






Article

Identification of Copy Number Variants in a Southern Chinese Cohort of Patients with Congenital Scoliosis

Wenjing Lai ^{1,†}, Xin Feng ^{1,†}, Ming Yue ¹ , Prudence W. H. Cheung ² , Vanessa N. T. Choi ¹, You-Qiang Song ¹ , Keith D. K. Luk ², Jason Pui Yin Cheung ^{2,*}  and Bo Gao ^{1,*} 

¹ School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; floralai@connect.hku.hk (W.L.); FengxinCassie@connect.hku.hk (X.F.); nervym7@connect.hku.hk (M.Y.); vntchoi@hku.hk (V.N.T.C.); songy@hku.hk (Y.-Q.S.)

² Department of Orthopaedics and Traumatology, The University of Hong Kong, Hong Kong, China; gnuhchp6@hotmail.com (P.W.H.C.); hrmoldk@hku.hk (K.D.K.L.)

* Correspondence: cheungjp@hku.hk (J.P.Y.C.); gaobo@hku.hk (B.G.); Tel.: +852-2255-4581 (J.P.Y.C.); +852-3917-6809 (B.G.)

† These authors contribute equally to this work.

Abstract: Congenital scoliosis (CS) is a lateral curvature of the spine resulting from congenital vertebral malformations (CVMs) and affects 0.5–1/1000 live births. The copy number variant (CNV) at chromosome 16p11.2 has been implicated in CVMs and recent studies identified a compound heterozygosity of 16p11.2 microdeletion and *TBX6* variant/haplotype causing CS in multiple cohorts, which explains about 5–10% of the affected cases. Here, we studied the genetic etiology of CS by analyzing CNVs in a cohort of 67 patients with congenital hemivertebrae and 125 family controls. We employed both candidate gene and family-based approaches to filter CNVs called from whole exome sequencing data. This identified 12 CNVs in four scoliosis-associated genes (*TBX6*, *NOTCH2*, *DSCAM*, and *SNTG1*) as well as eight recessive and 64 novel rare CNVs in 15 additional genes. Some candidates, such as *DHX40*, *NBPF20*, *RASA2*, and *MYSM1*, have been found to be associated with syndromes with scoliosis or implicated in bone/spine development. In particular, the *MYSM1* mutant mouse showed spinal deformities. Our findings suggest that, in addition to the 16p11.2 microdeletion, other CNVs are potentially important in predisposing to CS.

Keywords: congenital scoliosis; congenital vertebral malformation; copy number variant; CNV



Citation: Lai, W.; Feng, X.; Yue, M.; Cheung, P.W.H.; Choi, V.N.T.; Song, Y.-Q.; Luk, K.D.K.; Cheung, J.P.Y.; Gao, B. Identification of Copy Number Variants in a Southern Chinese Cohort of Patients with Congenital Scoliosis. *Genes* **2021**, *12*, 1213. <https://doi.org/10.3390/genes12081213>

Academic Editor: Luisa Politano

Received: 31 May 2021

Accepted: 3 August 2021

Published: 5 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Among all musculoskeletal disorders, scoliosis is one of the most common diseases, affecting around 3% of the world population, which can occur as an isolated defect or as a concomitant symptom in other diseases or syndromes [1]. Scoliosis is categorized into several main groups, including congenital scoliosis (CS), idiopathic scoliosis (IS), neuromuscular scoliosis, and degenerative scoliosis. CS, which usually has first onset at birth or shortly after birth, affects approximately 0.5–1 in 1000 live births [2–5]. Compared with IS, CS is generally more severe due to the high risk of progressive deformity and associated problems such as pulmonary compromise [6]. One of the most significant differences between CS and IS is that IS does not have an association with congenital vertebral malformation (CVM), whereas CVM is the major cause leading to CS. CVM can be classified into several subclasses, including failure of vertebral formation (e.g., hemivertebrae, wedged vertebrae), failure of vertebral segmentation (e.g., unilateral bar, block vertebrae), and mixed type. Of all CVMs, congenital hemivertebrae is the most common anomaly that causes CS [4,5].

During vertebral development, the paraxial mesoderm forms bilaterally paired blocks, named somites, along the anterior–posterior axis. The vertebral bodies are derived from somites formed in the presomitic mesoderm. This fundamental process is called somitogenesis. Once somitogenesis is disturbed, the resulting CVM may lead to spinal deformities.

The most commonly accepted mechanism governing somitogenesis is the clock and wave-front model, which is controlled and coordinated by several key signaling pathways, such as Notch, Wnt, Fgf and retinoic acid signaling pathways [7,8]. Genetic studies of human patients with CVM have identified a variety of mutations in components of Notch signaling pathway (e.g., *NOTCH2*, *DLL3*, *MESP2*, *LFNG*, *HES7*, and *RIPPLY2*) and also in several key transcription factors essential for somitogenesis (e.g., *TBX6*, *TBXT*, and *SOX9*). Nevertheless, the genetic basis for majority of patients with CS still remains unclear [1,9].

Copy number variation (CNV) is a type of structural variation of genome. With the advancement of genome-wide analysis tools, it has been revealed that CNVs are widespread in the human genome and account for a large fraction of human genetic diversity [10]. CNVs have been, so far, implicated in many disease states including scoliosis. Although a number of CNVs were found to be associated with adolescent idiopathic scoliosis (AIS) [11,12], there have not been many reports about CS-associated CNVs. The 16p11.2 microdeletion was found to be associated with CS [13], and recent studies demonstrated that a compound inheritance of a *TBX6*-containing 16p11.2 microdeletion and a *TBX6* mutation or hypomorphic haplotype accounted for 5–10% of patients with CS in different populations [14–17]. Additional CNVs, including 10q24.31, 17p11.2, 20p11, 22q11.2, and a few other regions, were respectively reported in individual patients with CVMs [18,19]. Besides 16p11.2 microdeletion, it is unknown whether other CNVs are prevalent in CS.

Here, we analyzed CNVs in a Southern Chinese cohort of patients with congenital hemivertebrae. CNVs were called from whole-exome sequencing (WES) data of 67 cases and 125 family members (controls). We identified 12 rare CNVs in 4 known scoliosis-associated genes and eight recessive CNVs in three genes. We also found 64 novel, rare CNVs in 14 genes that occurred in multiple patients but are very rare in our control group and the general population, suggesting a potential role for genetic susceptibility in the development of CS.

2. Materials and Methods

2.1. Patient Recruitment

The patients studied in this project were recruited from the Duchess of Kent Children's Hospital (DKCH), a tertiary scoliosis referral center in Hong Kong. The patients with CS were diagnosed by imaging such as plain standing whole-spine radiographs and computed tomography. A total of 67 patients with hemivertebrae were chosen for this study, of which 31 had single congenital hemivertebrae while 36 had multiple congenital hemivertebrae. Patients' personal data and medical records were collected under ethical privacy guidelines and approval. Ethics was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (HKU/HA HKW IRB Ref # UW 15-216), and written informed consent was obtained from all participants and/or parents/siblings.

