Published in final edited form as:

Spine Deform. 2015 July ; 3(4): 288–296. doi:10.1016/j.jspd.2015.01.005.

Sequencing of the *TBX6* Gene in Families with Familial Idiopathic Scoliosis

Erin E. Baschal^a, Kandice Swindle^a, Cristina M. Justice^b, Robin M. Baschal^c, Anoja Perera^d, Cambria I. Wethey^a, Alex Poole^a, Olivier Pourquié^e, Olivier Tassy^e, and Nancy H. Miller^{a,c,*}

^a Department of Orthopedics, University of Colorado Denver Anschutz Medical Campus, 13001 E 17th Place, Aurora, Colorado 80045, USA

^b Genometrics Section, National Human Genome Research Institute, National Institutes of Health, 333 Cassell Drive, Baltimore, Maryland 21224, USA

^c Musculoskeletal Research Center, Children's Hospital Colorado, 13123 East 16th Avenue, Aurora, Colorado, 80045

^d Molecular Biology Facility, Stowers Institute, 1000 E. 50th Street, Kansas City, Missouri 64110, United States

^e Department of Cell Biology and Development, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), 1 rue Laurent Fries, BP 10142, 67404 Illkirch CEDEX, France

Abstract

Study Design—A hypothesis-driven study was conducted in a familial cohort to determine the potential association between variants within the *TBX6* gene and Familial Idiopathic Scoliosis (FIS).

Objective—To determine if variants within exons of the *TBX6* gene segregate with the FIS phenotype within a sample of families with FIS.

Summary of Background Data—Idiopathic Scoliosis (IS) is a structural curvature of the spine whose underlying genetic etiology has not been established. IS has been reported to occur at a higher rate than expected in family members of individuals with congenital scoliosis (CS),

* **Correspondence:** Nancy H. Miller, University of Colorado Denver Anschutz Medical Campus, 12800 E. 19th Ave., MS8343, Aurora, CO 80045 Tel: 1-303-724-0046, Fax: 1-303-734-0394, Nancy.Hadley-Miller@ucdenver.edu.

Erin E. Baschal-Erin.Baschal@ucdenver.edu

Kandice Swindle-kandiceswindle@hotmail.com

Cristina M. Justice-cmj@mail.nih.gov

Robin M. Baschal-Robin.Baschal@childrenscolorado.org

Anoja Perera-AGP@stowers.org

Cambria I. Wethey-Cambria.Wethey@ucdenver.edu

Alex Poole-Alex.Poole@ucdenver.edu

Olivier Pourquié-pourquie@igbmc.fr

Oliver Tassy-otassy@igbmc.fr

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

suggesting that the two diseases might have a shared etiology. The *TBX6* gene on chromosome 16p, essential to somite development, has been associated with CS in a Chinese population. Previous studies have identified linkage to this locus in families with FIS, and specifically with rs8060511, located in an intron of the *TBX6* gene.

Methods—Parent-offspring trios from 11 families (13 trios, 42 individuals) with FIS were selected for Sanger sequencing of the *TBX6* gene. Trios were selected from a large population of families with FIS in which a genome-wide scan had resulted in linkage to 16p.

Results—Sequencing analyses of the subset of families resulted in the identification of five coding variants. Three of the five variants were novel; the remaining two variants were previously characterized and account for 90% of the observed variants in these trios. In all cases, there was no correlation between transmission of the *TBX6* variant allele and FIS phenotype. However, an analysis of regulatory markers in osteoblasts showed that rs8060511 is in a putative enhancer element.

Conclusions—Although this study did not identify any *TBX6* coding variants that segregate with FIS, we identified a variant that is located in a potential *TBX6* enhancer element. Therefore, further investigation of the region is needed.

Keywords

TBX6 gene; idiopathic scoliosis; congenital scoliosis; chromosome 16p11.2; DNA sequencing

1. INTRODUCTION

As one of the most common inherited disorders of the axial skeleton, idiopathic scoliosis (IS) affects 2% to 3% of otherwise normal children and adolescents. On spinal radiographs, a structural lateral rotatory curvature of the spine is present without visual vertebral anomalies. The familial nature of IS is well established, and multiple loci and genes have been implicated in IS etiology in the published literature [1-9].

