Supplementary Materials

Supplementary methods

Patient recruitment and phenotyping

From October 2009 to December 2016, 447 patients of Chinese Han ethnicity affected with EOS (scoliotic age of onset ≤ 10 years) and who underwent spinal surgery in Peking Union Medical College Hospital were consecutively recruited as part of the Deciphering Disorders Involving Scoliosis and COmorbidities (DISCO) study (http://www.discostudy.org/).

Physical examination, X-ray, computed tomography (CT), magnetic resonance imaging (MRI), and B-ultrasound scanner (BUS) were performed to give a prior etiological diagnosis for each patient:

(1) Congenital scoliosis (CS): Patients with structural malformation of the vertebrae revealed by X-ray or CT and not diagnosed with specific syndromes were assigned to the CS subgroup.

(2) Neuromuscular scoliosis (NMS): Patients without structural vertebrae and affected with neuropathic or muscular disorders (e.g., cerebral palsy, myelodysplasia, muscular dystrophy) were assigned to NMS subgroup

(3) Syndromic EOS: Patients who fulfilled the clinical diagnostic criteria of scoliosis-associated syndromes, such as Ehlers-Danlos syndrome, neurofibromatosis type I and achondroplasia, were assigned to syndromic EOS group.

(4) Idiopathic early-onset scoliosis (IEOS): Patients without clear causal agent and

underlying etiology were classified as IEOS.

The family history for each case was taken by trained orthopedic doctors from oral descriptions of the patients and/or their family members. If there is suspicious of the family history, the physical examination or spine X-rays were required to confirm the phenotype of family members. Detailed phenotyping of all the patients were performed by 3 orthopedic surgeons in a double-blinded fashion. Besides characterization of scoliosis, extensive clinical features were evaluated:
1) the Cobb angle upon initial measurement, 2) history of muscular complications, 3) number of organ systems involved (ocular, urinary, cardiovascular, digestive), 4) age of presentation, 5) presence of intraspinal defects, 6) presence of vertebral malformation, 7) presence of chest deformity, 8) number of spinal curvatures, 9) gender, 10) presence of kyphosis, 11) presence of lordosis, and 12) presence and degree of respiratory disturbance.

Exome sequencing

For the Chinese cohort, Illumina paired-end libraries were prepared from DNA samples. Following the library construction, four different capture kits were used dependent on the time of enrollment: xGEN targeted capture kit (IDT), seqcap pure capture kit (Nimblegen), VCRome SeqCap EZ Chice HGSC 96 Reactions (Roche), and All Exon V6+UTR r2 core design (91 Mb, Agilent). The mean coverage, percent of 20X coverage, and Q20 were presented in **Table S1**.

For the American cohort, DNA was extracted from whole-blood using the QIAamp DNA Blood Maxi Kit (QIAGEN, Valencia, CA). ES of 42 samples from 13 EOS families was performed at the Human Genome sequencing center (HGSC) at Baylor College of Medicine through the Baylor-Hopkins Center for Mendelian Genomics initiative. Using 0.5ug of DNA an Illumina paired-end pre-capture library was constructed according to the manufacturer's protocol (Illumina Multiplexing_SamplePrep_Guide_1005361_D) with modifications as described in the BCM-HGSC protocol (https://www.hgsc.bcm.edu/content/protocols-sequencing-library-construction). Six pre-captured libraries were pooled and then hybridized in solution to the HGSC VCRome 2.1 design [1] (42Mb NimbleGen, Cat. No. 06266380001) according to the manufacturer's protocol. The sequencing run was performed in paired-end mode using the Illumina HiSeq 2000 platform, with sequencing-by-synthesis reactions extended for 101 cycles from each end and an additional 7 cycles for the index read. Illumina sequence analysis was performed using the HGSC Mercury analysis pipeline [2 3] (https://www.hgsc.bcm.edu/software/mercury). The samples achieved 95% of the targeted exome bases covered to a depth of 20X or greater with the average coverage overall samples equal to 93X.

Variants were merged using VCFTools.v0.1.16 [4] per each family. We used ANNOVAR[5] to annotate frequency of each allele in the Exome Aggregation Consortium (ExAC) database, location of the variant within genes (including exonic, splice site, nearby intronic and intragenic), and predicted variantal consequence for all

variants. We filtered for segregating variants applying inheritance models derived from family pedigrees as shown in **Table S2**. Candidate variants per family were then restricted to rare variants (ExAC_MAF<=0.01) that altered the coding sequence (missense, nonsense, splice-site, frameshift and non-frameshift Indels). Resulting variants were further annotated with current information available from ClinVar [6] and the Online Mendelian Inheritance in Man (OMIM) [7] databases using custom perl scripts. Sequencing quality of selected variants was checked in the Integrative Genomics Viewer V2.3 (IGV) [8], and likely casual candidates were confirmed by Sanger sequencing.

Variant-calling and annotation

The variant-calling and annotation were performed by the in-house developed PUMP (Peking Union Medical college hospital Pipeline). Single-nucleotide variants and internal duplications and/or deletions (aka indels) were called using the HaplotypeCaller of the Genome Analysis Toolkit (GATK), version 3.4.0. Annotation of *de novo*, compound heterozygous, and recessive inherited variants were calculated with Gemini (version 0.19.1) for *in silico* subtraction of parental variants from the proband's variants, with accounting for read number information extracted from BAM files. Computational prediction tools (GERP++ [9], Combined Annotation Dependent Depletion (CADD) [10], SIFT [11], Polyphen-2 [12], and VariantTaster [13]) were used to predict the conservation and pathogenicity of candidate variants. All variants

were compared against publicly available databases such as the 1000 Genomes Project (http://www.internationalgenome.org/), the Exome variant server, NHLBI GO Exome Sequencing Project (ESP) (http://evs.gs.washington.edu/EVS/), and the Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org/).

Single nucleotide variants (SNVs) /indels interpretation

The protocol for interpretation of SNVs and indels was adapted from the American College of Medical Genetics and Genomics (ACMG) guidelines [14], and is summarized in **Figure S1**.

Variants annotated by the PUMP pipeline were first filtered against a population frequency of 0.001 based on 1000 Genomes (October 2013), the Exome Aggregation Consortium (ExAC; http://exac.broadinstitute.org), and the genome Aggregation Database (gnomAD , http://gnomad.broadinstitute.org/).

Pathogenicity of retained variants were evaluated according to variant type, reported evidence, *de novo* origin, and presence of *in trans* allele. Variants predicted to cause protein truncation, including stop-gain, frameshift, splice acceptor/donor variants, were classified as pathogenic variants. Reported evidence were assessed by 1) number of unrelated patients with corresponding disease phenotypes; 2) for missense variants/in-frame indel, functional validation using animal model or *in vitro* test. When a variant has previously been reported in more than 3 unrelated patients, or reported in one patient with well-designed functional validation, which are considered

as strong evidence, the variant is classified as a pathogenic variant. By contrast, a variant which has been reported in only one patient is considered to be supported by moderate evidence, thus classified as a likely pathogenic variant. *De novo* origin of a variant in a trio consisted of proband and unaffected parents also qualify it as a likely pathogenic variant. For genes with an autosomal recessive inheritance mode, a variant with another variant *in trans*, which together constitute a biallelic variant pair, was also classified as a likely pathogenic variant.

Anticipated mode of inheritance associated with the identified genes was then considered (**Figure S1**). For genes whose variants show an autosomal dominant or X-linked inheritance pattern, a heterozygous variant is sufficient to be potentially disease-causing. For genes whose variants show an autosomal recessive inheritance mode, biallelic variants revealed through trio exome sequencing were required to suspect a gene to be disease-causing. For *TBX6*-related scoliosis (TACS), a *TBX6* null allele (16p11.2 deletion or truncating variant) and an *in trans* T-C-A (rs2289292, rs3809624, rs3809627) haplotype were required to make a molecular diagnosis of TACS. For apparently homozygous variant, a previously reported algorithm Homozygous and hemizygous deletion finder (HMZDelFinder)[15] and BafCalculator tool (https://github.com/BCM-Lupskilab/BafCalculator/blob/master/BafCalculator.R) were used to check whether the 'homozygous state' was caused by an AOH (Absence of Heterozygosity) region. If the observed variant(s) is/are pathogenic and consistent with the expected mode of inheritance, gene-related phenotypes were compared to the patient phenotype. A likely

molecular diagnose was made if the phenotypic spectrum of the gene could explain the whole clinical presentation of the patient.

Variant in the 59 genes recommended by the ACMG was also reviewed for medically actionable secondary findings [16]. Only variants which could be classified as known pathogenic (KP) or expected pathogenic (EP) was reported. EP variants were only reported among genes for which reporting both EP and KP variants were recommended.

All per locus diagnostic variants and incidental findings were confirmed by orthogonal experiments using Sanger sequencing. Genomic regions encompassing the variants were amplified by PCR from genomic DNA obtained from probands and parents and purified using an Axygen AP-GX-50 kit and sequenced by Sanger sequencing on an ABI3730XL instrument.

Haplotyping of T-C-A (rs2289292, rs3809624, rs3809627) SNPs

Haplotyping of three SNPs (rs2289292, rs3809624, rs3809627) were performed on patients with heterozygous *TBX6* null variants.

(A) Long-range PCR to amplify the whole *TBX6* gene.

Forward primer: 5'-

TCGCGAATGCGTCGAGATGCTCCTTCGGGCGCGCCTCAGCGCTGAGCC-3'

Reverse primer: 5'-

TCGGTCCCGGCATCCGATTCAGGCAGCTGGAAACACAGGTCTAGATCT-3'

PCR reaction conditions:

94℃

2min

98℃ } 40

10sec cycles

68℃

4min

68℃

5min

4℃ ∞

PCR system(50μl):

10× PCR buffer for KOD-Plus-Neo	5μl
dNTPs Mixture (each 2mM)	5μl
Forward primer (10 μM)	1μl
Reverse primer (10 μM)	1μl
Template(200ng/μl)	3μl
KOD-Plus-Neo	1μl
25mM $MgSO_4$	4μl

dd H_2O	30μl

All of PCR products were verified by 1% gel and were collected by gel extraction using universal DNA purification kit (TIANGEN, Beijing, China).

