OPEN ACCESS

ORIGINAL ARTICLE

Distal chromosome 16p11.2 duplications containing *SH2B1* in patients with scoliosis

Brooke Sadler,[1] Gabe Haller,[2] Lilian Antunes,[1] Xavier Bledsoe,[1] Jose Morcuende,[3] Philip Giampietro,[4] Cathleen Raggio,[5] Nancy Miller,[6] Yared Kidane,[7] Carol A. Wise,[7] Ina Amarillo,[8] Nephi Walton,[9] Mark Seeley,[9] Darren Johnson,[9] Conner Jenkins,[9] Troy Jenkins,[9] Matthew Oetjens,[9] R. Spencer Tong,[10] Todd E Druley,[10] Matthew B. Dobbs,[2] Christina A. Gurnett[11]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2018-105877).

For numbered affiliations see end of article.

Correspondence to
Dr Christina A. Gurnett, Department of Neurology, Division of Pediatric Neurology, Washington University in St. Louis School of Medicine, St. Louis MO USA; gurnettc@wustl.edu

Received 15 November 2018
Revised 18 January 2019
Accepted 25 January 2019

To cite: Sadler B, Haller G, Antunes L, *et al*. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmedgenet-2018-105877

ABSTRACT

Introduction Adolescent idiopathic scoliosis (AIS) is a common musculoskeletal disorder with strong evidence for a genetic contribution. CNVs play an important role in congenital scoliosis, but their role in idiopathic scoliosis has been largely unexplored.

Methods Exome sequence data from 1197 AIS cases and 1664 in-house controls was analysed using coverage data to identify rare CNVs. CNV calls were filtered to include only highly confident CNVs with >10 average reads per region and mean log-ratio of coverage consistent with single-copy duplication or deletion. The frequency of 55 common recurrent CNVs was determined and correlated with clinical characteristics.

Results Distal chromosome 16p11.2 microduplications containing the gene *SH2B1* were found in 0.7% of AIS cases (8/1197). We replicated this finding in two additional AIS cohorts (8/1097 and 2/433), resulting in 0.7% (18/2727) of all AIS cases harbouring a chromosome 16p11.2 microduplication, compared with 0.06% of local controls (1/1664) and 0.04% of published controls (8/19584) ($p=2.28\times10^{-11}$, OR=16.15). Furthermore, examination of electronic health records of 92 455 patients from the Geisinger health system showed scoliosis in 30% (20/66) patients with chromosome 16p11.2 microduplications containing *SH2B1* compared with 7.6% (10/132) of controls ($p=5.6\times10^{-4}$, OR=3.9).

Conclusions Recurrent distal chromosome 16p11.2 duplications explain nearly 1% of AIS. Distal chromosome 16p11.2 duplications may contribute to scoliosis pathogenesis by directly impairing growth or by altering expression of nearby genes, such as *TBX6*. Individuals with distal chromosome 16p11.2 microduplications should be screened for scoliosis to facilitate early treatment.

INTRODUCTION

Scoliosis is defined as a spinal deformity consisting of a 10° or greater lateral curvature of the spine using the Cobb measurement method on a standing radiograph.[1] Known causes of scoliosis include congenital abnormalities, neuromuscular and connective tissue disorders. However, the majority of scoliosis cases are idiopathic (~80%). Adolescent idiopathic scoliosis (AIS) develops during late childhood and affects up to 3% of the population, with females disproportionately experiencing more severe scoliosis than males.[2] Twin studies show high rates of concordance for AIS and increased risk in first-degree relatives.[3] Genetic studies of common variants have shed important insight into the aetiology of AIS and include associations with variants near basonuclin-2,[4] adherens junctions-associated protein 1, BCL2 apoptosis regulator, ladybird homeobox 1 antisense RNA 1,[5] ladybird homeobox 1,[6] neural cell adhesion molecules,[7] paired homeobox 1[8] and G-protein-coupled receptor 126,[9] but these variants explain only a small fraction of the heritability of AIS. With the advances of sequencing technology, rare genetic variants are also being shown to contribute to AIS risk. Rare variants in *FBN1* and *FBN2*, which underlie Marfan syndrome and congenital contractural arachnodactyly, along with musculoskeletal collagen genes, were recently shown to be enriched in AIS cases and contribute to scoliosis severity.[10 11]

In addition to single nucleotide variants, CNVs resulting in large-scale duplications or deletions can also influence disease risk through alteration of gene dosage.[12] CNVs contribute to a wide range of disorders, including autism, schizophrenia and neuropsychiatric disorders[13 14] and influence bone mineral density[15] and numerous Mendelian and complex traits.[16] Scoliosis and other skeletal abnormalities were observed among patients with proximal chromosome 16p11.2 CNVs ascertained during evaluation for developmental disabilities.[17 18] Furthermore, proximal chromosome 16p11.2 deletions containing *TBX6* are strongly associated with congenital scoliosis.[19–21]

In the single prior study of CNVs in AIS, recurrent CNVs were identified in several of the 143 cases studied, suggesting a possible role in AIS pathogenesis.[22] However, this study was too small to determine a role for CNVs in AIS. In the current study, we performed a comprehensive, large-scale study to determine the contribution of recurrent CNVs to AIS risk by analysing exome sequence data from 1197 AIS cases and 1664 controls.

MATERIALS AND METHODS

Cohort description

Patients with juvenile or AIS were recruited from St. Louis Children's Hospital, St. Louis Shriners Hospital for Children, University of Iowa, University of Colorado, Hospital for Special Surgery and University of Wisconsin-Madison. All patients and/or parents provided informed consent. Inclusion criteria was spinal curve of >10°, although most

patients had severe scoliosis with a mean curve of 50°. Patients with developmental delay, multiple congenital anomalies or known or suspected underlying disorders, including Marfan syndrome or Ehlers-Danlos syndrome, were excluded from the study.