Axial skeletal formation in humans is the result of highly synchronized developmental molecular pathways in the process known as somitogenesis. T-box genes, members of a highly phylogenetically conserved gene family, are widely involved in these developmental processes. The T-box gene family encodes transcription factors characterized by a highly conserved N-terminal DNA-binding domain known as the T-box, whereas the C-terminus maintains far less sequence conservation. These transcription factors are intimately involved in the regulation of early mesoderm specification, somite formation, and left/right (L/R) body symmetry [10]. The brachyury (*T*) gene and the T-box 6 (*TBX6*) gene, members of the T-box gene family, have been targeted as potential genetic factors related to spinal abnormalities in humans due to their involvement in the somite patterning of vertebrate animal models.

Disturbances in axial skeletal development can result in ‘congenital’ scoliosis (CS), characterized by a spinal curvature that results from vertebral malformations visible on spinal radiographs, or in ‘idiopathic’ scoliosis, characterized by a lateral spinal curvature without visible vertebral malformations or any known developmental or neuromuscular

etiology. Clinically, deformities of vertebral growth manifesting as scoliosis become evident predominantly as a child matures and gains trunk stability. Clinical studies and studies in model organisms support the role of an underlying genetic contribution to both CS and IS to varying degrees [11,12]. CS is usually sporadic with occasional familial clusters. In comparison, genetic influences are clearly supported in IS through multiple studies, particularly in relation to families [4,13-15]. Purkiss et al. analyzed 237 families in which the proband had CS, and reported that 17.3% of these individuals had family members with IS, representing a much higher than expected rate of IS compared to the general population [12]. This observation raises the question of shared underlying genetic mechanisms or interactions that can result in susceptibility to both CS and IS.

A recent study has shown a significant association between genetic variants in the *TBX6* gene and CS in a Chinese population [16]. A mutation in *TBX6*, which results in haploinsufficiency, has been identified as a cause of spondylocostal dysostosis (SCD), a severe syndromic form of congenital vertebral malformations [17]. Through both linkage and association studies in a large FIS population, our own work has identified linkage to chromosome 16p, a locus which contains the *TBX6* gene [14,18]. In addition, the 16p11.2 region encompasses several structural variants [19,20]. There is a higher incidence of scoliosis and vertebral anomalies in patients with recurrent 16p11.2 rearrangements [21,22]. Given both the association between *TBX6* genetic variants and CS and the relationship between CS and IS, the current study utilizes traditional Sanger sequencing technology to examine the *TBX6* gene in a subset of 13 trios with FIS.

2. MATERIALS AND METHODS

2.1 Study Populations

Written informed consent was obtained from an initial study population of 202 families collected in accordance with a protocol approved by the Johns Hopkins School of Medicine Institutional Review Board. Families had at least two members with idiopathic scoliosis as ascertained by a single orthopedic surgeon [14,23]. Criteria for a diagnosis of scoliosis were clinical/family history and physical examination determining a spinal curvature in the coronal plane, standing anteroposterior spinal radiographs exhibiting 10° sagittal curvature by the Cobb method, and pedicle rotation with no evidence of congenital deformity or other genetic conditions [24-26].

The study population of 202 families was stratified according to likely mode of inheritance, as previously published [14]. A subset of families was identified as most likely exhibiting an autosomal dominant (AD) form of FIS inheritance (95 families, 552 individuals). Linkage and subsequent fine-mapping studies for chromosome 16p were completed in this subset of families. From this group of 95 families with a likely AD form of FIS, parent-offspring trios from 11 families (consisting of 13 trios, plus 3 unaffected siblings, for a total of 42 individuals) were selected for sequencing of the *TBX6* gene utilizing Sanger sequencing methodology. Selection criteria required that each trio contain both an affected proband and an affected parent, each with a scoliotic curvature of 25° by radiographic analyses, as well as an unaffected parent. Unaffected status was confirmed with radiographs. Three of the families had an additional unaffected sibling. The final study population consisted of 26

affected individuals (18 female, 8 male) and 16 unaffected individuals (8 female, 8 male). This subset underwent linkage analyses of fine-mapping single nucleotide polymorphisms (SNPs) on 16p to ensure that it was representative of the larger group prior to sequencing of the *TBX6* gene.