We amplified both the vectors and the insert DNA fragment, the above-mentioned products of PCR, to perform recombination cloning. The pUCSF-simple vector (Hitrobio) was used as template for amplifying the vectors. Both vector and insert DNA fragments were amplified using ExnaseII polymerase (Vazyme).

Recombination reaction system:

dd H_2O	Up to 20μl
5×CEII Buffer	4μl
Vector	50~200ng
Insert fragment	20~200ng
Exnase II	2μl

The PCR products were dealt with universal DNA purification kit (TIANGEN) for gel exaction. After ligation, the products were transfected to JM109 competent cells. The bacterial colonies were collected and verified by Sanger DNA sequencing.

Sequencing primer: 5'-TGTAAAACGACGGCCAGT-3'

Gene ontology (GO) clustering

Gene ontology (GO) clustering of 33 causal gene was performed using DAVID [17]. The gene list was input by HGNC-approved gene symbols. Significance threshold of

clustering was a P-value of 0.01.

Computational CNV calling based on ES data

Computational CNV calling was based on ES data independently using two software: XHMM [18] and CoNIFER [19] (copy number inference from exome reads).

For XHMM, coverage information was computed from BAM files using GATK. Then, extreme targets or samples were filtered out. Highly variable targets were also filtered out. After quality control, read depths for each sample were analyzed and a z score was calculated. CNVs were called according to the z-score [18].

For CoNIFER, exome sequencing reads from FASTQ files were firstly divided into nonoverlapping 36-bp constituents and aligned to targeted regions, which allow for up to two mismatches per 36-bp alignment. For each exon or targeted region, RPKM values were calculated and then transformed these into "ZRPKM" values based on the median and standard deviation of each exon across all samples. Then, ZRPKM values were inputted into the SVD transformation, where the first 12–15 singular values were removed. Finally, a centrally weighted 15-exon average was passed over the SVD-ZRPKM values in order to reduce false positives, and a 61.5 SVD-ZRPKM threshold was used to discover CNVs [19].

Copy number variant interpretation

CNVs annotated by PUMP pipeline were manually assessed and classified into three categories: benign, uncertain clinical significance, and recurrent pathogenic. A CNV is firstly checked in the recurrent CNV region reported in the Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources [20] (DECIPHER community, https://decipher.sanger.ac.uk/). We defined a CNV as recurrent pathogenic CNV by using the threshold for reciprocal overlap >80% with the reported CNVs or the exact CNV has already been reported in previous studies. The remaining CNVs (<80% overlap with reported CNVs) were checked in the Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/app/home) and the segmental duplication regions. We defined a CNV as benign if it is shown with a frequency >1% in the DGV database and/or with an overlap >80% in the segmental duplication regions. The variants of uncertain clinical significance CNV can be defined as the exact CNV is shown with a frequency <1% in the DGV database and overlap of a small region (<80%) with the segmental duplication regions.

Comparative genomic hybridization microarray (aCGH)

Pathogenic CNV identified by ES were validated using aCGH.

The genomic DNA extracted from peripheral blood of each subject and the sex-matched reference DNA were fragmented using AluI and RsaI enzyme. Agilent

SureTag DNA Labeling Kit was used for DNA labeling. Cy5-dUTP was used for dying the DNA of each subject and Cy3-dUTP was used for reference DNA. Labeled subject DNA and reference DNA were hybridized onto Agilent human CGH microarrays (including the formats of 1×1M and 8×80K).

DNA processing, microarray handling, and data analysis were conducted by following the Agilent oligonucleotide CGH protocol (version 6.0).

Phenotypic analysis using random forest

Eleven clinical features were selected and transformed into discrete variations: 1) Cobb angle (the max Cobb angle when an individual has more than one curve), 2) musculoskeletal complication, 3) number of multi-systemic defects (ocular, urinary, cardiovascular, digestive), 4) onset age, 5) presence of intraspinal defects, 6) presence of vertebral malformation, 7) presence of chest deformity, 8) number of curvatures, 9) gender, 10) presence of kyphosis, and 11) presence of lordosis.

All patients were divided into three group according to their diagnostic status: 1) undiagnosed; 2) diagnosed by genes or CNVs other than *TBX6* and 16p11.2 deletion (non-TACS); 3) *TBX6*-associated scoliosis (TACS).

Clinical feature data and patient grouping information were fed into a random forest (RF) classifier [21] (number of trees= 15; maximum depth = 3). The model was trained for 100 times with different random seeds to reduce the effect caused by the intrinsic randomness of RF algorithm. A mean reciprocal rank (MRR, 0-1) as output

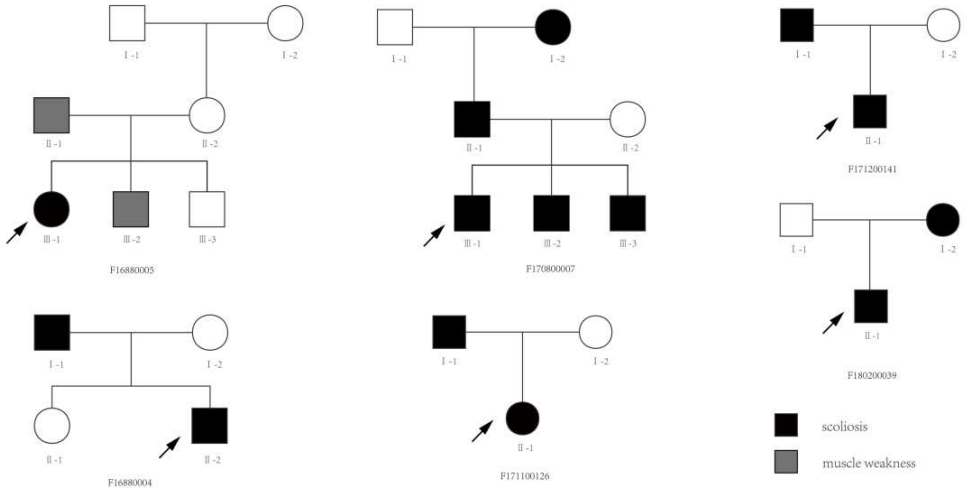
demonstrates the degree of relevance between clinical features and patient groups (the more relevant, the higher).

Return of molecular diagnosis

For patients who gained a molecular diagnosis, information about the syndrome and causal genetic variants were returned to the patients through the genetic counselling clinic. For patients who did not gain a molecular diagnosis, a negative result was delivered through either phone call or the genetic counselling clinic.

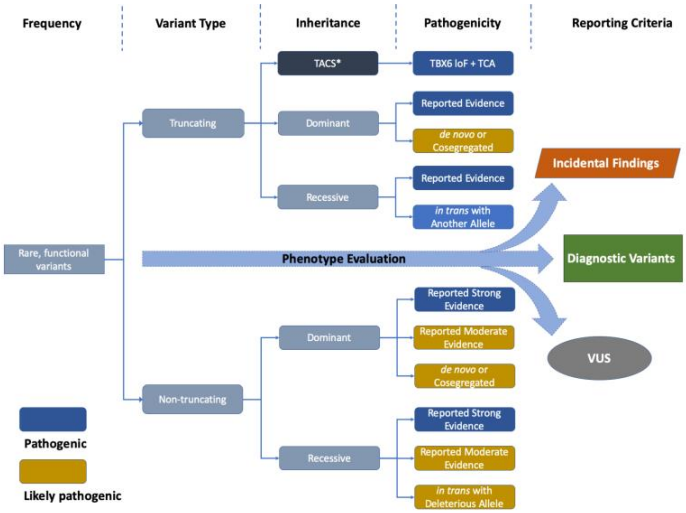
Supplementary figures

Figure S1. Six families in the Chinese cohort with likely dominant inheritance



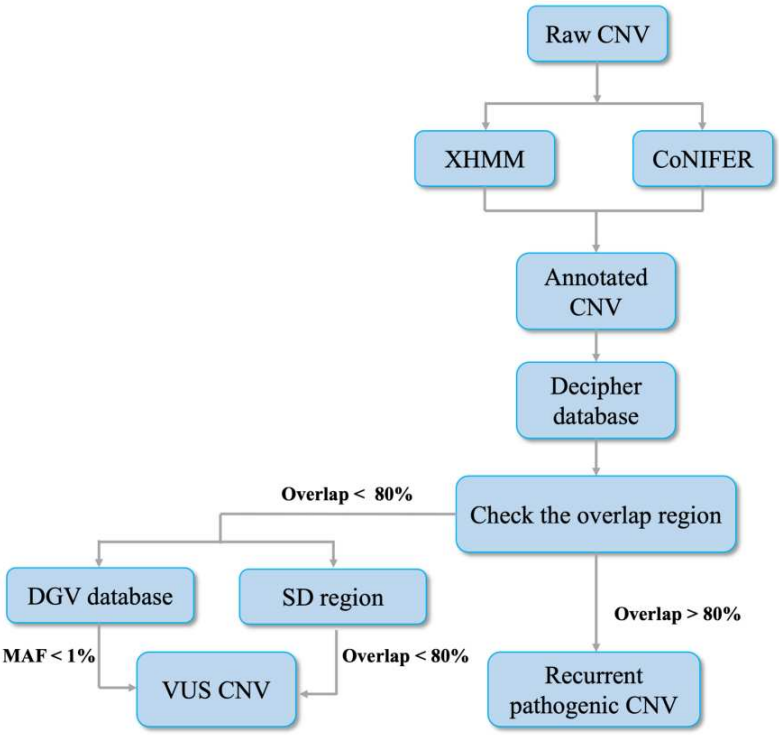
In 105 families recruited, 6 presented a likely dominant inheritance, and probands (arrows) of five families received a molecular diagnosis.

Figure S2. Workflow of single nucleotide variants /indels interpretation



Interpretation protocol of SNVs and indels were performed adapted from American College of Medical Genetics and Genomics (ACMG) guidelines, taking into consideration the variant pathogenicity, mode of inheritance, and patient phenotype. MAF, minor allele frequency; P, pathogenic; LP, likely pathogenic; TACS, *TBX6*-associated congenital scoliosis.

