supplemental Table 7- Primers Used for High-Throughput *POC5* Sequencing

ID PCR	Left primer sequence	Right primer sequence
PJ1210142_0001	AGGTTCCCTCTCAACACTTTGA	ACATCATGGAGACATCATGTTCA
PJ1210142_0002	TATTTTGATGCTGTAATCAGCAAC	GTGGGGTCTTTTAATCCCTCTG
PJ1210142_0003	GTTACAAAGCATGGTAGAGCTTGAA	CGGACCATTCATCCTGAAAGTA
PJ1210142_0004	ATGGTAGAGCTTGAAAAAGCCTCT	CTCAAGCAACTGCAGCAAAATA
PJ1210142_0005	CAGAGGGATTAAAAGACCCCACT	GATCTTAACAAATGTTTATTGGGTTAA
PJ1210142_0006	TCTGGAAGCTGAGGTACTACTTTCA	CACTGTGGGTGTTGAACATGTC
PJ1210142_0007	AATTAATTTCCCAACAGCAGAAA	CTCCCATGAGCTCAGTTGTTGT
PJ1210142_0008	AGTCTATGAACTCTCAGGAAAAAGACTT	GATTTTGCTGTGGATTTTCTGC
PJ1210142_0009	ACTCACCACTGTGACTGGATGA	ATTTTAAGTGCCTGTGTATTCTTCA
PJ1210142_0010	CCTGAGTAGCTGGGACTACAGG	GCATCTTCTGTTCACGTTCCTG
PJ1210142_0011	CTCCCAAAGTGCTGGGATTAC	CTCCGATGCCCTTACCAGTTAC
PJ1210142_0012	ACCATTTCTTCTGATGCAGCAG	TTTGTGATTTATAGGGATAGACTCCA
PJ1210142_0013	CGGCTGGTGGGGATG	TGTAAACATCTAAATTTTTGTTAGGACCA
PJ1210142_0014	TTCTTTTCCTTGAACACCAGGA	CACATGTGGAAGGAAGTAGTCTGA
PJ1210142_0015	ACCAATCACCAAAATCTCCTCA	TCTTCTTTGTATCTCAAATTGTTTTGC
PJ1210142_0016	ACCAAAATCTCCTCAAATCTTTTT	TTCTTCTTTGTATCTCAAATTGTTTTG
PJ1210142_0017	ACCGCGCCCAACTAATAATTT	TTGTCAAGCAAGAGCTGAAGAA
PJ1210142_0018	TTGTCTGCATTCTTGATTAAAGACC	TGACCAGTACTACCAGAGAACTTTACTG
PJ1210142_0019	AGATGCTTACCATAGCAACTTTGG	TGTACACTTACCACCATGTTATGTTT
PJ1210142_0020	CACTACGGAACGCCAGACTTT	TGTGAACTACGTAGCTTGCTTAACC
PJ1210142_0021	AAAACAACTATAATAAGTGATCCTGAGCA	TGAGAAAATCACTTTACCAATTGC
PJ1210142_0022	TGAATTCAGAAGTCTAACATCCATCA	TGATCCATACCTATCCTCTCTAGCA
PJ1210142_0023	TTATCTAGCAAGACATTTACTGAATCTCA	TTTTAAGAAGTGATATTAACGAACAAACT
PJ1210142_0024	TCTAGCAAGACATTTACTGAATCTCAGT	GTTTTAAGAAGTGATATTAACGAACAAAC
PJ1210142_0025	TGCTGAAGACCTACTGCATATGAA	TGATGAAAACCTTCAGAAGATGG
PJ1210142_0026	CTCAGGAAACAAAAGATTTTTAGAAA	CTTCTCACCCAGTCATGGATTT
PJ1210142_0027	CCTGAACTCCAAAGATCAAGCA	CGTCTTTTAGCCTTCCAGTATGG
PJ1210142_0028	GCTGGTGAGGAAGAGTCAGCTA	GGTATCTCATGCCAGTCTGTGC
PJ1210142_0029	AATATGCTGAAATTATTCCTCTTACTACA	ACTGCTACCTTCTTAAATTATGTGTGA
PJ1210142_0030	ATTCATCAAACCACAAAATGTTG	GTGTTCCTATATCCCCAGCATG
PJ1210142_0031	TGAAGTTCAAAGAAAAATCAAGCTG	TCATCCTAAGGGAGAATTGGTG
PJ1210142_0032	ATCACTGATCTGGACAGGCATT	TTGATTATTGGCCTCTCTTAGGAAG
PJ1210142_0033	AACCTTGACTATGAAGAATATCATGAA	AATAAGCATGGGACTCTATTATGGTAA
PJ1210142_0034	ATGACTGTGAAGCACAGGGTTC	CCAGCCCTCTTCAAACTGTTAA
PJ1210142_0035	CTCCAGAAATCTAAATCCATATTTTTG	CATTTTAGCTACCTCTCATGAATGC
PJ1210142_0036	GAAATAAGAGAAATTTAAAACATTTCATG	GAGGTGAAACAAATGTTCAAGAAA
PJ1210142_0037	TTGAAAAACTCCCTTGTAAATGG	GGAAAGGATTTTATCTTAAATATCAAGG
PJ1210142_0038	AAAAATCCTAGTTTTCCCTTACATTCA	TGCAGATTTGGATACTGTTGCA
PJ1210142_0039	CATCCCTCACTCCTGCTCACT	CACTTGCTGACACTGCAGCT
PJ1210142_0040	GCGCCAAGGAACTTTAAATCTC	GCAGATTGCTGAAACAAAGGAC
PJ1210142_0041	CGACCAAATCCCGACTCCT	CGCCCCCTACCAACCTG
PJ1210142_0042	TTCAAACTGCAGGGAGGAATTA	CTCCCGGAGCCGCTTAG