2.2 Initial Genomic Screen and Fine-Mapping SNP Panel

Blood samples were collected from all participants and genomic DNA was extracted according to standard purification protocols [27,28]. A genomic screen for 202 families was performed at the Center for Inherited Disease Research with a modified CHLC v.9 marker set consisting of 391 short tandem repeat markers, as previously published [14]. Linkage analyses identified a significant candidate region for FIS on 16p in a subset of the initial population identified as most likely exhibiting an AD form of FIS inheritance (95 families, 552 individuals) [14]. Custom SNP fine-mapping panels were generated from National Center for Biotechnology Information (NCBI) dbSNP database (Build 34) to cover the previously identified region of interest on 16p (174 SNPs, spanning approximately 17 megabases (Mb) from 18.1 Mb to 34.8 Mb, dbSNP minor allele frequencies of 0.25) [18]. Genotyping was performed on the Illumina BeadArray Platform and evaluated with Illumina's Gentrain software.

2.3 *TBX6* gene: PCR and Sequencing

Sequencing efforts were directed to the *TBX6* gene (chr16: 30,097,117 - 30,103,205 base pairs), lying in the linkage region on 16p11.2. The gene was sequenced in a stepwise fashion to conserve genomic DNA. First, PCR was performed to amplify the entire *TBX6* gene. Second, multiple PCR reactions were performed on the amplified *TBX6* DNA from step 1.

Amplification of the entire *TBX6* gene, from genomic DNA, was completed using TaKaRa LA Taq™ (Mg²⁺ plus Buffer, Takara Bio Inc) according to a company provided protocol. Primers were determined utilizing the human reference sequence from Ensembl (<http://www.ensembl.org/index.html>, Genome Assembly GRCh37) (**Supplemental Table 1**).

The PCR reaction for *TBX6* gene amplification consisted of 1X buffer, 200 μM dNTPs, 0.2 μM of each primer, 2.5 units of TaKaRa LA Taq™, and 200 ng of genomic DNA. Amplification conditions were as follows: 1 min at 94°C, followed by 30 cycles of 98°C for 10 sec, 68°C for 15 min, with a final extension of 72°C for 10 min. PCR products were separated on an agarose gel stained with ethidium bromide, and a DNA ladder was used to verify correct fragment sizes.

Following whole gene amplification, smaller, internal PCRs were set up, utilizing Failsafe PCR enzyme and buffer D from Epicentre Biotechnologies (**Supplemental Table 1**). These sequences contained the M13-21F and M13-24R primer tags (tgtaaacgacggccagt and aacagctatgaccatg, respectively) for sequencing purposes. The internal PCR Failsafe reaction consisted of: 1X Failsafe buffer, 1X Failsafe PCR Enzyme mix and 1 μL DNA (*TBX6* whole gene PCR product) in a total reaction volume of 15 μL. Amplification conditions were as follows: 95°C for 10 min, followed by 30 cycles of 95°C for 30 sec, 60°C for 30 sec, and 72°C for 1 min, with a final extension of 72°C for 10 min. Post-PCR cleanup was performed

using Agencourt Ampure SPRI reagents (Agencourt, Beverly, MA, USA) and eluted in 30 μ L of ddH₂O. A 10 μ L cleanup reaction utilized BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA), 0.5 μ L DMSO, 0.32 μ M sequencing primer (M13-21F or M13-24R), and 2 μ L DNA (post internal PCR product). Conditions for the sequencing reaction were as follows: 25 cycles of 96°C for 10 sec, 50°C for 5 sec, and 60°C for 4 min. Post-sequencing amplification purification was performed using Agencourt CleanSEQ SPRI reagents (Agencourt, Beverly, MA, USA) and eluted in 25 μ L of 0.01 mM EDTA. Products were then Sanger sequenced on an Applied Biosystems 3730x1 DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