Figure S3. Workflow of copy number variation (CNV) analysis

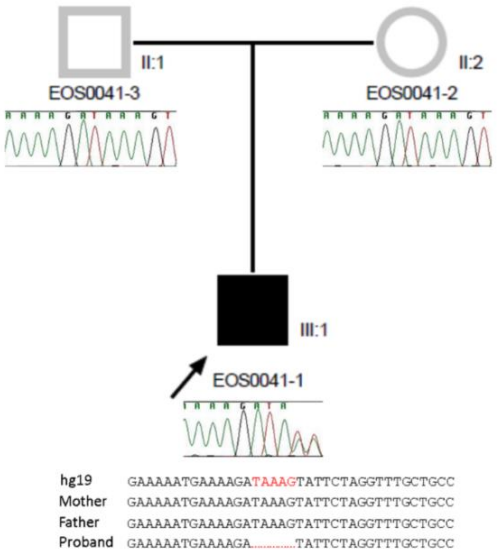


CNVs annotated by PUMP pipeline were manually assessed and classified into three categories: benign, uncertain clinical significance, and recurrent pathogenic.

Abbreviations: SD, segmental duplication; CNV, copy number variation; VUS, variation of unknown significance

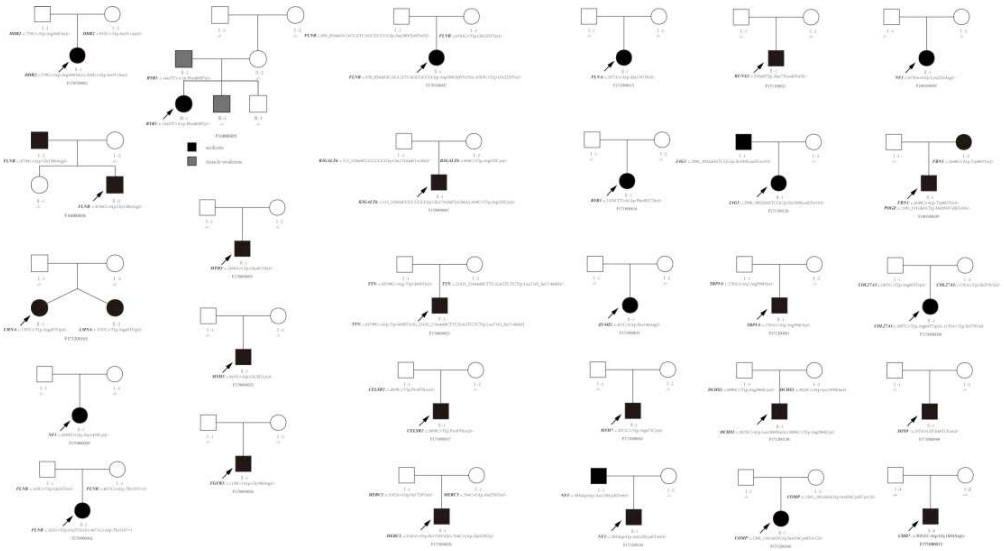
Figure S4. Sanger sequencing confirmed *de novo NSD1* variant in subject

EOS0041-1



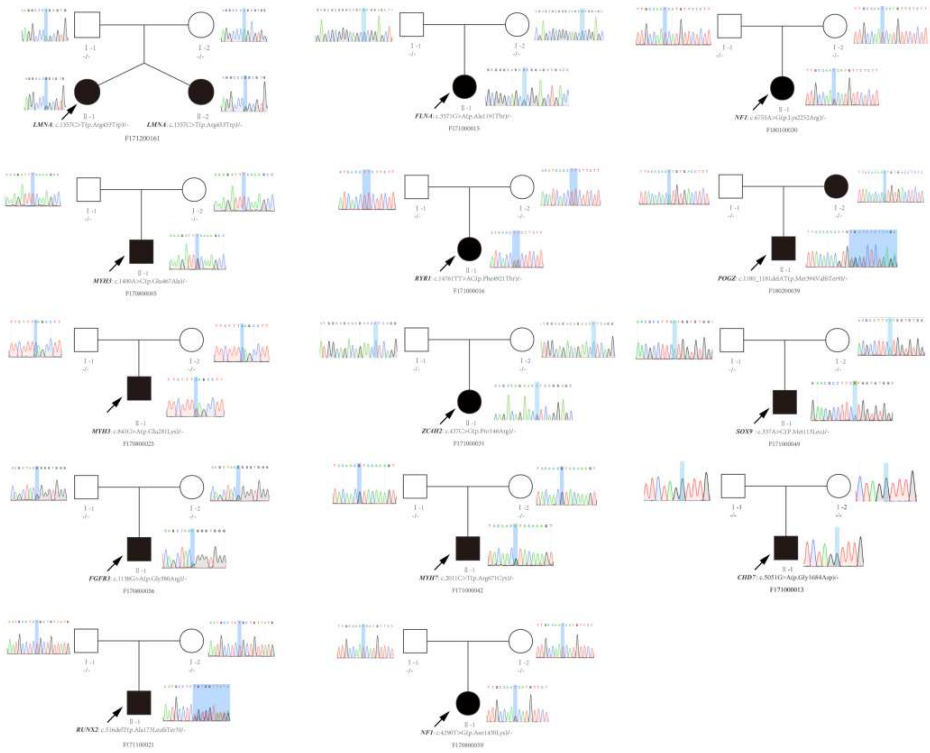
Sanger sequencing and alignment results of EOS0041-1 and his parents.

Figure S5. Pedigrees of 28 molecular diagnosed cases (non-TACS) with family members sequenced



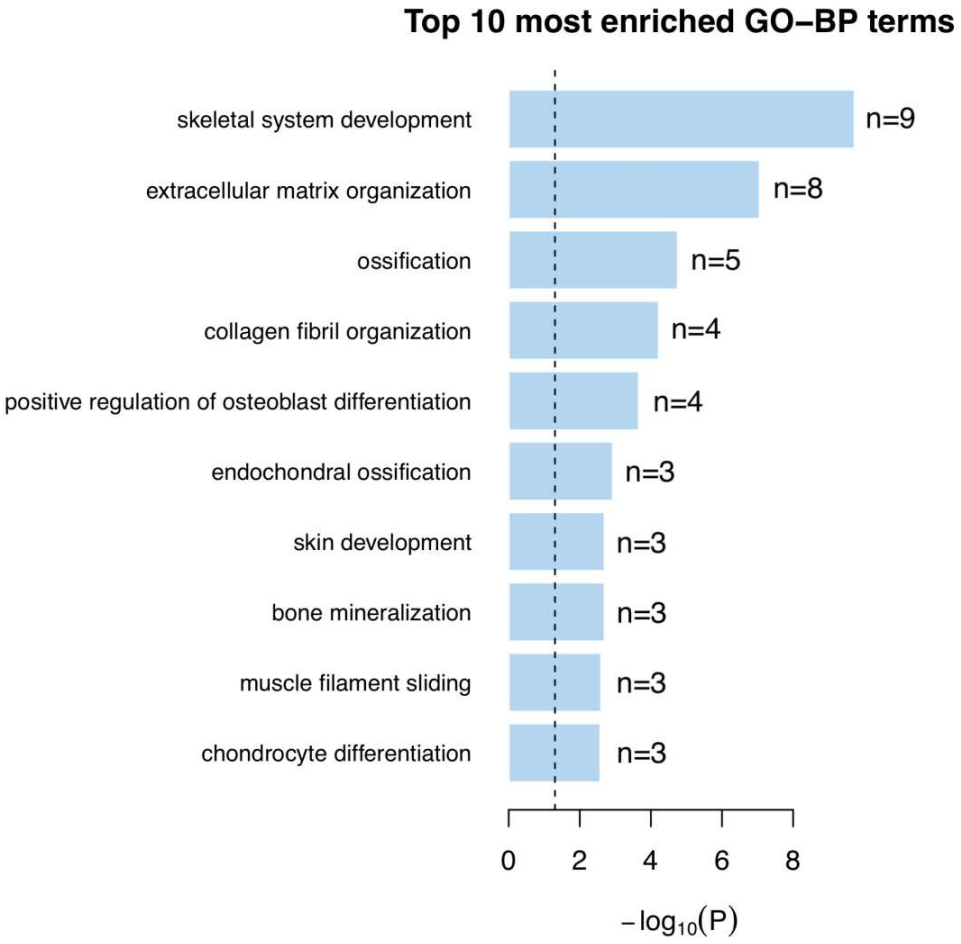
Of 46 patients diagnosed by genes other than TBX6, 28 have family members sequenced. Family ID, disease-causing genes and variants were labeled for each family. Probands in families with more than one patient are indicated by arrows.

Figure S6. Validation of *de novo* variants by Sanger sequencing



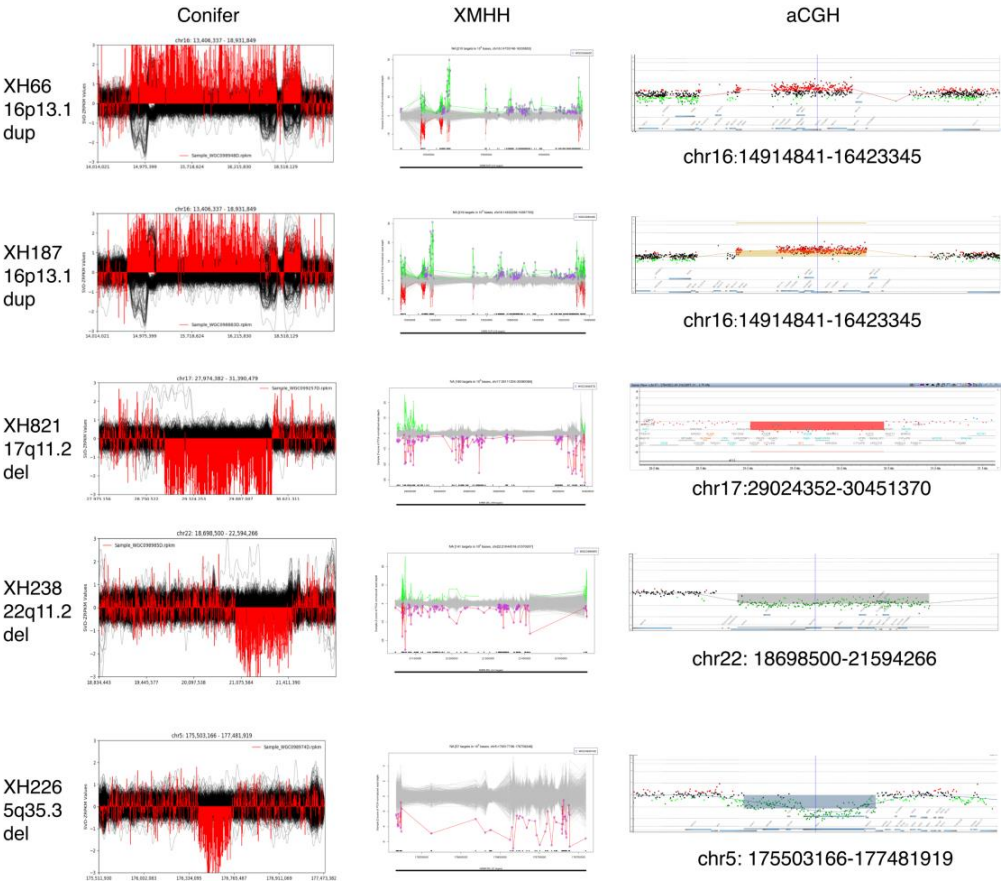
Thirteen diagnostic *de novo* variants were identified from thirteen families, and were validated using Sanger sequencing. Family ID, disease-causing genes and variants were labeled for each family.