2.2. Control Cohort

The control cohort studied in this project consisted of 125 participating family members of the recruited patients with CS. Only unaffected parents and siblings (without CS) were included. Accordingly, 58 out of 67 patients had family member(s) participating in this study, including 2 quintets, 14 quartets, 33 trios, and 9 duos.

2.3. Genomic DNA Extraction

Genomic DNAs were extracted from peripheral blood samples of 67 patients and 125 of their family members using Invitrogen™ ChargeSwitch gDNA Serum Kit. The purified genomic DNA was quantified by NanoDrop.

2.4. Whole-Exome Sequencing (WES) and Copy Number Variations (CNVs) Calling

WES was performed for all recruited patients with congenital hemivertebrae and participating family members by Novogene Co, Ltd. (Hong Kong, China), using the

Agilent SureSelect Human All Exon Kit on the Illumina sequencing platform. The WES data were processed as described previously [20]. The raw sequence data were first analyzed by fastp for quality control and filtering [21]. After filtering, the Q20 base of most samples was greater than 95%, and the Q30 base was greater than 90%. The sequence reads were mapped to the reference genome (GRCh37/hg19) by Burrows-Wheller Aligner v0.7.17 (BWA-MEM) [22] and further processed using SAMtools v1.10 to sort and index aligned reads [23]. The sam format files generated by BWA was converted to bam format files by SAMtools. CNVs were called from bam files with ExomeDepth v1.1.15, which is an R package based on a read depth algorithm [24]. ExomeDepth uses a robust statistical model to build an optimized reference set in maximizing the CNVs detection power. In this study, four healthy control family members (CS59A, CS71A, CS71B, and CS81A) were selected to generate the reference set.

2.5. CNVs Filtering

Several criteria were used to filter CNVs: (i) Bayes factor (BF) values were calculated for each variant. BF equals to the log₁₀ likelihood ratio of the alternative hypothesis (i.e., there is a CNV) over the null hypothesis (i.e., there is no CNV). $BF = \log_{10}(\text{alternative hypothesis}/\text{null hypothesis})$. BF value greater than 1 was regarded as a strong supporting evidence of CNV. CNVs with BF values smaller than 1 were excluded. (ii) As ExomeDepth cannot detect small size CNVs accurately, CNVs with size smaller than 100 bp were excluded. (iii) Because CNVs with high allele frequency in the general population are likely benign and less susceptible, the CNVs with the allele frequency greater than 0.01 were excluded (a minimum sample size of 100 is required). Database of Genomic Variants (DGV) and Genome Aggregation Database (gnomAD) were used. If available, the allele frequency in East Asian population was also checked. As different CNVs often overlap and have no clear boundaries, this filtration was conducted in a gene-based manner. If there were multiple CNVs covering the same gene, the maximum allele frequency was used for filtering. (iv) In a gene-based manner, the number of CNV recurrence was counted in patients and controls.

2.6. Real-Time Quantitative PCR (qPCR)

Real-time qPCR was performed to validate some of the detected CNVs. Briefly, ROX Reference Dye (0.4 μ L, 50X), forward and reverse primers (0.4 μ L each, 10 μ M), TB Green Premix Ex Taq (10 μ L, 2X, Tli RNaseH Plus, Takara), patients' genomic DNA (0.5 μ L, 10 ng/ μ L), and sterile ddH₂O (8.3 μ L) were mixed for qPCR, which was performed using Applied Biosystem™ StepOnePlus™ Real-Time PCR System. A locus outside of the detected CNV region of *NOTCH2*, *DSCAM* and *SNTG1* was used as reference locus (P1). P1 is near the region of chromosome 16p11.2 and previously used as a reference site to detect 16p11.2/TBX6 deletion [14,17]. Each sample was analyzed in triplicate. Quantities of the copy numbers of specific locus were determined by the delta Ct method. The $2^{-\Delta\Delta CT}$ method was used to analyze the relative changes. The qPCR primer sequences: *NOTCH2*-F: 5'-AGGAGGCGACCGAGAAGATG-3'; *NOTCH2*-R: 5'-CGATACTCACCATGCGCG-GG-3'; *DSCAM*-F: 5'-AGCGAACGTTCCCTATCGCTT-3'; *DSCAM*-R: 5'-TTTCACTTATGCGCCCTGGG-3'; *SNTG1*-F: 5'-GTCTACATGGGCTGGTG-TGA-3'; *SNTG1*-R: 5'-CTGGAGGTGCCAGAACTTG-3'; P1-F: 5'-GGGGAAGGAACTTATCATGAC-3'; P1-R: 5'-TCGTGTTTCCCTGTTGTACC-3'.

3. Results

3.1. CS Cohort and WES

In our cohort, we recruited a total of 92 patients with CS, in which vertebral malformations, such as hemivertebrae, unilateral bar, or block vertebrae, were identified. This operational definition thus excluded other types of scoliosis such as AIS. Because hemivertebrae is the most common type of vertebral malformation in CS and has the greatest potential for rapid progression (5–10 degrees/year) [25], 67 patients with congenital

hemivertebrae were first selected. Further, 125 healthy family members of 58 patients were enrolled for this study, including parents and siblings from two quintets, 14 quartets, 33 trios, and nine duos. WES was performed for all 67 patients and 125 participating family members (controls). The contaminating sequencing adaptors and low-quality reads were first removed and the filtered reads were then aligned to the reference human genome (GRCh37/hg19), sorted and indexed.

3.2. CNV Calling

CNVs were called from the sequence reads with the read-depth analysis tool ExomeDepth, which has high sensitivity and specificity at the exon level [24,26]. Four healthy parents who were not carriers of 16p11.2 microdeletion but whose children have been previously diagnosed with TBX6 compound heterozygosity [17] were selected to generate the reference set for ExomeDepth analysis. After CNV calling of the 67 patients with CS, a total of 15,671 CNVs were detected. On average, each patient carries around 234 CNVs. By counting repeatedly occurring CNVs among different cases, there were 6084 distinct CNVs. This strategy successfully identified TBX6-containing 16p11.2 microdeletion in four patients as previously reported [17]. For the control group, a total of 27,116 CNVs were detected from 125 family control members. On average, each control carried around 217 CNVs. By counting repeatedly occurring CNVs among different controls, there were 7171 distinct CNVs. Although more CNVs were detected in a few individuals (six patients and four controls), there was no significant difference between the patient group and the control group (Supplementary Figure S1). The average CNV numbers in patients and controls were similar to the previous report [24]. Afterwards, we analyzed all CNVs by employing both a candidate gene approach and family-based filtering and prioritization strategies. A workflow is shown in Figure 1.