Exome sequencing

Sequencing data for 1197 AIS samples was generated at the McDonnell Genome Institute (MGI) using IDT xGen Exome Panel V1 capture on Illumina HiSeq 4000 paired-end reads. The St Louis replication cohort of 443 AIS samples were captured using Agilent reagents and sequenced on HiSeq 1500/2000. For controls, we used MGI in-house control data from 1664 samples that included 860 male infertility cases, which were captured and sequenced using the identical IDT platform from AIS cases, and 637 amyotrophic lateral sclerosis cases and 167 multiple congenital anomaly cases captured using Agilent reagents and sequenced on HiSeq 1500/2000 (dbGAP accession number: phs001677).

Exome alignment and BAM file processing

The sequencing reads (ie, fastq files) from all AIS cases and controls samples were aligned to the human genome reference (GRCh37) using BWA-MEM (V.0.7.15). The resulting BAM files were sorted, and duplicates were marked using Picard MarkDuplicates (V.2.9.0). Realignment intervals for each BAM file was determined using GATK (V.3.5) RealignerTargetCreator using a list of known indel sites (Mills and 1kg indels data from the GATK resource bundle ftp://ftp.broadinstitute.org/bundle/b37/), and local realignment and base quality recalibration were then performed by GATK (V.3.5) IndelRealigner and BaseRecalibrator, respectively.

Exome CNV analysis

Copy number analysis was performed on AIS cases and control samples using the software package FishingCNV[23] that is designed to identify rare CNVs from exome-sequencing data without the need for a paired control. This program uses an algorithm to prioritise rare variants and compares coverage depth in a test sample against the background distribution, as well as principle components analysis (PCA) to remove batch effects. The program extracts coverage information from GATK processed BAM files,[24] then normalises read depth information at each exon[25] into reads per kilobase per million mapped reads (RPKM) files. To identify only the most confident CNV calls, we included only CNVs with an average RPKM of at least 10, the mean log-ratio coverage between background and sample of −1.2 to −0.8 (consistent with a deletion) or 0.4 to 0.8 (consistent with a duplication) and Holm-Bonferroni adjusted p value less than 10^{-3}.

To systematically assess our samples for known recurrent CNVs, we used a size overlap threshold of 40% to call known recurrent CNVs. This lower threshold was used because the exome sequence often does not cover the entire CNV and because FishingCNV occasionally breaks a CNV into smaller segments, particularly for larger CNVs. Data were also manually reviewed to identify CNVs that were broken into smaller segments and therefore removed by the above filter.

CNV analysis in replication cohort

Raw intensity data were obtained from 1390 subjects using the Illumina HumanCoreExome-24 V.1.0 BeadChip. CNVs were called using PennCNV.[26] B allele frequency (BAF) and log R ratio (LRR) values were computed in Genome Studio. Data were exported out and processed using the 'split_illumina_report.pl' script in PennCNV to obtain BAF and LRR values per SNP per subject. Population frequency of B allele (PFB) and Gcmodel data sets were derived using the 'compile_pfb' and 'cal_gc_snp' Perl modules in PennCNV, respectively.

After CNV calling, we merged CNVs if the gap between adjacent calls was ≤50% of the combined length. Then, we performed initial quality control checks as recommended by PennCNV. Subjects were excluded from downstream analysis if they had a LRR SD of >0.28, a BAF drift of >0.01 or an absolute waviness factor of >0.05. Then, we excluded spurious CNV calls pertaining to telomeric and centromeric regions. CNVs that overlapped with immunoglobulin encoding regions were also removed. We retained 1339 subjects for further analysis after quality control. Of these, 1097 were unique from the original dataset.

CNV validation

For validation of CNVs using qPCR, we designed three primer pairs to exonic sequence within the CNV. Primers were tested for quality and efficiency using PCR and RT-qPCR standard curves, respectively. Quality control (QC) was performed on the tested DNA using the Qubit Fluorimeter (Invitrogen). qPCR is performed on a 96-well high-sided low-profile PCR plates (Thermo Scientific) on the Applied Biosystems StepOne Real-Time PCR System. Positive and negative controls with known gains or losses in the gene of interest were included in each run. The *RPL27* housekeeping gene was used as a reference.

Geisinger dataset

We assessed the exome sequencing data of 92 445 participants in the MyCode program, data that was generated as part of the Geisinger-Regeneron DiscovEHR project.[27] We used copy number estimation using Lattice-Aligned Mixture Models software for calling CNVs from the exome data.[28] We identified all individuals in the DiscovEHR cohort with duplications of the chromosome 16p11.2 region that contained the *SH2B1* gene. Manual chart review was performed on all the patients with *SH2B1* duplications to assess for scoliosis. All images were reviewed by an orthopaedic surgeon who assessed them for scoliosis and calculated Cobb angles. Patients were identified as having scoliosis if they met any of the following criteria: (1) mention of significant scoliosis in notes or in an imaging report, (2) scoliosis was identified on the problem list or (3) imaging of the spine that showed a Cobb angle >10°.

RESULTS

To systematically identify the frequency of 55 clinically relevant recurrent CNVs described in Coe *et al*[29] in AIS, the discovery cohort data consisting of 1197 AIS cases and 1664 in-house controls was assessed for CNVs with >40% overlap in size with known recurrent CNVs (table 1).

A total of 8 AIS cases with distal chromosome 16p11.2 duplications were detected (online supplementary table S1). The smallest duplication was detected by manual review, because it did not pass the initial filter criteria due to minimal overlap with the known CNV (only 11%). Other than the distal chromosome 16p11.2 duplication, there were no other recurrent CNVs that were statistically enriched in the AIS cohort. Identical analysis of our in-house control samples revealed only one distal chromosome 16p11.2 duplication out of 1664 controls (table 2).