Figure S7. Gene ontology clustering for 34 disease-causing genes



Pathway clustering was performed based on gene ontology (GO) term. Top 10 most enriched GO Biological Process (GO-BP) were presented in the order of their p-value. Dash line demonstrates a p-value of 0.05.

Figure S8. Pathogenic CNVs identified in 16p13.1, 17q11.2, 22q11.2, and 5q35.3 region



Visualized diagram of five pathogenic CNVs in five patients were drawn by Copy Number Inference from Exome Reads (CoNIFER) and XHMM. Validation of CNV called from ES data were performed using comparative genomic hybridization array (aCGH).

Abbreviations: dup, duplication; del, deletion.

Reference

1. Bainbridge MN, Wang M, Wu Y, Newsham I, Muzny DM, Jefferies JL, Albert TJ, Burgess DL, Gibbs RA. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol 2011;**12**(7):R68 doi: 10.1186/gb-2011-12-7-r68.
2. Challis D, Yu J, Evani US, Jackson AR, Paithankar S, Coarfa C, Milosavljevic A, Gibbs RA, Yu F. An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics 2012;**13**:8 doi: 10.1186/1471-2105-13-8.
3. Reid JG, Carroll A, Veeraraghavan N, Dahdouli M, Sundquist A, English A, Bainbridge M, White S, Salerno W, Buhay C, Yu F, Muzny D, Daly R, Duyk G, Gibbs RA, Boerwinkle E. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. BMC Bioinformatics 2014;**15**:30 doi: 10.1186/1471-2105-15-30.
4. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G. The variant call format and VCFtools. Bioinformatics 2011;**27**(15):2156-8 doi: 10.1093/bioinformatics/btr330.
5. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;**38**(16):e164 doi: 10.1093/nar/gkq603.
6. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018;**46**(D1):D1062-D67 doi: 10.1093/nar/gkx1153.
7. OMIM®. Online Mendelian Inheritance in Man. Secondary Online Mendelian Inheritance in Man 2018. https://omim.org/.
8. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;**14**(2):178-92 doi: 10.1093/bib/bbs017.

9. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 2010;**6**(12):e1001025 doi: 10.1371/journal.pcbi.1001025.
10. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;**46**(3):310-5 doi: 10.1038/ng.2892.
11. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc 2016;**11**(1):1-9 doi: 10.1038/nprot.2015.123.
12. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods 2010;**7**(4):248-9 doi: 10.1038/nmeth0410-248.
13. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 2014;**11**(4):361-2 doi: 10.1038/nmeth.2890.
14. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;**17**(5):405-24 doi: 10.1038/gim.2015.30.
15. Gambin T, Akdemir ZC, Yuan B, Gu S, Chiang T, Carvalho CMB, Shaw C, Jhangiani S, Boone PM, Eldomery MK, Karaca E, Bayram Y, Stray-Pedersen A, Muzny D, Charng WL, Bahrambeigi V, Belmont JW, Boerwinkle E, Beaudet AL, Gibbs RA, Lupski JR. Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. Nucleic Acids Res 2017;**45**(4):1633-48 doi: 10.1093/nar/gkw1237.
16. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical

Genetics and Genomics. Genet Med 2017;**19**(2):249-55 doi: 10.1038/gim.2016.190.

17. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;**4**(1):44-57 doi: 10.1038/nprot.2008.211.
18. Fromer M, Purcell SM. Using XHMM Software to Detect Copy Number Variation in Whole-Exome Sequencing Data. Curr Protoc Hum Genet 2014;**81**:7 23 1-21 doi: 10.1002/0471142905.hg0723s81.
19. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, Project NES, Quinlan AR, Nickerson DA, Eichler EE. Copy number variation detection and genotyping from exome sequence data. Genome Res 2012;**22**(8):1525-32 doi: 10.1101/gr.138115.112.
20. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 2009;**84**(4):524-33 doi: 10.1016/j.ajhg.2009.03.010.
21. Breiman L. Random Forests. Machine Learning 2001;**45**(1):5-32 doi: 10.1023/a:1010933404324.

Supplementary Tables

Table S1. Demographic information of the American cohort

Proband ID	Ethnicity	Age at Scoliosis Onset	Gender	Clinical Features
EOS0001	NHW	8 months of age	Female	Treated in a Pavlik harness shortly after birth for some type of hip dysplasia, possibly a click. Still had some persistent acetabular dysplasia at age 14. Followed by cardiology for a possible valvular problem. Generalized ligamentous laxity; met 4 of the 5 Wynne-Davier criteria for hyperextensibility on clinical exam.
EOS0002	NHW	2 years old	Female	Normal MRI. Ligamentous laxity with positive thumb to wrist test, hyperextension at her elbows and knees, and dorsiflexion to her feet. Uses orthotics for her flexible flat feet. Incidental note of a cartilaginous cleft in the posterior element of L5 on x-ray. Noted to have a very thin body habitus.
EOS0003	AFR	3 years old	Male	Severe kyphoscoliosis. Normal MRI. Transitional vertebra at the lumbosacral junction. Significant back pain that wakes him from sleep most nights, occasional numbness and tingling of his fingers, bed wetting, and occasional accidental defecation. Followed by neurology for sleep dysregulation and headaches.
EOS0007	NHW	7 months old	Male	Delivered via C-section due to fetal distress. Flattening of the right occiput, slightly wide set eyes, and a very subtle flattening of the nasal bridge noted on clinical exam. MRI revealed no intracranial or intraspinal abnormality or congenital malformations.
EOS0020	NHW	6 months old	Female	Normal MRI. Small defect in the posterior elements of L5 noted on x-ray. 6 lumbar vertebrae.
EOS0022	SAS	1 year old	Male	Pectus carinatum and eczema noted on clinical exam. Normal MRI. Born via C-section due to maternal bradycardia. Mild diffuse osteopenia noted on one x-ray but this was not replicated on later x-rays.
EOS0025	NHW	3 years old	Female	Born via C-section due to transverse presentation. MRI showed no intrathecal abnormalities or congenital malformations. Radiopaque density in the region of the ductus arteriosus noted on several x-rays.
EOS0031	NHW	3 years old	Female	Born via C-section secondary to breech presentation with her twin positioned on top of her. Spent some time in the NICU. Experienced some delay in growth as well as in sensory and motor development for which she underwent therapy. Mother had gestational diabetes during pregnancy.
EOS0034	NHW	2 years old	Male	Posterior scalloping of the lumbar vertebrae noted on x-ray. Normal MRI.
EOS0035	NHW	9 months old	Male	Fraternal twin who is a product of in vitro fertilization. Some plagiocephaly and facial asymmetry with a large, high arching palate and palmar crease noted on clinical exam. This was evaluated by a geneticist who thought that the facial asymmetry was mostly positional and did not think there was an underlying syndrome driving his scoliosis. The geneticist did observe a triangular shaped face with slightly low-set ears. Normal MRI and eye exam. On x-ray, a sclerotic lesion was noted along the medial aspect of the right lesser trochanter that was thought to be a nonossifying fibroma or large bone island.
EOS0038	NHW	1.5 years old	Female	Spent 10 days in the NICU after birth due to respiratory complications. Minor development delay. Normal MRI.
EOS0041	NHW	14 months of age	Male	Normal MRI and CT scans. Premature birth and mild developmental delay. Atrial septal defect (ASD) and ventricular septal defect (VSD) that closed spontaneously. Trivial mitral valve regurgitation. Tracheomalacia that has resolved. Hyperreflexia. Chiari 1 malformation that was treated by decompression. Pineal brain tumor that was surgically removed at eight years of age. Bilateral genu valgum of his lower extremities that developed around twelve years of age and was treated with 8 plates.
EOS0045	NHW	In utero	Male	Normal MRI. Born via C-section and is an identical twin. Observed in the NICU after birth for low body temperature. Demonstrated a wide based out-toeing gait on clinical exam. Multiple vertebral anomalies noted on x-ray at T4-5, T7-8, and L1-2. Identical twin has not developed scoliosis.

Abbreviations: NHW: Non-Hispanic White; AFR: African American; SAS: South Asian; AR: Autosomal recessive; AD: Autosomal dominant

Table S2. Capture platforms and quality information of exome sequencing in the Chinese cohort

Platform	No. of Cases	Depth (×)	20× coverage	Q20 (%)	Q30 (%)
Agilentv6UTR	96	108	0.97	96.6	92.2
IDT	14	187	0.99	96.1	92.5
NimbleGen	98	96	0.94	95.2	89.4
VCRome	462	91	0.93	95	89.6
Total/Average	**670**	**96**	**0.94**	**95.3**	**90**

Four different capture platforms were used for 670 subjects depending on the time they were enrolled. Depth refers to the mean coverage of exome sequencing. 20× coverage refers to the ratio of bases covered by more than 20 reads.