Supplementary Table 8- Primers (5'-3') Used for Site-Directed Mutagenesis of Human POC5 Open Reading Frame

c.G1336A	Forward	ACCAGGGCTGCTTCC**A**CATCTTCTGTTCACG
	Reverse	CGTGAACAGAAGATG**T**GGAAGCAGCCCTGGT
c.G1363C	Forward	GAGGAGCCAGCG**T**GACTGCCGTTCC
	Reverse	GGAACGGCAGTC**A**CGCTGGCTCCTC
c.C1286T	Forward	TCACGTTCCTGTTTCT**C**CTCTTGGTGCAGGATC
	Reverse	GATCCTGCACCAAGAG**G**AGAAACAGGAACGTGA

All substitutions are noted in bold, underlined. All constructs were verified by sequencing.

References

1. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. 2002. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97-101.
2. Merner, N.D., Girard, S.L., Catoire, H., Bourassa, C.V., Belzil, V.V., Riviere, J.B., Hince, P., Levert, A., Dionne-Laporte, A., Spiegelman, D., et al. 2012. Exome sequencing identifies FUS mutations as a cause of essential tremor. *Am J Hum Genet* 91:313-319.
3. Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., et al. 2011. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* 43:860-863.
4. Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
5. 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.
6. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491-498.
7. 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.
8. Wang, K., Li, M., and Hakonarson, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38:e164.
9. Thiele, H., and Nurnberg, P. 2005. HaploPainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics* 21:1730-1732.
10. Cha, Y.R., and Weinstein, B.M. 2012. Use of PCR template-derived probes prevents off-target whole mount in situ hybridization in transgenic zebrafish. *Zebrafish* 9:85-89.
11. Thisse, C., and Thisse, B. 2008. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc* 3:59-69.
12. Bekri, A., Billaud, M., and Thelu, J. 2014. Analysis of NUAK1 and NUAK2 expression during early chick development reveals specific patterns in the developing head. *Int J Dev Biol* 58:379-384.