Table 1 Recurrent microdeletions and microduplications seen in our adolescent idiopathic scoliosis (AIS) cases

		Deletion				Duplication			
		Cases	Controls			Cases	Controls		
		AIS	Current study	Coe *et al*[29]		AIS	Current study	Coe *et al*[29]	
CNV	Chr:Position (hg19)	n=1197	n=1664	n=19 584	P value	n=1197	n=1664	n=19 584	P value
1q21.1	1:146573376–147393376	0	0	6	1.00	1	0	5	0.42
2q13	2:111383531–113093529	3	0	3	0.07	0	0	0	1.00
15q11.2	15:22798636–23088559	2	2	27	0.56	0	2	60	1.00
15q13.3	15:31132708–32482708	0	0	0	1.00	1	0	11	0.42
16p13.11	16:15502499–16292499	1	2	7	0.80	1	1	27	0.66
Distal 16p11.2	16:28772499–29112499	0	1	1	1.00	8	1	8	***0.005***
Proximal 16p11.2	16:29652499–30202499	1	0	6	0.42	0	0	9	1.00
HNPP/ CMT1A	17:14069275–15499275	1	1	8	0.66	0	0	5	1.00
17q12	17:34815887–36205887	1	0	2	0.42	1	1	3	0.66
DiGeorge/ VCFS	22:19020000–20290000	2	0	0	0.18	0	0	12	1.00

All distal chromosome 16p11.2 duplications were validated by qPCR. A large published study of CNVs identified this distal chromosome 16p11.2 duplication in only 8 of 19 584 controls.[29] Therefore, the distal chromosome 16p11.2 duplication was significantly enriched in our cases compared with both in-house controls (p=0.005, OR=8.55) and published controls ($p=1.05\times10^{-6}$, OR=16.34) even after Bonferonni correction for multiple tests.

Replication of chromosome 16p11.2 duplications in AIS

Two replication cohorts were assayed for chromosome 16p11.2 duplications containing the *SH2B1* gene to replicate the discovery cohort results. Replication cohort 1 consisted of 1097 AIS cases who were genotyped using the Illumina HumanCoreExome-24 V.1.0 BeadChip and had CNVs called using PennCNV.[26] Replication cohort 2 consisted of 433 AIS cases with exome data that were generated using Agilent reagents and were analysed for the presence of CNVs separately from the discovery cohort, which was captured using IDT probes. Eight AIS cases with the distal chromosome 16p11.2 duplication were identified in replication cohort 1 (n=1097). Two additional cases were identified in replication cohort 2 (n=433). When the discovery and replication cohorts were combined, the distal chromosome 16p11.2 duplication was present in 0.7% cases (18/2727) compared with 0.04% (8/19584) controls ($p=2.28\times10^{-11}$, OR=16.15) (table 2).

The locations of the distal chromosome 16p11.2 duplications that included *SH2B1* were mapped onto the genome (figure 1).

Table 2 Enrichment of distal chromosome 16p11.2 duplication in adolescent idiopathic scoliosis (AIS) cases

AIS cases		In-house controls	P value	Published controls[29]	P value
Discovery (IDT)	8/1197 (0.7%)	1/1664 (0.06%)	0.005	8/19584 (0.04%)	1.05×10^{-6}
Replication 1 (Texas)	8/1097 (0.9%)	0/746 (0.0%)	0.01	8/19584 (0.04%)	5.64×10^{-7}
Replication 2 (St. Louis)	2/433 (0.4%)	–	–	8/19584 (0.04%)	0.02
Combined	18/2727 (0.7%)	1/2410 (0.04%)	0.0001	8/19584 (0.04%)	2.28×10^{-11}

The chromosome 16p11.2 region is highly enriched for segmental microduplications and microdeletions with known breakpoints defining these regions (figure 1). The proximal 600 kb region (29.5–31.1 Mb) is defined by BP4-BP5 and has previously been associated with congenital scoliosis,[19] while the distal BP2-BP3 region (28.7–28.9 Mb) has not previously been associated with a skeletal phenotype. The duplications ranged from 80 to 340 exons in length and were consistent in size with known CNVs mediated by segmental duplications in this region. One duplication was slightly smaller than the typical BP2-BP3 CNV and involved only 80 exons in four genes, including *SH2B1*. These duplications were near, but did not include, the proximal chromosome 16p11.2 deletion (BP4-BP5). The proximal deletion was seen only once in the AIS cohort, shown in red on figure 1; however, further review revealed a subtle vertebral segmentation anomaly consistent with congenital scoliosis. Therefore, if we exclude this case, there were no AIS cases with proximal chromosome 16p11.2 deletions, suggesting no or minimal overlap with this genetic cause of congenital scoliosis. While the eight duplications identified in the replication cohort 1 appear slightly larger, their size was determined based on genotype data, rather than exome data.

Clinical phenotypes associated with chromosome 16p11.2 duplications

Clinical evaluations of the AIS cases with the distal chromosome 16p11.2 duplication revealed that all but two are women, which is consistent with the strong gender bias in severe AIS cases. The mean body mass index (BMI) of cases with the chromosome 16p11.2 duplication was lower than those without the duplication but the difference was not significant (21.1 vs 22.2) (p=0.17) (online supplementary table S2). The mean maximum curve severity in individuals with and without the chromosome 16p11.2 duplication was not different. As developmental disability was an exclusion for the study, all of the cases had normal neurodevelopment. One case had ADHD. Beighton scores of joint hypermobility[30] and Ghent systemic feature scores[31] were unremarkable in four cases but were not measured in others.

Analysis of scoliosis in chromosome 16p11.2 duplication carriers in Geisinger dataset

Finally, we examined a dataset of 92 445 patients from Geisinger Health System with exome data and discovered 66 patients with