3.3. CNVs in Candidate Genes

To identify CNVs associated with CS, we firstly used the candidate gene approach, and focused on CNVs that contained genes known to be involved in scoliosis or somitogenesis. After checking allele frequencies of CNVs in the Database of Genomic Variants (DGV) and the Genome Aggregation Database (gnomAD), a total of 12 rare CNVs that influence four candidate genes were found in 12 patients, including known TBX6-containing 16p11.2 heterozygous deletion in four cases [17]. We also identified two rare CNVs that contained NOTCH2, a key component in the Notch signaling pathway, in two patients, and six rare CNVs in AIS-associated genes, DSCAM [27] and SNTG1 [28,29], in six patients (Table 1). We then checked these CNVs in their available family members and found that they are either novel mutations (NOTCH2 in CS043, DSCAM in CS018 and CS036, and SNTG1 in CS048) or paternally inherited (DSCAM in CS050) (Table 1). We were unable to determine the inheritance patterns of other patients (NOTCH2 in CS033, DSCAM in CS053 and CS064) due to the lack of family members.

Among the identified rare CNVs, the TBX6-containing chromosome 16p11.2 microdeletion had been previously validated [17]. Here, we further examined CNVs that contained NOTCH2, DSCAM or SNTG1 genes. Indeed, qPCR analysis detected heterozygous deletions within the NOTCH2, DSCAM and SNTG1 loci (Supplementary Figure S2), indicating the reliability of CNVs called from WES data by ExomeDepth.

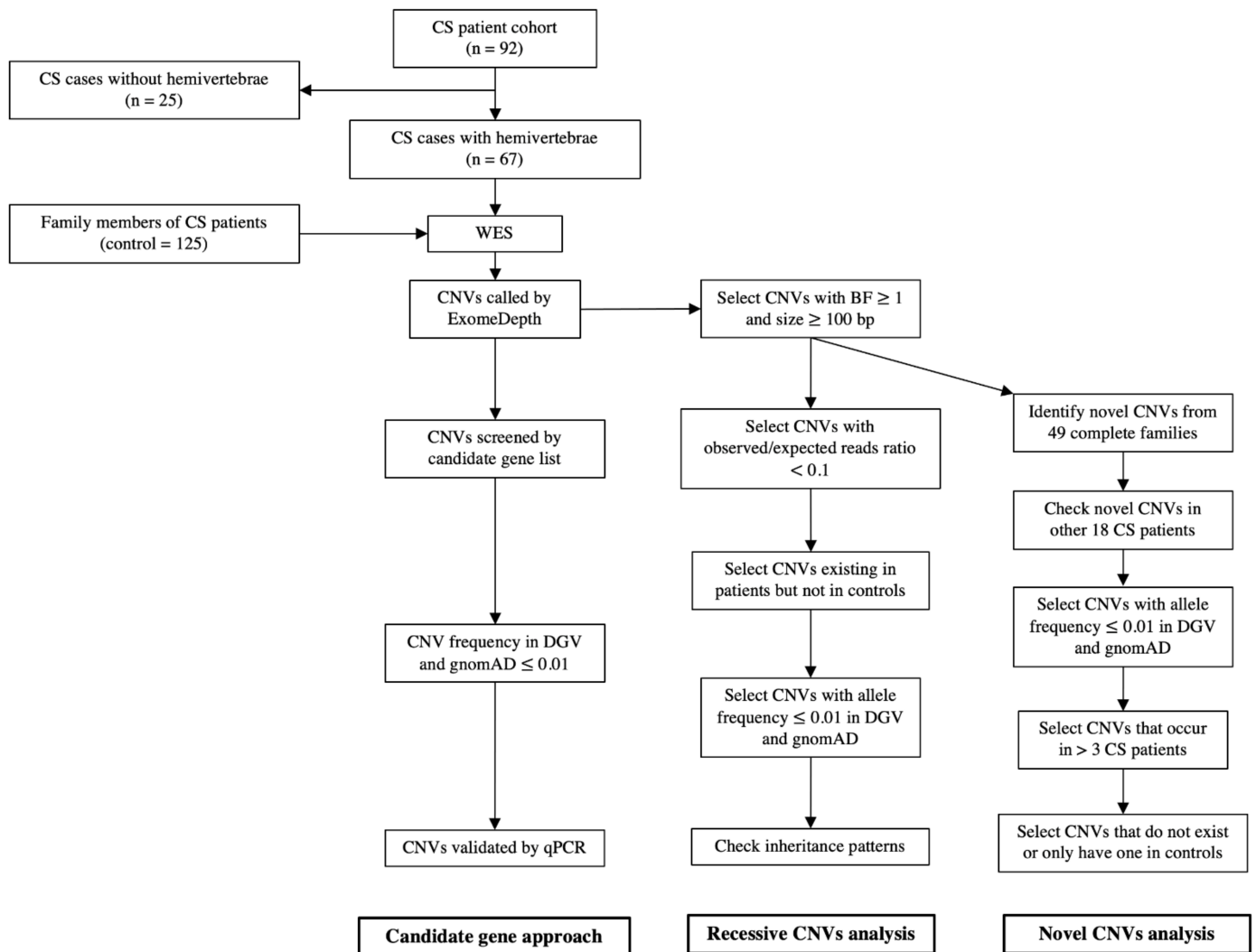


Figure 1. The workflow of CNV analysis. This strategy detected CNVs in several candidate genes and identified recessive and novel rare CNVs enriched in patients with CS.

3.4. Recessive CNVs in Patients with CS

We then searched for homozygous CNVs (observed/expected reads ratio < 0.1) in 67 patients and 125 controls. After excluding the homozygous CNVs that existed in both patients and controls, we identified unique homozygous CNVs in eight patients with CS. The heterozygous deletions of these loci are rare in DGV or gnomAD database (Table 2). Considering that homozygous CNVs might be inherited from parents, we further checked their inheritance pattern and found that they were either *novel* mutations or unknown due to lack of parents' data. These recessive CNVs contained three genes, *NBPF20* (Neuroblastoma Breakpoint Family Member 20), *FAM138C* (Family with Sequence Similarity 138 Member C), and *DHX40* (DEAH-Box Helicase 40). Interestingly, the *DHX40*-containing homozygous CNVs were detected in six patients but was not reported in DGV or gnomAD. *DHX40*-containing heterozygous CNVs are also very rare (Table 2). *FAM138C* is an RNA gene and *NBPF20* is a member of NBPF family characterized by tandemly repeats of DUF1220 domain, but their functions are unclear. *DHX40* encodes a member of the DExD/H-box RNA helicase superfamily that catalyzes the unwinding of double-stranded RNA and has an essential role in RNA metabolism [30].