2.4 Statistical Methods and Bioinformatic Analyses

In the initial genomic screen, model-independent linkage analyses were performed with SIBPAL (S.A.G.E. version 4.5) to screen for linkage between the trait and each marker in the AD subgroup [14,29]. Fine-mapping data generated with custom SNP panels was evaluated utilizing model-independent multipoint linkage analyses and association analyses including SIBPAL (S.A.G.E. version 6.2.0) and FBAT, for the 95 families determined to be most likely to exhibit an AD mode of inheritance and also for the subset of 13 trios chosen for further sequencing [30-32].

Sequence data were analyzed using the Phrap/Consed tool from the laboratory of Phil Green (<http://www.phrap.org/>), SeqScape software (Applied Biosystems, Foster City, CA, USA), and CodonCode Alignment software (version 3.7.1.1; CodonCode Corporation). Reads were aligned to a segment of chromosome 16 of the GRCh37 human reference assembly (NC_000016.9). Variants were called manually after alignment and then cross-referenced with dbSNP build 132. Variants were considered valid only if they could be verified by at least two sequencing reads for the same individual.

Haploview (version 4.2) was used to measure linkage disequilibrium between SNPs, using HapMap data (HapMap Version 3, Release R2, CEU+TSI) [33]. The University of California Santa Cruz (UCSC) Genome Browser was used to visualize SNP locations and sequenced areas [34], and custom tracks were created using the UCSC Table Browser [35,36]. All positions are reported using the GRCh37 human reference genome assembly. Regulatory regions were investigated using the publicly available ENCODE data [37]. These results were visualized through the WashU Epigenome Browser matplot function [38,39] and detailed data were accessed using the UCSC Table Browser [35,36]. The ENCODE data that we present used normal human osteoblasts, and were generated by the Bernstein laboratory at the Broad Institute [37].

3. RESULTS

3.1 SNP Fine-Mapping

Initial genomic screen data was analyzed for a 95 family subset (552 individuals) of the original group, those families determined to be most likely to exhibit an autosomal dominant (AD) mode of inheritance [14]. These results identified a significant region on chromosome 16p. Fine-mapping of this region was completed in this AD subset, using custom SNP

panels (**Supplemental Table 2**) [18]. When the affectation status was defined as 20° of lateral spinal curvature, multipoint linkage analyses identified multiple SNPs (27.83 Mb to 31.32 Mb) that were significant at $p = 0.001$. The most significant SNP was rs6565183 ($p=0.00059$), located at 30.38 Mb, slightly upstream of the *TBX6* gene (**Figure 1A**). Analyses were also completed using affectation status defined as 30° of lateral spinal curvature. Using this definition, rs6565183 was still significant ($p=0.00270$). A second SNP, rs8060511, which lies within an intron of the *TBX6* gene, was also significant at both thresholds ($p=0.00183$, affectation status of 20°; $p=0.00673$, 30°). Two SNPs in this region, both of which lie outside the *TBX6* gene, were significant by FBAT association (rs208600 at 22.96 Mb, $p=0.00644$ and rs4889606 at 31.01 Mb, $p=0.00290$).

From this group of 95 families with a likely AD form of FIS, 11 families (consisting of 13 trios, plus 3 unaffected siblings, for a total of 42 individuals) were selected for sequencing of portions of the *TBX6* gene utilizing Sanger sequencing methodology. Clinical details for the 42 individuals are provided in **Supplemental Table 3**. Within this small subset of 42 individuals, linkage analyses of the fine-mapping data identified significant SNPs in the region. The intronic SNP, rs8060511, was significant at both the 20° and 30° curve thresholds ($p=0.00489$ and $p=0.04413$, respectively). In addition, the most significant SNP in the larger group, rs6565183, upstream of the *TBX6* gene, was also significant in this group ($p=0.00410$ for 20° and $p=0.04443$ for 30°). We also investigated the linkage disequilibrium (LD) pattern across the linkage region. The two main SNPs of interest, rs8060511 and rs6565183, are not in significant LD with each other in the HapMap CEU + TSI data ($D'=0.244$) (**Supplemental Figure 1**).