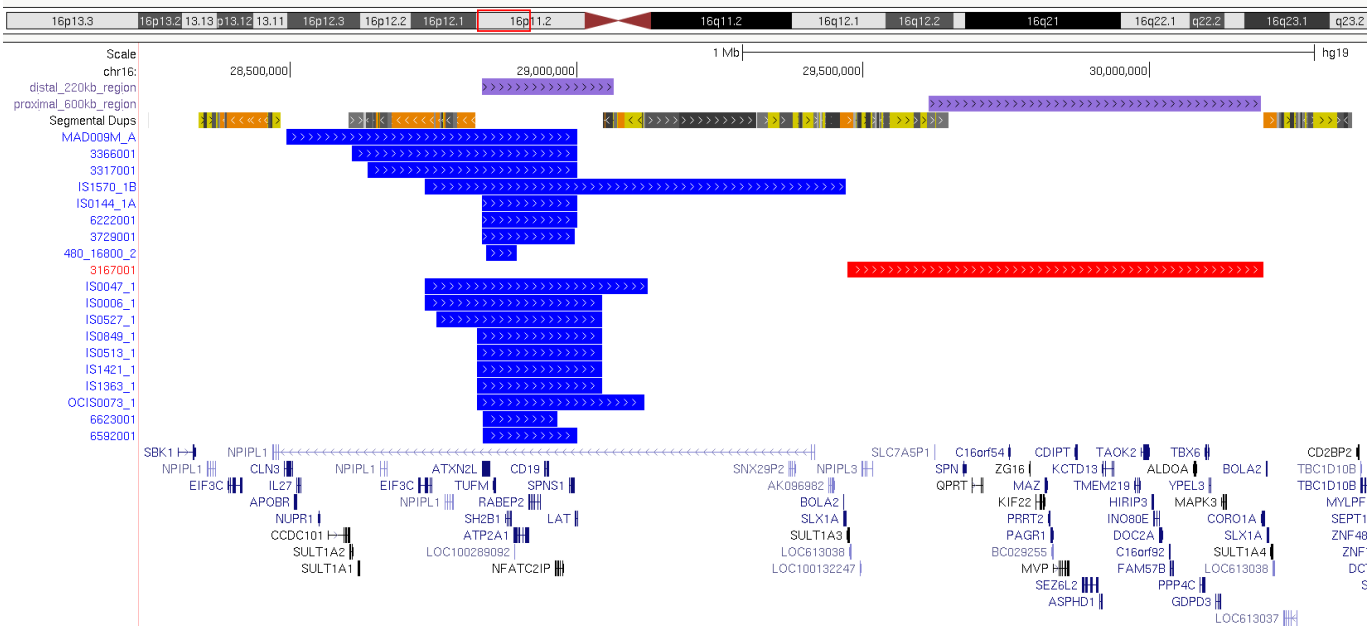

Figure 1 Locations of distal chromosome 16p11.2 duplications. Legend: locations of distal chromosome 16p11.2 duplications containing SH2B1 (blue) and proximal chromosome 16p11.2 deletions containing TBX6 (red) in our AIS cohort of 1197 patients and in replication cohort 1 (1097 AIS cases) and replication cohort 2 (433 AIS cases). Previously discovered locations of segmental duplications (orange) are shown along with locations of break points (BPs) resulting in recurrent proximal and distal CNV locations (purple). AIS, adolescent idiopathic scoliosis.

microduplications that include *SH2B1*. Of these 66 patients, 20 had evidence for scoliosis by imaging and/or report, suggesting a 30% lifetime risk of scoliosis. Their mean age was 58 years and 16 (80%) were women. Nineteen of the 20 patients had healthcare visits specifically for back pain and three out of 20 had back surgery for scoliosis. Of the 46 patients who were not classified as having scoliosis, 31 had imaging that included at least a section of the spine and showed no scoliosis or minor scoliosis (Cobb angle <10°) and 15 patients did not have any imaging that included the spine. One patient reported having scoliosis when younger but showed no scoliosis on current imaging, and two other patients had mild scoliosis on an imaging report from an outside hospital, but images were not available for assessment; therefore, all three were classified as not having scoliosis using the diagnostic criteria. In comparison, evaluation of 132 age-matched and sex-matched controls from the Geisinger dataset using the same criteria revealed that 7.6% controls (10/132) had scoliosis, suggesting an approximately fourfold increase in lifetime scoliosis risk in patients with the distal chromosome 16p11.2 duplication ($p=5.6\times10^{-4}$, OR=3.9).

DISCUSSION

Here we describe the occurrence of distal chromosome 16p11.2 duplications in nearly 1% of patients with AIS, a common paediatric musculoskeletal disorder whose aetiology is largely unexplained. The chromosome 16p11.2 region was previously proposed as a candidate locus for AIS based on linkage data.[32 33] This complex chromosomal region has undergone a recent, rapid integration of segmental duplications followed by adaptive evolution that has occurred since the split between the human/great ape lineage and orangutans, approximately 12 million years ago.[34] This region's richness in segmental duplications predisposes it to recurrent rearrangements.[35] The proximal 600 kb region (29.5–30.1 Mb) defined by breakpoints 4–5 (BP4-BP5; OMIM #611913) harbours the *TBX6* gene and has been previously associated with congenital scoliosis,[19] while the distal 220 kb BP2-BP3 region (28.7–28.9 Mb) reported here in AIS has not been previously associated with any skeletal phenotype. Deletion of the proximal region is associated with autism spectrum disorder, obesity and macrocephaly,[36] and deletion of the distal region has similarly associated phenotypes.[37] Developmental delay is associated with the distal chromosome 16p11.2 deletion and, to a lesser degree, with the duplication.[38 39] Distal chromosome 16p11.2 duplication (OMIM #614671) carriers also have lower BMI and reduced head circumference.[37] This distal region includes the *SH2B1* gene, which is involved in leptin and insulin signalling. Variants in *SH2B1* are associated with behavioural abnormalities and obesity[40] and deletion of which is associated with severe obesity, hyperphagia and insulin resistance,[41] suggesting that *SH2B1* is the major gene driving these growth phenotypes.