Table S3. Diagnostic yield in CS and syndromic EOS subtypes in the Chinese cohort

	CS			Syndromic EOS		
Subgroup	**CS I**	**CS II**	**CS III**	**NF1**	**ACH**	**EDS**
No. patients	130	41	253	4	3	1
Age of onset, mean	2.6y	3.5y	3.2y	3.2y	2.7y	1y
Male (%)	73/130 (56.2%)	16/41 (39%)	114/253 (45.0%)	2/4	3/3	1/1
TACS rate (%)	35/130 (26.9%)	0	6/253 (2.4%)	0	0	0
SNV/indel positive rate (%)	10/130 (7.7%)	6/41 (15%)	18/253 (7.1%)	3/4	3/3	1/1
CNV positive rate (%)	0	1/41 (2%)	3/253 (1.2%)	1/4	0	0
Overall diagnostic rate (%)	**45/130 (34.6%)**	**7/41 (17%)**	**27/253 (10.7%)**	**4/4**	**3/3**	**1/1**

Abbreviations: CS I, congenital scoliosis type I, vertebral malformations; CS II, congenital scoliosis type II, segmentation defects; CS III, congenital scoliosis type II, mixed type; NMS, neuromuscular scoliosis; NF1, neurofibromatosis type 1; ACH, achondroplasia; EDS, Ehlers-Danlos syndrome; TACS, TBX6-associated scoliosis; SNV, single nucleotide variant; CNV, copy number variation.

21 **Table S4. Summary of *TBX6* compound variants and patients' phenotype**

Sample ID	Verification Method	*TBX6* null allele	Haplotype *in trans*	Clinical diagnose	Skeletal Phenotype	Multi-system phenotype
XH286	Haplotyping/Sanger sequencing	c.1169dupC(p.His391AlafsTer96)	T-C-A	CS I	T8: segmented butterfly vertebra; L2: non-segmented hemivertebra at left side; L3: segmentation defect vertebra fused up to the vertebra above; scoliosis at left side;	N
XH101	Haplotyping/Sanger sequencing	c.1250dupT(p.Leu419SerfsTer68)	T-C-A	CS I	L2: hemivertebra at left side; R12: absent ribs at both sides; scoliosis at left side	Physiologic jaundice
XH1062	Sanger sequencing	c.621+1G>A	T-C-A	CS I	L1-L2: addition segmented hemivertebra at right side; scoliosis at right side;	N
XH122	Haplotyping/Sanger sequencing	c.266dupC(p.Val91GlyfsTer80)	T-C-A	CS I	T12: segmented hemivertebra at left side; R12: an absent rib at right side; L5-S1: addition vertebral; scoliosis at right side	Increased size of nasopharyngeal adenoids
XH799	Sanger sequencing	c.1061delC(p.Ala354GlyfsTer144)	T-C-A	CS I	L3: lamina fissure; L4: absent pedicle; scoliosis at left side	N
XH341	Sanger sequencing	c.1179_1180delAG(p.Gly395LeufsTer91)	T-C-A	CS I	T12: segmented hemivertebra at right side; R12: an absent rib at left side; scoliosis at right side	Hyperpigmentation
XH138	Haplotyping/Sanger sequencing	c.844C>T(p.Arg282Ter)	T-C-A	CS I	T10-T11: addition segmented hemivertebra at left side; R10: an additional rib at left side; L3: non-segmented hemivertebra at right side; L4: segmentation defect vertebra fused up to the vertebra above; scoliosis at right side; enlargement of the central canal;	N
XH1161	Sanger sequencing	c.1169dupC(p.His391AlafsTer96)	T-C-A	CS I	T12-L1: addition segmented hemivertebra at left side; scoliosis at left side;	N
XH170	Haplotyping/Sanger sequencing	c.704dupG(p.Met236HisfsTer44)	T-C-A	CS I	T12: segmented hemivertebra at left side; R12: an absent rib at left side; spinal bifida	N
XH586	Sanger sequencing	c.621+1G>C	T-C-A	CS I	T11: segmented hemivertebra at left side; R11: an absent rib at left side; scoliosis at left side; kyphosis	N
XH463	qPCR	16p11.2 deletion	T-C-A	CS I	L3: segmented hemivertebra at right side;	Hypertelorism; high arched palate; short fifth metacarpal
XH672	Array CGH	16p11.2 deletion	T-C-A	CS I	T5 and T6: fused lamina; T7: failure of formation; T8 and T11: segmented hemivertebra at right side: R8: an absent rib at left side; T9 and T10: failure of formation at left side; R11: an absent rib at left side; L1: segmented hemivertebra at left side and lamina fissure; L2 and L3: lamina fissure; scoliosis at right side	Mediastinum skewed to the right
XH785	Array-CGH	16p11.2 deletion	T-C-A	CS I	T6: fused lamina; L2: segmented hemivertebra at left side; L5 and S1: bifid spinel; scoliosis at left side; fat deposition in sacral canal	Delayed speech and language development
XH834	Array CGH	16p11.2 deletion	T-C-A	CS I	T8, L1 and L3: segmented butterfly vertebra; T10, T12 and L2: failure of formation; T11: segmented hemivertebra at right side; scoliosis at right side	Mental retardation; delayed speech and language development; motor deterioration
XH843	Array CGH	16p11.2 deletion	T-C-A	CS I	Chest deformity; T9 and T11: failure of formation; T10: segmented butterfly vertebra at left side; R11: an absent rib at right side; L3, L4 and L5: lamina fissure; scoliosis at left side	Hernia
XH402	Array CGH	16p11.2 deletion	T-C-A	CS III	C5: segmented hemivertebra at left side; T4: segmentation defect vertebra fused down to the vertebra below; T5: non-segmented hemivertebra with fusion both to the vertebrae above and below at right side; R5: an absent rib at left side; T6: segmentation defect vertebra fused up to the vertebra above; T7: segmentation defect vertebra fused down to the vertebra below; T8: non-segmented hemivertebra with fusion both to the vertebrae above and below at left side; R8: an absent rib at right side; L1-L2: failure of formation; split cervical spinal processes; scoliosis at right side;	N
XH25	qPCR	16p11.2 deletion	T-C-A	CS I	T11: segmented hemivertebra at left side; R11: an absent rib at right side; scoliosis at left side	N
XH140	Array-CGH	16p11.2 deletion	T-C-A	CS I	R10-R11: a fused and bifid rib with fusion at left side; T11: segmented hemivertebra at left side; R11: an absent rib at right side; T12: segmented wedged vertebra at left side; lamina fissure; hexadactyly; scoliosis at left side; kyphosis	N
XH149	Array CGH	16p11.2 deletion	T-C-A	CS I	T12: segmented hemivertebra at right side; R12: an absent rib at left side; scoliosis at right side	N
XH186	Array CGH	16p11.2 deletion	T-C-A	CS I	T11: segmented hemivertebra at left side; R11: an absent rib at right side; scoliosis at left side	N
XH228	Array CGH	16p11.2 deletion	T-C-A	CS I	T11: segmented hemivertebra at left side; R11: an absent rib at right side; scoliosis at left side	N

XH237	Array CGH	16p11.2 deletion	T-C-A	CS I	T6: segmentation defect vertebra fused up to the vertebra above; T12: segmented hemivertebra at left side; R12: an absent rib at right side; L3: segmented butterfly vertebra; pelvic asymmetry; scoliosis at left side; kyphosis;	Hyperthyroidism
XH259	Array CGH	16p11.2 deletion	T-C-A	CS III	T10: segmented hemivertebra at left side; T12: non-segmented hemivertebra with fusion at left side; R12: an absent rib at right side; L1: non-segmented hemivertebra with fusion up to the vertebra above; scoliosis at left side kyphosis	N
XH270	Array CGH	16p11.2 deletion	T-C-A	CS III	R3: bifid rib at left side; T11 and L1: segmented wedged vertebra at dorsal side; T12: segmented hemivertebra at dorsal side; R12: an absent rib at right side; pelvic tilt; scoliosis at right side; kyphosis	Pneumonia with decreased FEV1
XH468	Array CGH	16p11.2 deletion	T-C-A	CS I	T12: segmentation defect vertebra fusion at left side; scoliosis at left side	Anemia; thrombocytopenia
XH494	Array CGH	16p11.2 deletion	T-C-A	CS I	C2: deformation; L4-L5: an additional non-segmented hemivertebra with fusion at right side; scoliosis at right side	Maxillary and ethmoidal sinusitis
XH580	Array CGH	16p11.2 deletion	T-C-A	CS I	T10: segmented wedged vertebra at right side; T11: segmented hemivertebra at dorsal side; T12: segmented wedged vertebra at right side; scoliosis at right side; kyphosis	N
XH605	Array CGH	16p11.2 deletion	T-C-A	CS I	T12: segmented hemivertebra at right side; R12: an absent rib at left side; lumbarization of sacrum; scoliosis at right side	N
XH636	Array CGH	16p11.2 deletion	T-C-A	CS III	L1: non-segmented butterfly vertebra; L2: segmentation defect vertebra fused up to the vertebra above; scoliosis at right side; kyphosis	N
XH693	Array CGH	16p11.2 deletion	T-C-A	CS I	L1: segmented butterfly vertebra; scoliosis at left side	N
XH480	Array CGH	16p11.2 deletion	T-C-A	CS III	T9: non-segmented wedged vertebra with fusion at right side; T10: non-segmented hemivertebra with fusion at right side; R10: an absent rib at left side; T11: segmented wedged vertebra at right side; lumbarization of sacrum; scoliosis at right side	N
XH522	Array CGH	16p11.2 deletion	T-C-A	CS I	T10: segmented hemivertebra at right side; R10: an absent rib at left side; scoliosis at right side	N
XH529	Array CGH	16p11.2 deletion	T-C-A	CS I	T12: segmented hemivertebra at dorsal side; scoliosis at left side; sacrum bifida occulta	N
XH930	Array CGH	16p11.2 deletion	T-C-A	CS I	T7: segmented butterfly vertebra; T9 and T12: segmented hemivertebra at left side; R9 and R12: an absent rib at right side; scoliosis at left side	Sinus tachycardia
XH330	Array CGH	16p11.2 deletion	T-C-A	CS I	T10: segmented hemivertebra at left side; R10: an absent rib at left side; T12: segmented hemivertebra at right side; R12: an absent rib at left side; scoliosis at right side	N
XH978	Array-CGH	16p11.2 deletion	T-C-A	CS I	T4: segmented hemivertebra at left side; R4: an absent rib at right side; T12: segmented hemivertebra at right side; R12: an absent rib at left side; scoliosis at right side; ankle clonus	N
XH678	Array CGH	16p11.2 deletion	T-C-A	CS I	T11: segmented butterfly vertebra; scoliosis at right side; kyphosis	N
XH1154	Array-CGH	16p11.2 deletion	T-C-A	CS III	T9: segmentation defect vertebra fused down to the vertebra below; T10: non-segmented hemivertebra with fusion at left side; T11: segmentation defect vertebra fused up to the vertebra above; scoliosis at left side	N
XH1134	Array-CGH	16p11.2 deletion	T-C-A	CS I	T7, T8 and T9: segmented wedged vertebra at right side; scoliosis at right side	N
XH265	Array CGH	16p11.2 deletion	T-C-A	CS I	L1: segmented hemivertebra at left side	N
XH138	Array CGH	16p11.2 deletion	T-C-A	CS I	L3: non-segmented hemivertebra; L3-L4 segmentation defect	N