Table 1. CNVs found with candidate gene approach (N.D., not determined; N.A., not applied).

| Gene | Patient | Type | Chr | Start | End | Size (bp) | Bayes Factor | Reads Ratio (Observed/Expected) | Exons Annotation (hg19) (Gene_exon) | Inheritance Pattern | Highest Frequency in DGV (Sample Size >100) | gnomAD_Structural_Variants Frequency (Heterozygous Loss) |
|--------|---------|----------|-----|-------------|-------------|-----------|--------------|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|---------------------------------------------|----------------------------------------------------------|
| NOTCH2 | CS033 | deletion | 1 | 120,611,949 | 120,612,020 | 71 | 4.62 | 0.657 | NOTCH2_1 | N.D. | 0.0037 | 0.00037 (0 in East Asia) |
| | CS043 | deletion | 1 | 120,539,621 | 120,612,020 | 72,399 | 8.46 | 0.722 | NOTCH2_1-4 | De novo | 0.0037 | 0.00037 (0 in East Asia) |
| DSCAM | CS018 | deletion | 21 | 41,452,080 | 41,452,267 | 187 | 4.65 | 0.429 | DSCAM_25 | De novo | 0.00049 | N.A. |
| | CS036 | deletion | 21 | 41,452,080 | 41,452,267 | 187 | 6.02 | 0.415 | DSCAM_25 | De novo | 0.00049 | N.A. |
| | CS050 | deletion | 21 | 41,452,080 | 41,452,267 | 187 | 5.8 | 0.468 | DSCAM_25 | Paternal | 0.00049 | N.A. |
| | CS053 | deletion | 21 | 41,452,080 | 41,452,267 | 187 | 5.45 | 0.494 | DSCAM_25 | N.D. | 0.00049 | N.A. |
| | CS064 | deletion | 21 | 41,452,080 | 41,452,267 | 187 | 6.1 | 0.463 | DSCAM_25 | N.D. | 0.00049 | N.A. |
| SNTG1 | CS048 | deletion | 8 | 51,503,440 | 51,571,223 | 67,783 | 6.37 | 0.43 | SNTG1_13-15 | De novo | 0.0002 | 0.000046 (0 in East Asia) |
| TBX6 | CS059 | deletion | 16 | 29,674,601 | 30,199,897 | 525,296 | 644 | 0.555 | SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_2-4,KIF22_1-13,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-1,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-3,CORO1A_4-10,NPIPL3_3-1,SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_3-4,KIF22_2-12,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-2,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-11,BOLA2B_3-1,SLX1A_1-5,SULT1A3_3-9,RP11-347C12.3_5-2,NPIPL3_1,SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_3-4,KIF22_2-12,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-2,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-10,NPIPL3_1,SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_3-4,KIF22_2-12,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-2,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-10 | N.D. | 0.0005 | 0.0001462 (0 in East Asia) |
| | CS071 | deletion | 16 | 29,495,011 | 30,218,221 | 723,210 | 754 | 0.572 | SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_3-4,KIF22_2-12,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-2,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-11,BOLA2B_3-1,SLX1A_1-5,SULT1A3_3-9,RP11-347C12.3_5-2,NPIPL3_1,SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_3-4,KIF22_2-12,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-2,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-10,NPIPL3_1,SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_3-4,KIF22_2-12,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-2,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-10 | De novo | 0.0005 | 0.0001462 (0 in East Asia) |
| | CS078 | deletion | 16 | 29,498,516 | 30,199,897 | 701,381 | 690 | 0.578 | SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_3-4,KIF22_2-12,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-2,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-10,NPIPL3_1,SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_3-4,KIF22_2-12,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-2,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-10 | N.D. | 0.0005 | 0.0001462 (0 in East Asia) |
| | CS081 | deletion | 16 | 29,498,516 | 30,199,897 | 701,381 | 645 | 0.558 | SPN_2,AC009133.19_2-3,QPRT_1-4,C16orf54_2,ZG16_3-4,KIF22_2-12,MAZ_1-5,PRRT2_2-3,PAGR1_1-3,CTD-2574D22.6_1-2,MVP_2-15,CDIPT_6-2,SEZ6L2_16-1,ASPHD1_1-3,KCTD13_6-1,TMEM219_1-4,TAOK2_2-16,HIRIP3_7-2,INO80E_1-7,DOC2A_11-2,C16orf92_2-3,FAM57B_5-1,ALDOA_8-16,PPP4C_2-9,TBX6_9-2,YPEL3_4-1,GDPD3_10-1,MAPK3_8-1,CORO1A_2-10 | N.D. | 0.0005 | 0.0001462 (0 in East Asia) |

Table 2. Recessive CNVs unique in patients with CS (N.D., not determined; N.A., not applied).

| Gene | Patient | Type | Chr | Start | End | Size (bp) | Bayes Factor | Reads Ratio (Observed/Expected) | Exons Annotation (hg19) (Gene_exon) | Inheritance Pattern | Highest Frequency in DGV (Sample Size > 100) | gnomAD_ Structural Variants Frequency (Heterozygous Loss) |
|----------------|---------|----------|-----|-------------|-------------|-----------|--------------|------------------------------------|----------------------------------------|------------------------|----------------------------------------------------|--------------------------------------------------------------------|
| <i>NBPF20</i> | CS047 | deletion | 1 | 148,261,458 | 148,262,366 | 908 | 5.27 | 0.04 | NBPF20_98-99 | De novo | 0 | 0 (0 in East Asia) |
| <i>FAM138C</i> | CS048 | deletion | 9 | 35,061 | 35,519 | 458 | 6.51 | 0 | FAM138C_1-2 | * De novo | 0.0074 | N.A. |
| <i>DHX40</i> | CS004 | deletion | 17 | 57,656,834 | 57,657,240 | 406 | 5.38 | 0 | DHX40_9-10 | N.D. | 0.0000922 | 0.0025 (0.008152 in East Asia) |
| | CS035 | | | | | | 4.91 | | | N.D. | | |
| | CS043 | | | | | | 6 | | | * De novo | | |
| | CS050 | | | | | | 7.23 | | | * De novo | | |
| | CS053 | | | | | | 7.02 | | | N.D. | | |
| CS057 | 6.29 | De novo | | | | | | | | | | |

* The CNV is also not present in the healthy siblings.

3.5. Novel CNVs in Patients with CS

We also sought to identify CS-associated novel CNVs and first analyzed the data from 49 complete families (two quintets, 14 quartets or 33 trios). The detected novel CNVs were then checked in the other 18 patients (nine singlets and nine duos). Eventually, we identified 64 CNVs in 14 genes that occurred in more than three patients but did not exist or was very rare (<1%) in family control group. Those with high CNV allele frequency (>1%) in the general population were also filtered out. This strategy successfully identified the known *TBX6*-containing CNVs in four patients [17] and *DHX40*-containing homozygous CNVs in six patients. Interestingly, we also found there are four additional heterozygous *DHX40* CNVs (Table 3 and Supplementary Table S1). Most of the identified novel CNVs were heterozygous loss, and one was gain of one copy. Our CNV shortlist includes genes involved in ubiquitination (*NAE1*, *MYSM1*), enzymatic activities (*MME*, *PHKB*), ion/small molecule transportation (*SCN7A*, *ABCA6*), meiosis (*MNS1*, *SPO11*), spermatogenesis (*GMCL1*), GTPase activity (*RASA2*), TNF signaling (*NSMAF*), or with unknown function (*LRRC40*).