The complex relationship between the proximal and distal chromosome 16p11.2 regions is exemplified by data demonstrating that perturbation of the distal region may impact expression of genes in the proximal region, including *TBX6*.[37] The two regions are reciprocally engaged in complex chromatin looping, bringing together regulatory elements and genes, which is disturbed in carriers of the proximal CNV. The observed duplication could also be affecting nearby topologically associated domains and, as such, have an effect on expression and coregulation of more distant genes in the region. Thus, duplications of distal chromosome 16p11.2 may influence scoliosis risk by indirectly altering *TBX6* expression. However, a role for *TBX6* is considered less likely because none of the AIS cases with the distal chromosome 16p11.2 duplication had evidence for structural vertebral anomalies that are hallmarks of congenital scoliosis associated with *TBX6* dysregulation.[19] Likewise, we detected only one proximal chromosome 16p11.2 deletion that includes *TBX6* in our large cohort, and this patient should have been excluded from our study as he had subtle congenital scoliosis on closer review. Additional support against a role for *TBX6* in AIS pathogenesis is also provided by prior work showing the

lack of segregating *TBX6* coding variants in idiopathic scoliosis families.[42]

While the distal chromosome 16p11.2 duplication may influence scoliosis risk by indirectly altering nearby genes such as *TBX6*, direct involvement of single genes within the interval may also be causative. The smallest duplication interval in our AIS cases included only *ATXN2L*, Tu translation elongation factor, mitochondrial (*TUFM*), *SH2B1* and ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 1 (*ATP2A1*). *SH2B1* is a strong candidate due to its previous associations with growth, obesity and BMI. The distal duplication is also associated with low BMI.[37] When combined with the observation that patients with AIS are significantly more likely be underweight than healthy controls,[43] the enrichment of this duplication in our sample is intriguing and suggests the possibility that genetic abnormalities that alter growth, such as the distal chromosome 16p11.2 duplication, directly influence either susceptibility to scoliosis or its progression. Notably, *SH2B1* is one of only two genes in this particular CNV with a probability of loss intolerance (pLI) score of greater than 0.9,[44] indicating a strong dependence on gene dosage. The other gene with a high pLI score is *ATXN2L* (ataxin 2 like), which encodes a protein of unknown function. Other genes in the region include *TUFM* , which is associated with human combined oxidative phosphorylation deficiency, and *ATP2A1*, an enzyme involved in muscular excitation and contraction. Delineation of the exact mechanism by which the chromosome 16p11.2 microduplication influences scoliosis risk will require additional study.

To grossly estimate the penetrance of scoliosis associated with this distal chromosome 16p11.2 microduplication, we used a genotype-first approach to analyse a very large cohort of patients in the Geisinger Health System database with paired exome data and found a 30% overall lifetime risk of scoliosis, which represents a >4-fold lifetime risk of scoliosis compared with controls from the same healthcare system. In comparison, prior studies determined that this microduplication is mildly associated with cognitive impairment, with penetrance estimates of only ~11%.[45] Incomplete penetrance of CNVs associated with neuropsychiatric disorders and cognitive impairment is well established, and we have confirmed this is also true for scoliosis. However, chromosome 16p11.2 microduplications appear to be more strongly associated with scoliosis than any other phenotype.

Limitations of the present study include the propensity of the program to fragment large CNVs into many smaller CNVs, as well as potential coverage biases, both of which relate to the use of exome sequence data to call CNVs rather than whole genome or chromosomal microarray data. The size of our cohort is relatively small compared with association studies of autism and other neuropsychiatric disorders, which often used ~20 000 samples to identify associations due to the low frequency of these CNVs. Of note, the distal chromosome 16p11.2 duplication is a significant risk factor that has an OR that is much higher than any other previously described for AIS, which explains why we were able to detect it with this relatively modest sample size. In addition to the distal chromosome 16p11.2 duplication, we also identified a small number of other recurrent CNVs in AIS cases (online supplementary table S3), although due to their rarity, the numbers are too small to statistically confirm enrichment in AIS. Our numbers are also likely too small to determine whether the distal chromosome 16p11.2 duplication influences clinical factors such as curve severity. While we did not detect a difference in scoliosis curve severity in cases with and without the distal chromosome 16p11.2 duplication, our cohort consists predominantly of severe AIS cases, and therefore, future studies with a broader representation of curve severity in cases are needed to determine if this CNV has any prognostic role in predicting curve progression.

While it is not yet standard of care to perform genetic testing in otherwise healthy patients with AIS, discovery of this distal chromosome 16p11.2 duplication as an AIS risk factor in nearly 1% of patients will add to the benefits of genetic testing for this patient population, which may also include identifying associated aortic aneurysm disease gene variants,[11] anaesthesia and surgical bleeding risk factors and determination of recurrence risk in family members. For now, the results of this study support scoliosis screening for individuals identified with distal chromosome 16p11.2 microduplications in order to increase the utilisation and effectiveness of early treatment interventions such as bracing.

Author affiliations
[1]Department of Neurology, Washington University in Saint Louis School of Medicine, St. Louis, Missouri, USA
[2]Department of Orthopedic Surgery, Washington University in Saint Louis School of Medicine, St. Louis, Missouri, USA
[3]Department of Orthopaedic Surgery and Rehabilitation, University of Iowa Roy J and Lucille A Carver College of Medicine, Iowa City, Iowa, USA
[4]Department of Genetics, St. Christopher's Hospital for Children, Philadelphia, Pennsylvania, USA
[5]Orthopedic Surgery, Pediatrics, Hospital for Special Surgery, New York City, New York, USA
[6]Department of Orthopedics, University of Colorado at Denver - Anschutz Medical Campus, Aurora, Colorado, USA
[7]Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, Texas, USA
[8]Department of Pathology and Immunology, Washington University in Saint Louis School of Medicine, St. Louis, Missouri, USA
[9]Genomic Medicine, Geisinger Health System, Danville, Pennsylvania, USA
[10]Department of Pediatrics, Washington University in Saint Louis School of Medicine, St. Louis, Missouri, USA
[11]Department of Neurology, Division of Pediatric Neurology, Washington University in St. Louis School of Medicine, St. Louis, Missouri, USA

Acknowledgements We would like to thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by National Cancer Institute (NCI) Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by the Institute for Clinical and Translational Sciences (ICTS) / Clinical and Translational Science Awards Program (CTSA) Grant #UL1RR024992 from the National Center for Research Resources, a component of the National Institutes of Health (NIH) and NIH Roadmap for Medical Research. Funded with support from University of Missouri Spinal Cord Injury Research Program, Shriners Hospital for Children and the Children's Discovery Institute of Washington University and St. Louis Children's Hospital, and Hope Center DNA/RNA Purification Core at Washington University School of Medicine. Computations were performed using the facilities of the Washington University Center for High Performance Computing, which were partially funded by NIH grants 1S10RR022984-01A1 and 1S10OD018091-01. We would like to thank the patients and their families for their participation, as well as Drs Munish Gupta, Keith Bridwell, Mike Kelly, Scott Luhmann, Brian Kelly, Luke Zebala, Lawrence Lenke and Christi Abeln.