22 Abbreviations: ddPCR, digital droplet PCR; WGS, whole genome sequencing; CS I, congenital scoliosis type I, vertebral malformations; CS II, congenital scoliosis type II, segmentation defects); CS III, congenital scoliosis type II

23

24 **Table S5. Summary of identified pathogenic/ likely pathogenic variants and patients' phenotype in the Chinese cohort**

Family ID	Case ID	Onset-Age of Scoliosis	Gender	Clinical diagnose	Musculoskeletal phenotype	Systematic phenotype	Gene	Variant	Variant type	Zygosity	Origin	Inheritance	Classification	Molecular diagnosis
F16880004	XHO2	9	M	CS I	Congenital scoliosis; cubitus valgus	N	*FLNB*	c.4756G>A (p.Gly1586Arg)	Missense	Het	P	AD	P	Larsen syndrome
F16880005	XHO1	2	F	NMS	Scoliosis; short stature; muscular hypotonia	Growth retardation; nephroptosis	*RYR1*	c.14423T>A (p.Phe4808Tyr)	Missense	Het	P	AD	LP	Central core disease
F170700002	XH1	1	F	CS I	T11 and T12: segmented wedged vertebra at left side; scoliosis at left side	N	*DDR2*	c.779G>A(p.Arg260Gln)	Missense	Het	P	AR	LP	Spondylometaepiphyseal dysplasia
F170700002	XH1	1	F	CS I	T11 and T12: segmented wedged vertebra at left side; scoliosis at left side	N	*DDR2*	c.932G>A(p.Ser311Asn)	Missense	Het	M	AR	LP	Spondylometaepiphyseal dysplasia
F171000013	XH15	0	M	CS III	T4-T7: nonsegmented wedge vertebrae; T12: segmental defect	N	*CHD7*	c.5051G>A(p.Gly1684Asp)	Missense	Het	*de novo*	AD	LP	CHARGE syndrome
F170700007	XH193	3	F	CS II	Chest deformity; C3 and T5: block vertebrae fused down to the vertebra below; C4: block vertebra fused both to the vertebrae above and below; C5: block vertebra fused up to the vertebra above; T6-T10: segmentation defect vertebra fused both to the vertebrae above and below; T11: segmentation defect vertebra fused up to the vertebra above; scoliosis at right side; lordosis	Mitral valve prolapse; hypoventilation	*FLNB*	c.839_854delACACCATCAGCGCCGG (p.Asp280GlyfsTer5)	Frameshift	Het	P	AR	P	Spondylocarpotarsal synostosis syndrome
F170700007	XH193	3	F	CS II	Chest deformity; C3 and T5: block vertebrae fused down to the vertebra below; C4: block vertebra fused both to the vertebrae above and below; C5: block vertebra fused up to the vertebra above; T6-T10: segmentation defect vertebra fused both to the vertebrae above and below; T11: segmentation defect vertebra fused up to the vertebra above; scoliosis at right side; lordosis	Mitral valve prolapse; hypoventilation	*FLNB*	c.6763C>T(p.Gln2255Ter)	Nonsense	Het	M	AR	P	Spondylocarpotarsal synostosis syndrome
F170800005	XH475	0	M	CS II	Chest deformity; T3: segmentation defect vertebra fused down to the vertebra below; T4: segmentation defect vertebra fused both to the vertebrae above and below; R4 and R8: absent ribs at left side; T7: segmentation defect vertebra fused down to the vertebra below; T8: block vertebra fused both to the vertebrae above and below; T9: block vertebra fused up to the vertebra above; L5: missing vertebra; scoliosis at right side	Arthrogryposis of hands	*MYH3*	c.1400A>C(p.Glu467Ala)	Missense	Het	*de novo*	AD	LP	Arthrogryposis, distal, type 8
F170800025	XH657	1	M	CS III	T6: block vertebra fused down to the vertebra below; T7: non-segmented wedged vertebra with fusion up to the vertebra above at left side; high scapula; arthrogryposis multiplex congenita; scoliosis at left side; kyphosis	Cleft palate; hernia of the abdominal wall	*MYH3*	c.841G>A(p.Glu281Lys)	Missense	Het	*de novo*	AD	LP	Arthrogryposis, distal, type 8
F170800047	XH1077	1	M	CS II	T1-T12: segmentation defect; scoliosis at left side	Mitral insufficiency; mitral valve prolapse; renal cysts	*B3GALT6*	c.513_520delCGCCCGCG(p.Glu174AlafsTer266)	Frameshift	Het	P	AR	P	Spondyloepimetaphyseal dysplasia with joint laxity
F170800047	XH1077	1	M	CS II	T1-T12: segmentation defect; scoliosis at left side	Mitral insufficiency; mitral valve prolapse; renal cysts	*B3GALT6*	c.694C>T(p.Arg232Cys)	Missense	Het	M	AR	LP	Spondyloepimetaphyseal dysplasia with joint laxity
F170800053	XH1049	4	M	NMS	Skeletal muscle fatty infiltration; muscle weakness; scoliosis at right side; multiple minicore myopathy	Restrictive ventilatory defect	*TTN*	c.43799G>A(p.Trp14600Ter)	Nonsense	Het	P	AR	P	Myopathy
F170800053	XH1049	4	M	NMS	Skeletal muscle fatty infiltration; muscle weakness; scoliosis at right side; multiple minicore myopathy; restrictive ventilatory defect	Restrictive ventilatory defect	*TTN*	c.21433_21444delCTTCAAGTGTCT(p.Leu7145_Ser7148del)	Inframe-deletion	Het	M	AR	LP	Myopathy
F170800056	XH1000	8	M	ACH	Disproportionate short-limb short stature; anomaly of cartilage; abnormal form of the vertebral bodies; spinal canal stenosis; cubitus valgus; scoliosis at left side; kyphosis; claudication	Hydrocephalus	*FGFR3*	c.1138G>A(p.Gly380Arg)	Missense	Het	*de novo*	AD	P	Achondroplasia,Hypochondroplasia
F170800059	XH488	0	F	NFS	Scoliosis at right side	Cafe-au-lait spot	*NF1*	c.4290T>G(p.Asn1430Lys)	Missense	Het	*de novo*	AD	P	Neurofibromatosis, type 1
F170900002	XH200	1	F	CS III	Chest deformity; T1: block vertebra fused down to the vertebra below; T2 and T3: failure of formation; T4-T10: block vertebra; scoliosis at right side	N	*FLNB*	c.163G>T(p.Glu55Ter)	Nonsense	Het	P	AR	P	Spondylocarpotarsal synostosis syndrome
F170900002	XH200	1	F	CS III	Chest deformity; T1: block vertebra fused down to the vertebra below; T2 and T3: failure of formation; T4-T10: block vertebra; scoliosis at right side	N	*FLNB*	c.4671G>A(p.Thr1557=)	Splicing	Het	M	AR	LP	Spondylocarpotarsal synostosis syndrome
F171000015	XH921	8	F	CS I	Scoliosis at left side; kyphosis; developmental dysplasia of the hip; tibia valga	Cafe-au-lait spots	*FLNA*	c.3571G>A(p.Ala1191Thr)	Missense	Het	*de novo*	XL	LP	Terminal osseous dysplasia
F171000016	XH30	1	M	CS II	T4 and T12: segmentation defect vertebra fused down to the vertebra below; T5 and L2: segmentation defect vertebra fused up to the vertebra above; L1: segmentation defect vertebra fused both to the vertebrae above and below	N	*RYR1*	c.14761TT>AC(p.Phe4921Thr)	Missense	Het	*de novo*	AD	P	Central core disease