Table 3. Novel CNVs enriched in patients with CS (N.A., not applied).

| Gene | Type | Chr | Size (bp) | Count in 67 Patients | Count in 125 Controls | Highest Frequency in DGV (Sample Size > 100) | gnomad_Structural_Variants Frequency (Heterozygous Loss) | gnomad_East Asia_Structural_Variants Frequency (Heterozygous Loss) |
|---------------|-------------|-----|-----------------|----------------------|-----------------------|----------------------------------------------|----------------------------------------------------------|--------------------------------------------------------------------|
| <i>LRRC40</i> | deletion | 1 | 30,080–383,938 | 4 | 0 | 0.009556907 | 0.0000461 | 0.000 |
| <i>SCN7A</i> | deletion | 2 | 544–11,914 | 4 | 0 | 0.002257336 | 0.0000461 | 0.000 |
| <i>MME</i> | deletion | 3 | 279–55,232 | 4 | 0 | 0.000798722 | 0.0000462 | 0.000 |
| <i>NAE1</i> | deletion | 16 | 6724–71,803 | 4 | 0 | N.A. ^b | 0 | 0.000 |
| <i>TBX6</i> | deletion | 16 | 525,296–723,210 | 4 | 0 | 0.0005 | 0.0001462 | 0.000 |
| <i>DHX40</i> | deletion | 17 | 107–2354 | 10 ^a | 1 | 0.0000922 | 0.0025 | 0.008152 |
| <i>GMCL1</i> | deletion | 2 | 11,694–24,032 | 5 | 1 | 0.001303781 | 0.0000479 | 0.000 |
| <i>MYSM1</i> | deletion | 1 | 190–10,053 | 4 | 1 | 0.0000922 | 0.0000461 | 0.0004139 |
| <i>RASA2</i> | deletion | 3 | 2892–100,149 | 4 | 1 | 0.00086881 | 0 | 0.000 |
| <i>NSMAF</i> | deletion | 8 | 203–12,325 | 4 | 1 | 0.001145475 | 0 | 0.000 |
| <i>MNS1</i> | deletion | 15 | 52,610–323,156 | 4 | 1 | 0.00518807 | 0.0000922 | 0.000 |
| <i>PHKB</i> | deletion | 16 | 7697–99,254 | 4 | 1 | 0.001198083 | 0.0000481 | 0.000 |
| <i>SPO11</i> | deletion | 20 | 730–33,608 | 4 | 1 | 0.00064226 | 0 | 0.000 |
| <i>ABCA6</i> | duplication | 17 | 560–13,580 | 4 | 1 | 0.0009219 | 0 | 0.000 |

^a *DHX40* has 6 homozygous (listed in Table 2) and 4 heterozygous CNVs. ^b No CNV with sample size more than 100 is found within the *NAE1* locus. Note: Detailed information of these novel CNVs is shown in Table S1.

4. Discussion

CS is a genetically heterogeneous disorder with evidence for multiple causative genes. However, the genetic causes of the majority of patients still remain unknown. As most cases of CS are of sporadic etiology, CNVs may have greater influence than single nucleotide variations (SNVs) [31]. This was well exemplified by the *TBX6*-containing 16p11.2 microdeletion in previous CS studies [14–17]. Here, we systematically investigated CNVs in a cohort of patients with congenital hemivertebrae and their family controls. We identified the well-known CNVs at chromosome 16p11.2, as well as a number of new CNVs that are potentially associated with CS. Haploinsufficiency of Notch signaling pathway has been demonstrated to cause CS [32] and mutations in *NOTCH2* caused Alagille syndrome and Hajdu–Cheney syndrome, both of which showed abnormal curvature of the spine [33,34]. In our study, we found one short and one long CNVs at *NOTCH2* locus in two patients, spanning one and four exons of *NOTCH2*, respectively. As no significant coding SNV could be detected in *NOTCH2* of these two patients, it is unclear whether heterozygous loss of *NOTCH2* is sufficient to cause CS or other non-coding *NOTCH2* SNVs or environmental factors [32] may contribute. Interestingly, CNVs in two AIS-associated genes, *DSCAM* [27] and *SNTG1* [28,29], were found in six patients, suggesting CS and AIS may be genetically related to each other.

An intriguing finding in our analysis is the identification of CNVs spanning various exons of *DHX40* in ten patients, including six homozygous and four heterozygous CNVs. Most of them are novel mutations. *DHX40* belongs to the conserved DExD/H-box RNA

helicase family, which facilitates the ATP-dependent unwinding of RNA secondary structures [30]. However, the biological functions of each member remained poorly understood. Interestingly, the *DHX40* mutant mice were described to exhibit abnormal bone structure and bone mineralization (Mouse Genome Informatics, MGI: 1914737), indicating a role of *DHX40* in bone development. The mutant of its family member *DHX35* was described to have abnormal vertebrae morphology and scoliosis in mice (MGI: 1918965). Patients carrying *DHX37* mutations showed developmental delay and intellectual disability as well as vertebral anomalies [30]. These observations in the mouse and human might suggest a potential link between DHX family members and CVM.

Although the function of *NBPF20* is unknown, it locates at chromosome 1q21.1, which microdeletion is associated with a variety of phenotypes including skeletal malformations such as scoliosis [35,36]. This region also contains other NBPF family members, such as *NBPF10*, whose genetic variants were implicated in Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome (OMIM # 277000) [37], a disease associated with CS [38].

Among the candidate genes of the identified novel CNVs, *RASA2* (RAS P21 Protein Activator 2) and *MYSM1* (Myb-Like, SWIRM, and MPN domains 1) are of particular interest. *RASA2* encodes a GAP (GTPase-activating protein) protein and functions as a suppressor of RAS by promoting its intrinsic GTPase activity. Rare variants in *RASA2* have been found associated with Noonan syndrome [39]. As scoliosis occurs frequently in Noonan syndrome [40], *RASA2* is a potential candidate gene for CS. It would be interesting to investigate the *RASA2* mutant mouse phenotype. *MYSM1* is a deubiquitinase reported to be essential for bone formation [41] and its mutant mice have truncated and kinky tails [42–44], which are often associated with vertebral malformations [45]. Indeed, an X-ray from the International Mouse Phenotyping Consortium (MGI: 2444584) exhibited a spinal deformity in the *MYSM1* mutant mouse (Supplementary Figure S3), indicating a potential role of *MYSM1* in spinal development and predisposition to CS. Further detailed phenotypic analysis of mutant animals is needed to validate its pathogenicity in CS.