Contributors BS, GH, MD and CAG conceived of the study; BS and GH carried out statistical analyses, LA compiled exome data; XB compiled supplemental table data; CAW, YK, JM, PG, CR and NM compiled and provided additional exome data from University of Iowa, University of Colorado and University of Texas; IA provided additional advice on CNV location methods; NW, MS, DJ, CJ, TJ and MO provided and analysed Geisinger dataset for CNVs and phenotypic data, and ST and TD validated CNVs. All authors edited the manuscript.

Funding Research reported in this publication was supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases under Award Number R01AR067715, Eunice Kennedy Shriver National Institutes of Child Health and Human Development of the National Institutes of Health under the Award Number P01HD084387, the Marfan Foundation Faculty Grant award #81831, Washington University Institute of Clinical and Translational Sciences grant UL1 TR002345 from the National Center for Advancing Translational Sciences of the National Institutes of Health, Washington University Musculoskeletal Research Center (NIH/NIAMS P30 AR057235) and the Eunice Kennedy Shriver National Institute of Child Health &

Human Development of the National Institutes of Health under Award Number U54 HD087011 to the Intellectual and Developmental Disabilities Research Center at Washington University.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests None declared.

Ethics approval The institutional review board at each institution approved this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Author note Accession number for the AIS exome data reported in this paper are dbGAP: phs001677.