F171000017	XH33	3	M	CS III	C7: non-segmented hemivertebra with fusion to the vertebra at left side; T1 and T2: segmentation defect vertebra fused both to the vertebrae above and below; T3: segmentation defect vertebra fused up to the vertebra above; diastematomyelia	Mitral valve prolapse; restrictive ventilatory defect	*CELSR1*	c.2609C>T(p.Pro870Leu)	Missense	Het	P	AD	P	Neural tube defects
F171000020	XH45	5	M	CS III	T10: segmentation defect vertebra fused down to the vertebra below; T11 and T12: segmentation defect vertebra fused both to the vertebrae above and below; L1: non-segmented hemivertebra with fusion down to the vertebra; L2: segmentation defect vertebra fused up to the vertebra above; diastematomyelia; tethered cord	N	*HERC1*	c.5182A>G(p.Ile1728Val)	Missense	Het	P	AR	LP	Macrocephaly, dysmorphic facies, and psychomotor retardation
F171000020	XH45	5	M	CS III	T10: segmentation defect vertebra fused down to the vertebra below; T11 and T12: segmentation defect vertebra fused both to the vertebrae above and below; L1: non-segmented hemivertebra with fusion down to the vertebra; L2: segmentation defect vertebra fused up to the vertebra above; scoliosis at left side; kyphosis; diastematomyelia; tethered cord	N	*HERC1*	c.764C>G(p.Ala255Gly)	Missense	Het	M	AR	LP	Macrocephaly, dysmorphic facies, and psychomotor retardation
F171000031	XH95	7	F	CS III	T7: segmentation defect vertebra fused down to the vertebra below; T8: non-segmented hemivertebra with fusion at right side; R8: an absent rib at left side; atrial septal defect; T9: segmentation defect vertebra fused up to the vertebra above; scoliosis at right side	Atrial septal defect	*ZC4H2*	c.437C>G(p.Pro146Arg)	Missense	Het	*de novo*	XL	LP	Wieacker-Wolff syndrome
F171000042	XHO95	0	M	CS II	C4-C6: Segmentation defect and fused lamina; short neck	Cardiac hypertrophy	*MYH7*	c.2011C>T(p.Arg671Cys)	Missense	Het	*de novo*	AD	P	Cardiomyopathy, hypertrophic
F171100021	XH4	1	M	CS I	Chest deformity; T6-T10: failure formation; R12: thin ribs at both sides; bilateral abnormalities of the clavicles; midclavicular aplasia; scoliosis at right side	Renal aplasia	*RUNX2*	c.516delT(p.Ala173LeufsTer3)	Frameshift	Het	*de novo*	AD	P	Cleidocranial dysplasia
F171100126	XH1186	0	F	CS I	T5-T7 and T10: segmented butterfly vertebra; motor deterioration	Patent foramen ovale; hypoplasia affecting the eye; inguinal hernia	*JAG1*	c.2998_3002delATCGC(p.Ile1000LeufsTer10)	Frameshift	Het	P	AD	P	Alagille syndrome 1
F171200061	XH1193	0	M	CS III	T1-T9 and L2-S1: failure of formation; T10-L1: segmentation defect, failure of formation; genu valgum; pes varus; scoliosis at left side	N	*TRPV4*	c.1781G>A(p.Arg594His)	Missense	Het	P	AD	P	Spondylometaphyseal dysplasia, Kozlowski type
F171200138	XH1234	1	M	CS I	T4 and T6: segmented hemivertebrae at right side; R4 and R6: absent ribs at left side; scoliosis at right side	N	*DCHS1*	c.3025C>A(p.Leu1009Met)	Missense	Het	M	AR	LP	Van Maldergem syndrome 1
F171200138	XH1234	1	M	CS I	T4 and T6: segmented hemivertebrae at right side; R4 and R6: absent ribs at left side; scoliosis at right side	N	*DCHS1*	c.8890C>T(p.Arg2964Cys)	Missense	Het	P	AR	LP	Van Maldergem syndrome 1
F171200141	XH20	5	M	NFS	T11: segmented wedged vertebra at right side; dysplasia of clavicles; scoliosis at right side; kyphosis	Cafe-au-lait spots	*NF1*	c.383dupA(p.Asn128LysfsTer6)	Frameshift	Het	P	AD	P	Neurofibromatosis, type 1
F171200160	XHO4	2	F	IEOS	Idiopathic Scoliosis	N	*COMP*	c.1360_1361delAG(p.Ser454CysfsTer12)	Frameshift	Het	M	AD	P	Epiphyseal dysplasia, multiple, type 1
F171200161	XH12	4	F	NMS	Sternocleidomastoid amyotrophy; restricted neck movement due to contractures; rectus femoris muscle atrophy; scoliosis at right side	Mitral valve prolapse; mitral insufficiency; restrictive ventilatory defect	*LMNA*	c.1357C>T(p.Arg453Trp)	Missense	Het	*de novo*	AD	P	Emery-Dreifuss muscular dystrophy 2
F180100030	XH23	0	M	NFS	Chest deformity; T1-T5: segmentation defect and failure of formation; scoliosis at right side; kyphosis	Cafe-au-lait spots	*NF1*	c.6755A>G(p.Lys2252Arg)	Missense	Het	*de novo*	AD	P	Neurofibromatosis,Neurofibromatosis 1
F180200039	XH9	2	M	CS II	T10-L4: segmentation defect; scoliosis at right side	Intellectual disability	*POGZ*	c.1180_1181delAT(p.Met394ValfsTer9)	Frameshift	Het	*de novo*	AD	P	White-Sutton syndrome
F180200039	XH9	2	M	CS II	T10-L4: segmentation defect; scoliosis at right side	Intellectual disability	*FBN1*	c.2649G>A(p.Trp883Ter)	Nonsense	Het	M	AD	P	Marfan syndrome
F171000049	XHR24	1	M	CS III	C2-C7 segmentation defect; C3 nonsegmented wedged vertebra; Micrognathia; Limitation of joint mobility; Atlantoaxial abnormality	Abnormality of the myocardium;Cleft palate	*SOX9*	c.337A>C(P.Met113Leu)	Missense	Het	*de novo*	AD	P	Acampomelic dysplasia
F171000028	XH86	0	F	CS III	T11 segmentation defect; T12 non-segmented hemi-vertebra; hypotonia of both lower limbs; diastematomyelia; short stature	N	*COL27A1*	c.1807C>T(p.Arg603Trp)	Missense	Het	P	AR	LP	Steel syndrome
F171000028	XH86	0	F	CS III	T11 segmentation defect; T13 non-segmented hemi-vertebra; hypotonia of both lower limbs; diastematomyelia; short stature	N	*COL27A1*	c.1135A>T(p.Ile379Val)	Missense	Het	M	AR	LP	Steel syndrome
NA	XH26	5	M	CS I	T12: segmented hemivertebra at left side; R12: an absent rib at right side; scoliosis at left side	N	*COMP*	c.1120G>A(p.Asp374Asn)	Missense	Het	NA	AD	LP	Multiple epiphyseal dysplasia
NA	XH66	2	F	CS III	Chest deformity; C1-L5: flattened vertebrae; developmental dysplasia of the hip; coxa vara; osteogenesis imperfecta of the upper femora; spinal metaphysis dysplasia; elbow extension restricted; scoliosis at left side; kyphosis; dwarfism; barrel chest; intermittent claudication	N	*TRPV4*	c.1781G>A(p.Arg594His)	Missense	Het	NA	AD	P	Spondylometaphyseal dysplasia, Kozlowski type
NA	XH179	1	M	CS III	T9: segmentation defect vertebra fused down to the vertebra below; T10: non-segmented hemivertebra with fusion both to the vertebrae above and below at	Cryptorchidism; hydrocele testis	*TCF12*	c.1966C>T(p.Gln656Ter)	Nonsense	Het	NA	AD	P	Craniosynostosis 3

					right side; R10: an absent rib at right side; T11: non-segmented wedged vertebra with fusion up to the vertebra above at left side; scoliosis at left side								
NA	XH284	0	M	ACH	T12: segmented hemivertebra; bifid sacrum; kyphosis	Hydrocephalus; dwarfism	*FGFR3*	c.1138G>A(p.Gly380Arg)	Missense	Het	NA	AD	P	Achondroplasia, Hypochondroplasia
NA	XH352	3	F	CS I	T11 and T12: segmented wedged vertebra at left side; developmental dysplasia of the hip; scoliosis at left side; kyphosis	Right eye cataract; ichthyosis	*EBP*	c.23dupT(p.Leu8PhefsTer17)	Frameshift	Het	NA	XL	P	Chondrodysplasia punctate, X-linked dominant
NA	XH364	0	F	CS III	T4: segmented butterfly vertebra; T7: non-segmented hemivertebra with fusion down to the vertebra below at right side; R7: an absent rib at left side; T8: segmentation defect vertebra fused up to the vertebra above; scoliosis at right side; kyphosis	Atrial septal defect; ectopic right kidney	*PTPN11*	c.844A>G(p.Ile282Val)	Missense	Het	NA	AD	LP	Noonan syndrome 1
NA	XH603	1	F	IEOS	Scoliosis at left side	N	*COL5A2*	c.402+1G>C	Splicing	Het	NA	AD	P	Ehlers-Danlos syndrome, classic type 2
NA	XH696	0	F	CS II	Chest deformity; T9-L1: segmentation defect; R11: ribs absent at both sides; scoliosis at left side; lordosis; arthrogryposis multiplex congenita	Malignant hyperthermia; congenital bilateral ptosis; bilateral lagophthalmos	*RYR1*	c.7523G>A(p.Arg2508His)	Missense	Het	NA	AD	LP	Central core disease
NA	XH765	0	M	ACH	Cauda equina syndrome; L1-L3: segmented wedged vertebra with fusion down to the vertebra; scoliosis at right side; kyphosis	Dwarfism; ; renal cysts; gallbladder polyps; hydrocephalus	*FGFR3*	c.1138G>A(p.Gly380Arg)	Missense	Het	NA	AD	P	Achondroplasia, Hypochondroplasia
NA	XH565	1	F	CS III	T7: segmentation defect vertebra fused down to the vertebra below; T7-8: non-segmented hemivertebra with fusion both to the vertebrae at right side; T8: segmentation defect vertebra fused up to the vertebra above; scoliosis at right side	Tetralogy of Fallot; tricuspid valve regurgitation; accessory spleen	*COL11A1*	c.1245+1G>A	Splicing	Het	NA	AD	P	Stickler syndrome, type II
NA	XH874	6	M	CS I	T12 and L1: segmented wedged vertebra with fusion down to the vertebrae; kyphosis; sacral canal cyst	Hydronephrosis	*PLOD1*	c.644-1G>C	Splicing	Het	NA	AR	P	Ehlers-Danlos syndrome, type VI
NA	XH874	6	M	CS I	T12 and L1: segmented wedged vertebra with fusion down to the vertebrae; kyphosis; sacral canal cyst	Hydronephrosis	*PLOD1*	c.2008C>T(p.Arg670Ter)	Nonsense	Het	NA	AR	P	Ehlers-Danlos syndrome, type VI
NA	XH886	1	M	EDS	Genu valgum; scoliosis at left side; kyphosis;joint hypermobility	Right atrial isomerism; cryptorchidism; dilated fourth ventricle	*PLOD1*	c.1095C>T(p.Gly365=)	Splicing	Hom	NA	AR	P	Ehlers-Danlos syndrome, type VI
NA	XH893	7	F	CS III	Chest deformity; C5: segmentation defect vertebra fused down to the vertebra below; C6-T4: segmentation defect vertebra fused both to the vertebrae above and below, failure of formation; R3-R6: a fused and bifid rib with fusion at left side; T5: non-segmented wedged vertebra with fusion up to the vertebra above at left side; R12: an absent rib at left side; bilateral Hoffmann sign; scoliosis at right side	Patent ductus arteriosus; right upper pulmonary nodules	*CHD7*	c.8962delG(p.Asp2988MetfsTer3)	Frameshift	Het	NA	AD	P	CHARGE syndrome
NA	XH895	1	F	CS I	Chest deformity; hypoplasia of clavicles; scoliosis at right side; persistent open anterior fontanelle	Ocular hypertelorism; high arched palate	*RUNX2*	c.674G>A(p.Arg225Gln)	Missense	Het	NA	AD	P	Cleidocranial dysplasia
NA	XH1040	7	F	CS III	T2: segmentation defect vertebra fused down to the vertebra below; T3: non-segmented butterfly vertebra with fusion both to the vertebrae above and below; T4-T7: segmentation defect vertebra fused both to the vertebrae above and below; T8: segmentation defect vertebra fused up to the vertebra above; scoliosis at right side; diastematomyelia	Mitral valve prolapse; abnormality of refraction	*SUFU*	c.1447delC(p.Leu483TyrfsTer20)	Frameshift	Het	NA	AD	P	Basal cell nevus syndrome
NA	XH1129	3	F	CS III	Chest deformity; T5: segmentation defect vertebra fused down to the vertebra below, failure of formation; T6-T9: segmentation defect vertebra fused both to the vertebrae above and below, failure of formation; T10: segmentation defect vertebra fused up to the vertebra above, failure of formation; S1: spina bifida occulta; scoliosis at right side	Atrial septal defect; abnormality of refraction	*BMP2*	c.766delA(p.Ser256AlafsTer5)	Frameshift	Het	NA	AD	P	Short stature, facial dysmorphism, and skeletal anomalies with cardiac anomalies
NA	XH1139	0	F	CS III	Kyphosis; upper extremity joint dislocation; herniation of intervertebral nuclei	Prominent umbilicus	*COL5A2*	c.1401G>A(p.Pro467=)	Splicing	Het	NA	AD	LP	Ehlers-Danlos syndrome
NA	XH1156	0	F	CS III	T5, T7 and T8: butterfly vertebra; T7-T8: segmentation defect; scoliosis at left side; lower limb asymmetry	N	*MMP13*	c.1415_1416delAA(p.Ter472CysfsTer9)	Frameshift	Het	NA	AD	P	Spondyloepimetaphyseal dysplasia, Missouri type