Although there are a few reports of heterozygous point mutations in CS patients [17,46–48], the dominant negative effect was only demonstrated by a novel *TBXT* mutation [17]. Considering the low familial recurrence rate in CS, recessive or compound heterozygous mutations are more likely to be the major cause of CS. In this regard, heterozygous CNVs are not sufficient to induce CS. Their pathogenicity may be explained by the following genetic models. First, in our cases, the patients carrying heterozygous CNVs may have additional risk variant or haplotype on the other allele. This possibility has been well exemplified by the 16p11.2/*TBX6* mutations and haplotype. However, further analysis of risk variant/haplotype in our study is severely limited by our dataset from WES as they may reside in non-coding regions that regulate gene transcription. We did not detect significant deleterious mutations in the coding regions of these genes. Second, additional mutations in other relevant genes may increase the risk of CS (polygenic model). Other possibilities include environmental contributions and novel mutations in somatic tissues. Environmental factors, such as short-term gestational hypoxia, have been found to cause CS in combination with haploinsufficiency of Notch signaling pathway genes [32]. On the other hand, somatic mutations may serve as the “second hit” in addition to the heterozygous germline CNV mutations (first hit). This genetic model has been well demonstrated in other diseases as well as dystrophic scoliosis caused by *NF1* [49–52]. Testing the above models will require whole genome sequencing, more comprehensive data analysis, and the isolation of malformed vertebral tissues in future studies.

5. Conclusions

In this study, we investigated the genetic basis of CS by analyzing CNVs in a cohort of CS families. Based on the candidate gene approach and family-based filtering of CNVs, we identified both known CS-associated genes and a set of new susceptibility genes, some of which (e.g., *DHX40*, *RASA2*, and *MYSM1*) warrant further investigations in larger cohorts as well as functional characterization. Given the well-defined example of the *TBX6*

compound inheritance and the complex genetic nature of CS, future studies examining the combined effects of SNVs and CNVs and somatic tissues may help better decipher the genetic etiology and heterogeneity of CS.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12081213/s1>, Figure S1: Total number CNVs in each patient and family control, Figure S2: Quantitative PCR analysis for validating heterozygous deletion, Figure S3: The X-ray of *MYSM1* mutant mouse from IMPC. Table S1: Novel CNVs enriched in CS patients.

Author Contributions: Conceptualization, B.G., J.P.Y.C. and K.D.K.L.; methodology, W.L., X.F., M.Y., V.N.T.C. and P.W.H.C.; software, W.L., M.Y. and X.F.; validation, W.L., X.F. and B.G.; resources, B.G., J.P.Y.C., Y.-Q.S. and K.D.K.L.; data curation, J.P.Y.C. and B.G.; manuscript preparation, W.L., X.F., J.P.Y.C. and B.G.; supervision, B.G., J.P.Y.C. and Y.-Q.S.; project administration, B.G. and J.P.Y.C.; funding acquisition, B.G., J.P.Y.C. and K.D.K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Hong Kong Health and Medical Research Fund (06171406) to B.G., and the Tam Sai Kit Endowment Fund to K.D.K.L.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (HKU/HAHKW, IRB Reference Number UW 15-216).

Informed Consent Statement: Informed consents was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request due to restrictions.

Acknowledgments: Nil.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Giampietro, P.F. Genetic Aspects of Congenital and Idiopathic Scoliosis. *Scientifica* **2012**, *2012*, 1–15. [[CrossRef](#)]
2. Shands, A.R.; Eisberg, H.B. The incidence of scoliosis in the state of Delaware; a study of 50,000 minifilms of the chest made during a survey for tuberculosis. *J. Bone Jt. Surg. Am. Vol.* **1955**, *37*, 1243–1249. [[CrossRef](#)]
3. Wynne-Davies, R. Congenital vertebral anomalies: Aetiology and relationship to spina bifida cystica. *J. Med. Genet.* **1975**, *12*, 280–288. [[CrossRef](#)]
4. McMaster, M.J.; Ohtsuka, K. The natural history of congenital scoliosis. A study of two hundred and fifty-one patients. *J. Bone Jt. Surg. Am. Vol.* **1982**, *64*, 1128–1147. [[CrossRef](#)]
5. McMaster, M.J.; Singh, H. Natural History of Congenital Kyphosis and Kyphoscoliosis. A Study of One Hundred and Twelve Patients. *J. Bone Jt. Surg. Am. Vol.* **1999**, *81*, 1367–1383. [[CrossRef](#)] [[PubMed](#)]
6. Cheng, J.C.; Castelein, R.M.; Chu, W.C.; Danielsson, A.J.; Dobbs, M.B.; Grivas, T.B.; Gurnett, C.A.; Luk, K.D.; Moreau, A.; Newton, P.O.; et al. Adolescent idiopathic scoliosis. *Nat. Rev. Dis. Primers* **2015**, *1*, 15030. [[CrossRef](#)] [[PubMed](#)]
7. Pourquié, O. Vertebrate Segmentation: From Cyclic Gene Networks to Scoliosis. *Cell* **2011**, *145*, 650–663. [[CrossRef](#)]
8. Maroto, M.; Bone, R.A.; Dale, J.K. Somitogenesis. *Development* **2012**, *139*, 2453–2456. [[CrossRef](#)]
9. Giampietro, P.; Raggio, C.; Blank, R.D.; McCarty, C.; Broeckel, U.; Pickart, M. Clinical, Genetic and Environmental Factors Associated with Congenital Vertebral Malformations. *Mol. Syndr.* **2012**, *4*, 94–105. [[CrossRef](#)]
10. Sebat, J.; Lakshmi, B.; Troge, J.; Alexander, J.; Young, J.; Lundin, P.; Månér, S.; Massa, H.; Walker, M.; Chi, M.; et al. Large-Scale Copy Number Polymorphism in the Human Genome. *Science* **2004**, *305*, 525–528. [[CrossRef](#)]
11. Buchan, J.G.; Alvarado, D.M.; Haller, G.; Aferol, H.; Miller, N.H.; Dobbs, M.B.; Gurnett, C.A. Are copy number variants associated with adolescent idiopathic scoliosis? *Clin. Orthop. Relat. Res.* **2014**, *472*, 3216–3225. [[CrossRef](#)] [[PubMed](#)]
12. Sadler, B.; Haller, G.; Antunes, L.; Bledsoe, X.; Morcuende, J.; Giampietro, P.; Raggio, C.; Miller, N.; Kidane, Y.; Wise, C.A.; et al. Distal chromosome 16p11.2 duplications containing SH2B1 in patients with scoliosis. *J. Med. Genet.* **2019**, *56*, 427–433. [[CrossRef](#)] [[PubMed](#)]
13. Al-Kateb, H.; Khanna, G.; Filges, I.; Hauser, N.; Grange, D.K.; Shen, J.; Smyser, C.D.; Kulkarni, S.; Shinawi, M. Scoliosis and vertebral anomalies: Additional abnormal phenotypes associated with chromosome 16p11.2 rearrangement. *Am. J. Med. Genet. Part A* **2014**, *164*, 1118–1126. [[CrossRef](#)]
14. Wu, N.; Ming, X.; Xiao, J.; Wu, Z.; Chen, X.; Shinawi, M.; Shen, Y.; Yu, G.; Liu, J.; Xie, H.; et al. TBX6 Null Variants and a Common Hypomorphic Allele in Congenital Scoliosis. *N. Engl. J. Med.* **2015**, *372*, 341–350. [[CrossRef](#)]