REFERENCES

1 Cobb J. Outline for the Study of Scoliosis. *Instru Course Lect* 1948;5:261–5.
2 Lenke L, Dobbs M. *The Adult and Pediatric Spine*. Philadelphia: PA: Lippincott, Williams & Wilkins, 2004.
3 Gorman KF, Julien C, Moreau A. The genetic epidemiology of idiopathic scoliosis. *Eur Spine J* 2012;21:1905–19.
4 Ogura Y, Kou I, Miura S, Takahashi A, Xu L, Takeda K, Takahashi Y, Kono K, Kawakami N, Uno K, Ito M, Minami S, Yonezawa I, Yanagida H, Taneichi H, Zhu Z, Tsuji T, Suzuki T, Sudo H, Kotani T, Watanabe K, Hosogane N, Okada E, Iida A, Nakajima M, Sudo A, Chiba K, Hiraki Y, Toyama Y, Qiu Y, Shukunami C, Kamatani Y, Kubo M, Matsumoto M, Ikegawa S. A Functional SNP in BNC2 Is Associated with Adolescent Idiopathic Scoliosis. *Am J Hum Genet* 2015;97:337–42.
5 Zhu Z, Tang NL, Xu L, Qin X, Mao S, Song Y, Liu L, Li F, Liu P, Yi L, Chang J, Jiang L, Ng BK, Shi B, Zhang W, Qiao J, Sun X, Qiu X, Wang Z, Wang F, Xie D, Chen L, Chen Z, Jin M, Han X, Hu Z, Zhang Z, Liu Z, Zhu F, Qian BP, Yu Y, Wang B, Lee KM, Lee WY, Lam TP, Qiu Y, Cheng JC. Genome-wide association study identifies new susceptibility loci for adolescent idiopathic scoliosis in Chinese girls. *Nat Commun* 2015;6:8355.
6 Takahashi Y, Kou I, Takahashi A, Johnson TA, Kono K, Kawakami N, Uno K, Ito M, Minami S, Yanagida H, Taneichi H, Tsuji T, Suzuki T, Sudo H, Kotani T, Watanabe K, Chiba K, Hosono N, Kamatani N, Tsunoda T, Toyama Y, Kubo M, Matsumoto M, Ikegawa S. A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. *Nat Genet* 2011;43:1237–40.
7 Sharma S, Gao X, Londono D, Devroy SE, Mauldin KN, Frankel JT, Brandon JM, Zhang D, Li QZ, Dobbs MB, Gurnett CA, Grant SF, Hakonarson H, Dormans JP, Herring JA, Gordon D, Wise CA. Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. *Hum Mol Genet* 2011;20:1456–66.
8 Sharma S, Londono D, Eckalbar WL, Gao X, Zhang D, Mauldin K, Kou I, Takahashi A, Matsumoto M, Kamiya N, Murphy KK, Cornelia R, Group TSC. Japan Scoliosis Clinical Research G, Herring JA, Burns D, Ahituv N, Ikegawa S, Gordon D, Wise CA. A PAX1 enhancer locus is associated with susceptibility to idiopathic scoliosis in females. *Nat Commun* 2015;6:6452.
9 Kou I, Takahashi Y, Johnson TA, Takahashi A, Guo L, Dai J, Qiu X, Sharma S, Takimoto A, Ogura Y, Jiang H, Yan H, Kono K, Kawakami N, Uno K, Ito M, Minami S, Yanagida H, Taneichi H, Hosono N, Tsuji T, Suzuki T, Sudo H, Kotani T, Yonezawa I, Londono D, Gordon D, Herring JA, Watanabe K, Chiba K, Kamatani N, Jiang Q, Hiraki Y, Kubo M, Toyama Y, Tsunoda T, Wise CA, Qiu Y, Shukunami C, Matsumoto M, Ikegawa S. Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nat Genet* 2013;45:676–9.
10 Haller G, Alvarado D, Mccall K, Yang P, Cruchaga C, Harms M, Goate A, Willing M, Morcuende JA, Baschal E, Miller NH, Wise C, Dobbs MB, Gurnett CA. A polygenic burden of rare variants across extracellular matrix genes among individuals with adolescent idiopathic scoliosis. *Hum Mol Genet* 2016;25:202–9.
11 Haller G, Alvarado DM, Willing MC, Braverman AC, Bridwell KH, Kelly M, Lenke LG, Luhmann SJ, Gurnett CA, Dobbs MB. Genetic Risk for Aortic Aneurysm in Adolescent Idiopathic Scoliosis. *J Bone Joint Surg Am* 2015;97:1411–7.
12 Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. *Nature* 2006;444:444–54.
13 Cook EH, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature* 2008;455:919–23.
14 Morrow EM. Genomic copy number variation in disorders of cognitive development. *J Am Acad Child Adolesc Psychiatry* 2010;49:1091–104.
15 Chew S, Dastani Z, Brown SJ, Lewis JR, Dudbridge F, Soranzo N, Surdulescu GL, Richards JB, Spector TD, Wilson SG. Copy number variation of the APC gene is associated with regulation of bone mineral density. *Bone* 2012;51:939–43.
16 Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med* 2010;61:437–55.
17 Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA, Probst FJ, Craigen WJ, Graham BH, Pursley A, Clark G, Lee J, Proud M, Stocco A, Rodriguez DL, Kozel BA, Sparagana S, Roeder ER, McGrew SG, Kurczynski TW, Allison LJ, Amato S, Savage S, Patel A, Stankiewicz P, Beaudet AL, Cheung SW, Lupski JR. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J Med Genet* 2010;47:332–41.
18 Al-Kateb H, Khanna G, Filges I, Hauser N, Grange DK, Shen J, Smyser CD, Kulkarni S, Shinawi M. Scoliosis and vertebral anomalies: additional abnormal phenotypes associated with chromosome 16p11.2 rearrangement. *Am J Med Genet A* 2014;164A:1118–26.
19 Wu N, Ming X, Xiao J, Wu Z, Chen X, Shinawi M, Shen Y, Yu G, Liu J, Xie H, Gucev ZS, Liu S, Yang N, Al-Kateb H, Chen J, Zhang J, Hauser N, Zhang T, Tasic V, Liu P, Su X, Pan X, Liu C, Wang L, Shen J, Shen J, Chen Y, Zhang T, Zhang J, Choy KW, Wang J, Wang Q, Li S, Zhou W, Guo J, Wang Y, Zhang C, Zhao H, An Y, Zhao Y, Wang J, Liu Z, Zuo Y, Tian Y, Weng X, Sutton VR, Wang H, Ming Y, Kulkarni S, Zhong TP, Giampietro PF, Dunwoodie SL, Cheung SW, Zhang X, Jin L, Lupski JR, Qiu G, Zhang F. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *N Engl J Med* 2015;372:341–50.
20 Lefebvre M, Duffourd Y, Jouan T, Poe C, Jean-Marçais N, Verloes A, St-Onge J, Riviere JB, Petit F, Pierquin G, Demeer B, Callier P, Thauvin-Robinet C, Faivre L, Thevenon J. Autosomal recessive variations of TBX6, from congenital scoliosis to spondylocostal dysostosis. *Clin Genet* 2017;91:908–12.
21 Takeda K, Kou I, Kawakami N, Iida A, Nakajima M, Ogura Y, Imagawa E, Miyake N, Matsumoto N, Yasuhiko Y, Sudo H, Kotani T, Nakamura M, Matsumoto M, Watanabe K, Ikegawa S. Japan Early Onset Scoliosis Research Group. Compound Heterozygosity for Null Mutations and a Common Hypomorphic Risk Haplotype in TBX6 Causes Congenital Scoliosis. *Hum Mutat* 2017;38:317–23.
22 Buchan JG, Alvarado DM, Haller G, Aferol H, Miller NH, Dobbs MB, Gurnett CA. Are copy number variants associated with adolescent idiopathic scoliosis?. *Clin Orthop Relat Res* 2014;472:3216–25.
23 Shi Y, Majewski J. FishingCNV: a graphical software package for detecting rare copy number variations in exome-sequencing data. *Bioinformatics* 2013;29:1461–2.
24 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
25 Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008;5:621–8.
26 Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007;17:1665–74.
27 Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, O'Dushlaine C, Van Hout CV, Staples J, Gonzaga-Jauregui C, Metpally R, Pendergrass SA, Giovanni MA, Kirchner HL, Balasubramanian S, Abul-Husn NS, Hartzel DN, Lavage DR, Kost KA, Packer JS, Lopez AE, Penn J, Mukherjee S, Gosalia N, Kanagaraj M, Li AH, Mitnaul LJ, Adams LJ, Person TN, Praveen K, Marcketta A, Lebo MS, Austin-Tse CA, Mason-Suares HM, Bruse S, Mellis S, Phillips R, Stahl N, Murphy A, Economides A, Skelding KA, Still CD, Elmore JR, Borecki IB, Yancopoulos GD, Davis FD, Faucett WA, Gottesman O, Ritchie MD, Shuldiner AR, Reid JG, Ledbetter DH, Baras A, Carey DJ. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* 2016;354:aaf6814.
28 Packer JS, Maxwell EK, O'Dushlaine C, Lopez AE, Dewey FE, Chernomorsky R, Baras A, Overton JD, Habegger L, Reid JG. CLAMMS: a scalable algorithm for calling common and rare copy number variants from exome sequencing data. *Bioinformatics* 2016;32:133–5.
29 Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, Friend KL, Baker C, Buono S, Vissers LE, Schuurs-Hoeijmakers JH, Hoischen A, Pfundt R, Krumm N, Carvill GL, Li D, Amaral D, Brown N, Lockhart PJ, Scheffer IE, Alberti A, Shaw M, Pettinato R, Tervo R, de Leeuw N, Reijnders MR, Torchia BS, Peeters H, O'Roak BJ, Fichera M, Hehir-Kwa JY, Shendure J, Mefford HC, Haan E, Gécz J, de Vries BB, Romano C, Eichler EE. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet* 2014;46:1063–71.
30 Beighton P, Horan F. Orthopaedic aspects of the ehlers-danlos syndrome. *J Bone Joint Surg Br* 1969;51:444–53.
31 Loeys BL, Dietz HC, Braverman AC, Callewaert BL, De Backer J, Devereux RB, Hilhorst-Hofstee Y, Jondeau G, Faivre L, Milewicz DM, Pyeritz RE, Sponseller PD, Wordsworth P, De Paepe AM. The revised Ghent nosology for the Marfan syndrome. *J Med Genet* 2010;47:476–85.
32 Miller NH, Justice CM, Marosy B, Doheny KF, Pugh E, Zhang J, Dietz HC, Wilson AF. Identification of candidate regions for familial idiopathic scoliosis. *Spine* 2005;30:1181–7.