Abbreviations: CS I, congenital scoliosis type I, vertebral malformations; CS II, congenital scoliosis type II, segmentation defects; CS III, congenital scoliosis type II, mixed type; NMS, neuromuscular scoliosis; NFS, neurofibromatosis; ACH, achondroplasia; EDS, Ehlers-Danlos syndrome; IEOS, idiopathic early-onset scoliosis; M, male; F, female; AD, autosomal dominant; AR, autosomal recessive; P, pathogenic; LP, likely pathogenic; N, None; NA, not applicable

Table S6. Summary of pathogenic CNVs and patient phenotype

Patient No.	Sex	Age of presentation	Phenotype	CNV region (hg19)	Reported associated phenotypes	Involved gene	Variant type
XH66	F	2	CS III, kyphoscoliosis, wedge vertebrae in L1-L5, incomplete osteogenesis, dysplastic spine metaphysis, dysplasia of double hips, introverted dwarf	16p13.1	Associated with aortic dissections, scoliosis, schizophrenia as well as autism	*C16orf45/KIAA0430/NDE1/MYH11/FOPNL/ABCC1ABCC6*	duplication
XH187	F	3	CS II, SD between T6/T7 & T8/T9 & T9/T10, clubfeet	16p13.1	Associated with aortic dissections, scoliosis, schizophrenia as well as autism	*C16orf45/KIAA0430/NDE1/MYH11/FOPNL/ABCC1ABCC6*	duplication
XH821	M	10	NFS, Cafe-au-lait spots, left inguinal hernia	17q11.2	Recurrent deletion including NF1 gene, neurofibromatosis type I characterized by cafe-au-lait spots, Lisch nodules in the eye, and fibromatous tumors of the skin	*CRLF3/ATAD5/TEFM/ADAP2/RNF135/NF1/OMG/EVI2B/EVI2A/RAB11FIP4/COPRS/UTP6/SUZ12/LRRC37B*	deletion
XH238	F	10	CS III, segmentation defects in T2/T3 & T5/T6 & T7/T8, wedge vertebrae in T2 & T6 &T7, diastematomyelia, restrictive ventilation dysfunction	22q11.2	22q11.2 deletion syndrome with a range of congenital malformations	*PI4KA/SERPIND1/SNAP29/CRKL/LZTR1/THAP7/SLC7A4/POM121L7/P2RX6/BCRP2/GGT2*	deletion
XH226	F	0	CS III, wedge vertebrae in T6 & T7 & T8 & T9, SD between T11-T12, left 3rd-4th ribs fusion, hydrocephalus, PDA, ASD	5q35.3	Recurrent deletion including NSD1, Sotos syndrome-1 (SOTOS1)	*FGFR4/NSD1/RAB24/PRELID1/MXD3/LMAN2*	deletion

Abbreviations: F, female; M, male; CS I, congenital scoliosis type I, vertebral malformations; CS II, congenital scoliosis type II, segmentation defects); CS III, congenital scoliosis type II, mixed type); NFS, neurofibromatosis.

Table S7. Clinically relevant secondary findings

Sample ID	Clinical diagnosis	Inheritance	Zygosity	CHR	POS	Variant_type	Gene_name	Disease	HGVS_nomenclature	Nucleotide change	Amino acid change	gnomAD_frequency	ExAC_AF	Reference	Origin
XH574	CS III	Autosomal dominant	Het	17	41203079	splice_donor	*BRCA1*	Breast-ovarian cancer, familial, 1	NM_007298.3	c.2020+1G>C	N	0	0	1108453; 7163244	NA
XH66	IS	Autosomal dominant	Het	17	41197766	frameshift	*BRCA1*	Breast-ovarian cancer, familial, 1	NM_007298.3	c.2209delA	p.Ser737ValfsTer2	0	0	23522120	NA
XH649	CS III	Autosomal dominant	Het	13	32930691	frameshift	*BRCA2*	Breast-ovarian cancer, familial, 2	NM_000059.3	c.7567_7568delCT	p.Leu2523GlufsTer15	3.23E-05	0	NA	NA
XH598	CS III	Autosomal dominant	Het	5	112151233	missense	*APC*	Thyroid cancer, non-medullary	NM_000038.5	c.876G>C	p.Leu292Phe	0	0	26530882	NA
XH663	CS III	Autosomal dominant	Het	10	43610044	missense	*RET*	Thyroid carcinoma, medullary	NM_020975.4	c.1996A>G	p.Lys666Glu	0	0	15858153	NA
XH1134	CS I	Autosomal dominant	Het	16	2132436	splice_acceptor	*TSC2*	Tuberous sclerosis-2	NM_000548.3	c.3815-1G>A	N	0	0	NA	NA
XH351	CS III	Autosomal dominant	Het	19	11224300	stop_gained	*LDLR*	Hypercholesterolaemia	NM_000527.4	c.1448G>A	p.Trp483Ter	0	0	7903864	Maternal
XH242	CS III	Autosomal dominant	Het	19	11230807	frameshift_variant	*LDLR*	Hypercholesterolaemia	NM_000527.4	c.1885_1886insCA	p.Phe629SerfsTer37	0	0	22883975	NA
XH206	CS III	Autosomal dominant	Het	19	38991282	missense_variant	*RYR1*	Malignant hyperthermia	NM_000540.2	c.7360C>T	p.Arg2454Cys	0	0.0000083	10484775	NA
XH984	CS III	Autosomal dominant	Het	7	44185217	missense_variant	*GCK*	Diabetes mellitus	NM_000162.3	c.1132G>C	p.Ala378Pro	0	0	1696533; 122035290	NA

Abbreviations: F, female; M, male; CS I, congenital scoliosis type I, vertebral malformations; CS II, congenital scoliosis type II, segmentation defects); CS III, congenital scoliosis type II, mixed type); IEOS, idiopathic early-onset scoliosis. VAF, variant allele frequency.

Table S8. Dual diagnosis resulting in blended phenotype in two cases

Case ID	Clinical diagnosis	Detailed phenotype	Genetic/Genomic changes	Inheritance	Associated disease	Reported phenotype
XH66	CS III	Flattened vertebrae	NM_021625.4(*TRPV4*):c.1781G>A(p.Arg594His)	U	Spondylometaphyseal dysplasia, Kozlowski type	**Spondyloepiphyseal dysplasia, Platyspondyly**
		Scoliosis, developmental dysplasia of the hip, dwarfism	16p13.1 Duplication	U	NA	Attention deficit/hyperactivity disorder, skeletal hypermobility, craniosynostosis and polydactyly
XHM9	CS II	Intellectual disability	NM_015100.3(*POGZ*):c.1180_1181delAT(p.Met394ValfsTer9)	*de novo*	White-Sutton syndrome	**Short stature, Joint laxity, Brachydactyly, delayed psychomotor development**
		Scoliosis	NM_000138.4(*FBN1*):c.2649G>A(p.Trp883Ter)	Maternal	Marfan syndrome	Tall stature, scoliosis, Flexible joints, heart problems

Detailed phenotypes of these two patients were categorized into each causal gene involving their dual diagnosis. Abbreviations: CS II, congenital scoliosis type II, segmentation defects); CS III, congenital scoliosis type II, mixed type; U, unknown; NA, not applicable.

Table S9. Relevance of selected clinical features ranked by random forest

Clinical feature	MRR
Age of presentation	0.53
Number of organ systems involved*	0.5
Cobb angle	0.49
Presence and degree of respiratory disturbance	0.45
Presence of intraspinal defects	0.33
Presence of vertebral malformation	0.3
Presence of chest deformity	0.26
History of muscular complications	0.17
Number of curvatures	0.12
Gender	0.11
Presence of kyphosis	0.1
Presence of lordosis	0.09

Abbreviations: MRR, mean reciprocal rank, which demonstrates the degree of relevance between that feature and each of the patient groups (the more relevant, the higher the rank score)

*organ systems include defects in ocular, urinary, cardiovascular, and digestive systems.

Table S10. Phenotypic comparison among three diagnostic groups

	Undiagnosed	TACS	P-value	Diagnosed nut not TACS	P-value
Age of presentation, y	3.33±3.17	1.56±2.17	**0.001**	2.50±2.52	**0.038**
Number of organ systems involved	0.34±0.59	0.07±0.26	**0.002**	0.54±0.68	**0.026**
Cobb angle, °	58.82±28.09	51.41±19.85	**0.032**	57.36±20.37	0.655

Abbreviations: TACS, TBX6-associated scoliosis, SD standard deviation. P-value was calculated using Student's t-test (for the Cobb angle) and Wilcoxon Signed Rank Test (for the age of presentation and the number of organ systems involved).