15. Takeda, K.; Kou, I.; Kawakami, N.; Iida, A.; Nakajima, M.; Ogura, Y.; Imagawa, E.; Miyake, N.; Matsumoto, N.; Yasuhiko, Y.; et al. Compound Heterozygosity for Null Mutations and a Common Hypomorphic Risk Haplotype in TBX6 Causes Congenital Scoliosis. *Hum. Mutat.* **2017**, *38*, 317–323. [[CrossRef](#)] [[PubMed](#)]
16. Lefebvre, M.; Duffourd, Y.; Jouan, T.; Poe, C.; Jean-Marçais, N.; Verloes, A.; St-Onge, J.; Riviere, J.-B.; Petit, F.; Pierquin, G.; et al. Autosomal recessive variations of TBX6, from congenital scoliosis to spondylocostal dysostosis. *Clin. Genet.* **2016**, *91*, 908–912. [[CrossRef](#)]
17. Feng, X.; Cheung, J.P.Y.; Je, J.S.H.; Cheung, P.W.H.; Chen, S.; Yue, M.; Wang, N.; Choi, V.N.T.; Yang, X.; Song, Y.; et al. Genetic variants of TBX6 and TBXT identified in patients with congenital scoliosis in Southern China. *J. Orthop. Res.* **2020**, *39*, 971–988. [[CrossRef](#)]
18. Liu, Z.L.; Wu, N.; Wu, Z.H.; Zuo, Y.Z.; Qiu, G.X. Advances in congenital vertebral malformation caused by genomic copy number variation. *Zhonghua Wai Ke Za Zhi* **2016**, *54*, 313–316.
19. Gao, X.; Gotway, G.; Rathjen, K.; Johnston, C.; Sparagana, S.; Wise, C.A. Genomic Analyses of Patients with Unexplained Early-Onset Scoliosis. *Spine Deform.* **2014**, *2*, 324–332. [[CrossRef](#)]
20. Li, C.; Wang, N.; Schaffer, A.A.; Liu, X.; Zhao, Z.; Elliott, G.; Garrett, L.; Choi, N.T.; Wang, Y.; Wang, Y.; et al. Mutations in COMP cause familial carpal tunnel syndrome. *Nat. Commun.* **2020**, *11*, 1–16. [[CrossRef](#)]
21. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)]
22. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
23. Li, H.; Handsaker, R.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
24. Plagnol, V.; Curtis, J.; Epstein, M.; Mok, K.; Stebbings, E.; Grigoriadou, S.; Wood, N.; Hambleton, S.; Burns, S.; Thrasher, A.; et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* **2012**, *28*, 2747–2754. [[CrossRef](#)]
25. Burnei, G.; Gavrilu, S.; Vlad, C.; Georgescu, I.; Ghita, R.A.; Dughilă, C.; Japie, E.M.; Onilă, A. Congenital scoliosis: An up-to-date. *J. Med. Life* **2015**, *8*, 388–397.
26. Ellingford, J.M.; Campbell, C.; Barton, S.; Bhaskar, S.; Gupta, S.; Taylor, R.L.; Sergouniotis, P.I.; Horn, B.; Lamb, J.; Michaelides, M.; et al. Validation of copy number variation analysis for next-generation sequencing diagnostics. *Eur. J. Hum. Genet.* **2017**, *25*, 719–724. [[CrossRef](#)]
27. Sharma, S.; Gao, X.; Londono, D.; Devroy, S.E.; Mauldin, K.N.; Frankel, J.T.; Brandon, J.M.; Zhang, D.; Li, Q.-Z.; Dobbs, M.B.; et al. Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. *Hum. Mol. Genet.* **2011**, *20*, 1456–1466. [[CrossRef](#)] [[PubMed](#)]
28. Bashiardes, S.; Veile, R.; Allen, M.; Wise, C.A.; Dobbs, M.; Morcuende, J.A.; Szappanos, L.; Herring, J.A.; Bowcock, A.M.; Lovett, M. SNTG1, the gene encoding gamma1-syntrophin: A candidate gene for idiopathic scoliosis. *Hum. Genet.* **2004**, *115*, 81–89. [[CrossRef](#)]
29. Tassano, E.; Ronchetto, P.; Severino, M.; Divizia, M.; Lerone, M.; Uccella, S.; Nobili, L.; Tavella, E.; Morerio, C.; Coviello, D.; et al. Scoliosis with cognitive impairment in a girl with 8q11.21q11.23 microdeletion and SNTG1 disruption. *Bone* **2021**, *150*, 116022. [[CrossRef](#)]
30. Paine, I.; Posey, J.; Grochowski, C.M.; Jhangiani, S.N.; Rosenheck, S.; Kleyner, R.; Marmorale, T.; Yoon, M.; Wang, K.; Robison, R.; et al. Paralog Studies Augment Gene Discovery: DDX and DHX Genes. *Am. J. Hum. Genet.* **2019**, *105*, 302–316. [[CrossRef](#)]
31. Lupski, J.R. Genomic rearrangements and sporadic disease. *Nat. Genet.* **2007**, *39*, S43–S47. [[CrossRef](#)] [[PubMed](#)]
32. Sparrow, D.B.; Chapman, G.; Smith, A.J.; Mattar, M.Z.; Major, J.A.; O'Reilly, V.C.; Saga, Y.; Zackai, E.H.; Dormans, J.P.; Alman, B.A.; et al. A Mechanism for Gene-Environment Interaction in the Etiology of Congenital Scoliosis. *Cell* **2012**, *149*, 295–306. [[CrossRef](#)]
33. McDaniell, R.; Warthen, D.M.; Sanchez-Lara, P.A.; Pai, A.; Krantz, I.D.; Piccoli, D.A.; Spinner, N.B. NOTCH2 Mutations Cause Alagille Syndrome, a Heterogeneous Disorder of the Notch Signaling Pathway. *Am. J. Hum. Genet.* **2006**, *79*, 169–173. [[CrossRef](#)]
34. Gray, M.J.; Kim, C.A.; Bertola, D.R.; Arantes, P.R.; Stewart, H.; Simpson, M.A.; Irving, M.D.; Robertson, S.P. Serpentine fibula polycystic kidney syndrome is part of the phenotypic spectrum of Hajdu–Cheney syndrome. *Eur. J. Hum. Genet.* **2011**, *20*, 122–124. [[CrossRef](#)] [[PubMed](#)]
35. Haldeman-Englert, C.R.; Jewett, T. 1q21.1 Recurrent Microdeletion. In *Gene Reviews(R)*; Adam, M.P., Ardinger, H.H., Eds.; University of Washington: Seattle, WA, USA, 1993.
36. Gamba, B.F.; Zechi-Ceide, R.M.; Kokitsu-Nakata, N.M.; Vendramini-Pittoli, S.; Rosenberg, C.; Santos, A.C.K.; Ribeiro-Bicudo, L.; Richieri-Costa, A. Interstitial 1q21.1 Microdeletion Is Associated with Severe Skeletal Anomalies, Dysmorphic Face and Moderate Intellectual Disability. *Mol. Syndr.* **2016**, *7*, 344–348. [[CrossRef](#)]
37. Pan, H.X.; Luo, G.N.; Wan, S.Q.; Qin, C.L.; Tang, J.; Zhang, M.; Du, M.; Xu, K.K.; Shi, J.Q. Detection of de novo genetic variants in Mayer-Rokitansky-Kuster-Hauser syndrome by whole genome sequencing. *Eur. J. Obstet. Gynecol. Reprod. Biol. X* **2019**, *4*, 100089. [[CrossRef](#)]
38. Fisher, K.; Esham, R.H.; Thorneycroft, I. Scoliosis associated with typical Mayer-Rokitansky-Küster-Hauser syndrome. *South. Med. J.* **2000**, *93*, 243–246. [[CrossRef](#)]