33 Miller NH, Justice CM, Marosy B, Swindle K, Kim Y, Roy-Gagnon MH, Sung H, Behneman D, Doheny KF, Pugh E, Wilson AF. Intra-familial tests of association between familial idiopathic scoliosis and linked regions on 9q31.3-q34.3 and 16p12.3-q22.2. *Hum Hered* 2012;74:36–44.
34 Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, Eichler EE. Positive selection of a gene family during the emergence of humans and African apes. *Nature* 2001;413:514–9.
35 Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S, Bader PI, Hamati A, Reitnauer PJ, Smith R, Stockton DW, Muhle H, Helbig I, Eichler EE, Ballif BC, Rosenfeld J, Tsuchiya KD. Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genet Med* 2010;12:641–7.
36 Migliavacca E, Golzio C, Männik K, Blumenthal I, Oh EC, Harewood L, Kosmicki JA, Loviglio MN, Giannuzzi G, Hippolyte L, Maillard AM, Alfaiz AA, van Haelst MM, Andrieux J, Gusella JF, Daly MJ, Beckmann JS, Jacquemont S, Talkowski ME, Katsanis N, Reymond A. 16p11.2 European Consortium. A Potential Contributory Role for Ciliary Dysfunction in the 16p11.2 600 kb BP4-BP5 Pathology. *Am J Hum Genet* 2015;96:784–96.
37 Loviglio MN, Leleu M, Männik K, Passeggeri M, Giannuzzi G, van der Werf I, Waszak SM, Zazhytska M, Roberts-Caldeira I, Gheldof N, Migliavacca E, Alfaiz AA, Hippolyte L, Maillard AM, Van Dijck A, Kooy RF, Sanlaville D, Rosenfeld JA, Shaffer LG, Andrieux J, Marshall C, Scherer SW, Shen Y, Gusella JF, Thorsteinsdottir U, Thorleifsson G, Dermitzakis ET, Deplancke B, Beckmann JS, Rougemont J, Jacquemont S, Reymond A. 2p15 Consortium16p11.2 Consortium. Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes. *Mol Psychiatry* 2017;22:836–49.
38 D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L, Hanson E, Maillard AM, Faucett WA, Mace A, Pain A, Bernier R, Chawner SJ, David A, Andrieux J, Aylward E, Baujat G, Caldeira I, Conus P, Ferrari C, Forzano F, Gerard M, Goin-Kochel RP, Grant E, Hunter JV, Isidor B, Jacquette A, Jonch AE, Keren B, Lacombe D, Le Caignec C, Martin CL, Mannik K, Metspalu A, Mignot C, Mukherjee P, Owen MJ, Passeggeri M, Rooryck-Thambo C, Rosenfeld JA, Spence SJ, Steinman KJ, Tjernagel J, Van Haelst M, Shen Y, Draganski B, Sherr EH, Ledbetter DH, van den Bree MB, Beckmann JS, Spiro JE, Reymond A, Jacquemont S, Chung WK. Cardiff University Experiences of Children With Copy Number Variants S,. p11.2 European C, Simons Variation in Individuals Project C. Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. *JAMA psychiatry* 2016;73:20–30.
39 Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L, Macé A, Ferrari C, Kutalik Z, Andrieux J, Aylward E, Barker M, Bernier R, Bouquillon S, Conus P, Delobel B, Faucett WA, Goin-Kochel RP, Grant E, Harewood L, Hunter JV, Lebon S, Ledbetter DH, Martin CL, Männik K, Martinet D, Mukherjee P, Ramocki MB, Spence SJ, Steinman KJ, Tjernagel J, Spiro JE, Reymond A, Beckmann JS, Chung WK, Jacquemont S. Simons VIP Consortium16p11.2 European Consortium. A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J Med Genet* 2012;49:660–8.
40 Doche ME, Bochukova EG, Su HW, Pearce LR, Keogh JM, Henning E, Cline JM, Saeed S, Dale A, Cheetham T, Barroso I, Argetsinger LS, O'Rahilly S, Rui L, Carter-Su C, Farooqi IS. Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *J Clin Invest* 2012;122:4732–6.
41 Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, Saeed S, Hamilton-Shield J, Clayton-Smith J, O'Rahilly S, Hurles ME, Farooqi IS. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 2010;463:666–70.
42 Baschal EE, Swindle K, Justice CM, Baschal RM, Perera A, Wethey CI, Poole A, Pourquié O, Tassy O, Miller NH. Sequencing of the *TBX6* Gene in Families with Familial Idiopathic Scoliosis. *Spine Deform* 2015;3:288–96.
43 Lonner BS, Toombs CS, Husain QM, Sponseller P, Shufflebarger H, Shah SA, Samdani AF, Betz RR, Cahill PJ, Yaszay B, Newton PO. Body Mass Index in Adolescent Spinal Deformity: Comparison of Scheuermann's Kyphosis, Adolescent Idiopathic Scoliosis, and Normal Controls. *Spine Deform* 2015;3:318–26.
44 Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnström K, Mallick S, Kirby A, Wall DP, MacArthur DG, Gabriel SB, DePristo M, Purcell SM, Palotie A, Boerwinkle E, Buxbaum JD, Cook EH, Gibbs RA, Schellenberg GD, Sutcliffe JS, Devlin B, Roeder K, Neale BM, Daly MJ. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014;46:944–50.
45 Rosenfeld JA, Coe BP, Eichler EE, Cuckle H, Shaffer LG. Estimates of penetrance for recurrent pathogenic copy-number variations. *Genet Med* 2013;15:478–81.