39. Chen, P.-C.J.; Yin, J.; Yu, H.-W.; Yuan, T.; Fernandez, M.; Yung, C.; Trinh, Q.M.; Peltekova, V.D.; Reid, J.G.; Tworog-Dube, E.; et al. Next-generation sequencing identifies rare variants associated with Noonan syndrome. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11473–11478. [[CrossRef](#)]
40. Lee, C.-K.; Chang, B.S.; Hong, Y.M.; Yang, S.W.; Seo, J.B. Spinal deformities in Noonan syndrome: A clinical review of sixty cases. *J. Bone Jt. Surg. Am. Vol.* **2001**, *83*, 1495–1502. [[CrossRef](#)]
41. Li, P.; Yang, Y.M.; Sanchez, S.; Cui, D.C.; Dang, R.J.; Wang, X.Y.; Lin, Q.X.; Wang, Y.; Wang, C.; Chen, D.F.; et al. Deubiquitinase MYSM1 Is Essential for Normal Bone Formation and Mesenchymal Stem Cell Differentiation. *Sci. Rep.* **2016**, *6*, 1–11. [[CrossRef](#)]
42. Jiang, X.-X.; Nguyen, Q.; Chou, Y.; Wang, T.; Nandakumar, V.; Yates, P.; Jones, L.; Wang, L.; Won, H.; Lee, H.-R.; et al. Control of B Cell Development by the Histone H2A Deubiquitinase MYSM1. *Immunity* **2011**, *35*, 883–896. [[CrossRef](#)]
43. Liakath-Ali, K.; Vancollie, V.; Heath, E.H.; Smedley, D.P.; Estabel, J.; Sunter, D.; DiTommaso, T.; White, J.K.; Ramirez-Solis, R.; Smyth, I.; et al. Novel skin phenotypes revealed by a genome-wide mouse reverse genetic screen. *Nat. Commun.* **2014**, *5*, 1–11. [[CrossRef](#)] [[PubMed](#)]
44. DiTommaso, T.; Jones, L.K.; Cottle, D.L.; Gerdin, A.-K.; Vancollie, V.E.; Watt, F.M.; Ramirez-Solis, R.; Bradley, A.; Steel, K.P.; Sundberg, J.P.; et al. Identification of Genes Important for Cutaneous Function Revealed by a Large Scale Reverse Genetic Screen in the Mouse. *PLoS Genet.* **2014**, *10*, e1004705. [[CrossRef](#)] [[PubMed](#)]
45. Giampietro, P.F.; Blank, R.D.; Raggio, C.L.; Merchant, S.; Jacobsen, F.S.; Faciszewski, T.; Shukla, S.K.; Greenlee, A.R.; Reynolds, C.; Schowalter, D.B. Congenital and Idiopathic Scoliosis: Clinical and Genetic Aspects. *Clin. Med. Res.* **2003**, *1*, 125–136. [[CrossRef](#)]
46. Wu, N.; Wang, L.; Hu, J.; Zhao, S.; Liu, B.; Li, Y.; Du, H.; Zhang, Y.; Li, X.; Yan, Z.; et al. A Recurrent Rare SOX9 Variant (M469V) is Associated with Congenital Vertebral Malformations. *Curr. Gene Ther.* **2019**, *19*, 242–247. [[CrossRef](#)]
47. Lin, M.; Zhao, S.; Liu, G.; Huang, Y.; Yu, C.; Zhao, Y.; Wang, L.; Zhang, Y.; Yan, Z.; Wang, S.; et al. Identification of novel FBN1 variations implicated in congenital scoliosis. *J. Hum. Genet.* **2019**, *65*, 221–230. [[CrossRef](#)]
48. Al Dhaheri, N.; Wu, N.; Zhao, S.; Wu, Z.; Blank, R.D.; Zhang, J.; Raggio, C.; Halanski, M.; Shen, J.; Noonan, K.; et al. KIAA1217: A novel candidate gene associated with isolated and syndromic vertebral malformations. *Am. J. Med. Genet. Part A* **2020**, *182*, 1664–1672. [[CrossRef](#)]
49. Qian, F.; Germino, G.G. Mistakes happen: Somatic mutation and disease. *Am. J. Hum. Genet.* **1997**, *61*, 1000–1005. [[CrossRef](#)]
50. Watnick, T.J.; Torres, V.E.; Gandolph, M.A.; Qian, F.; Onuchic, L.F.; Klinger, K.W.; Landes, G.; Germino, G.G. Somatic Mutation in Individual Liver Cysts Supports a Two-Hit Model of Cystogenesis in Autosomal Dominant Polycystic Kidney Disease. *Mol. Cell* **1998**, *2*, 247–251. [[CrossRef](#)]
51. Lamlum, H.; Ilyas, M.; Rowan, A.; Clark, S.; Johnson, V.; Bell, J.; Frayling, I.; Efstathiou, J.; Pack, K.; Payne, S.; et al. The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: A new facet to Knudson's 'two-hit' hypothesis. *Nat. Med.* **1999**, *5*, 1071–1075. [[CrossRef](#)] [[PubMed](#)]
52. Margraf, R.L.; VanSant-Webb, C.; Mao, R.; Viskochil, D.H.; Carey, J.; Hanson, H.; D'Astous, J.; Grossmann, A.; Stevenson, D.A. NF1 Somatic Mutation in Dystrophic Scoliosis. *J. Mol. Neurosci.* **2019**, *68*, 11–18. [[CrossRef](#)]