Using the fine-mapping SNP results, the negative log of the p-value was plotted for the 95 families determined to be most likely to exhibit an AD mode of inheritance. The same analysis was completed for the 13 trios chosen for sequencing (**Figure 2**). The peaks are similar, indicating that the families selected for sequencing are a representative subset of the larger group of 95 families.

3.2 Sequencing

The *TBX6* gene, located on 16p11.2 (chr16: 30,097,117 to 30,103,205 base pairs), was sequenced in 11 families (13 trios, 42 individuals). All *TBX6* exons were sequenced, including surrounding intronic DNA (**Figure 1B**, black boxes indicate regions sequenced in this study). Five distinct sequence variants were identified within this study population, four in coding exons, and one in an intron (**Table 1**). At least one variant was identified in all 13 trios and 29 of the 42 individuals. Three of the five sequence variants are novel, whereas two variants have been previously characterized (rs3833842 and rs2289292). Only two of the four coding variants result in amino acid changes (Gly395Ser and Tyr421His). These were both found in unaffected parents from unrelated trios, and were not transmitted to probands in either case. Thus, there is no correlation between transmission of the risk alleles for any of the five variants and FIS phenotype.

3.3 Regulatory Elements

We investigated whether our SNPs of interest were located in putative regulatory elements, using publicly available ENCODE data (**Table 2** and **Supplemental Table 4**). We analyzed data generated in normal human osteoblasts, as this cell type could potentially be relevant for FIS. One SNP, rs8060511, is in a potential regulatory element within an intron of *TBX6* (**Figure 3A**). Specifically, in the ENCODE data, the peak in the region surrounding rs8060511 is significantly associated with multiple markers of open chromatin and putative enhancer elements (**Table 2**), including H3K4me1, H3K4me2, H2A.Z, and H3K27ac. Additionally, the most significant SNP in the linkage peak, rs6565183, is upstream of the *TBC1D10B* gene, and is also located in a region of promoter/enhancer signature (H3K27ac, H3K4me1, H2A.Z, H3K4me2, H3K4me3, H3K79me2, and P300) (**Figure 3B** and **Table 2**).

4. DISCUSSION

The *TBX6* gene on chromosome 16p11.2, related to somite development critical to the axial skeleton, was targeted for sequencing in a select group of families with familial idiopathic scoliosis (FIS). Previous studies suggest that the locus on 16p is a candidate region within a subgroup of families affected with FIS [14,18]. Although multiple *TBX6* variants were identified in the families sequenced in this study, none of the *TBX6* sequence variants segregated with the FIS phenotype.

The 16p11.2 locus encompasses several unique genomic structural variants, including a recurrent interstitial deletion, microduplications, and copy number variants [19,20,40]. These structural variants have been shown to relate to cognitive function, behavior, and neurodevelopmental disorders such as autism and schizophrenia. Additionally, several reports identify an increased incidence of scoliosis and vertebral anomalies in patients with 16p11.2 rearrangements [21,22,40-44]. The 16p11.2 rearrangements in these patients with scoliosis and/or vertebral anomalies overlap with our linkage peak and typically include the *TBX6* gene, but not our SNP that is upstream of *TBX6* (rs6565183) [21,22,40-44]. The complexity of the 16p11.2 region contributes to the difficulty in sequencing the complete *TBX6* gene. Several repetitive elements lie within *TBX6* introns, therefore, only a portion of the intronic regions were sequenced in this study (**Figure 1B**) [45].

In this study, we did not identify sequence variants in the *TBX6* gene that segregated with the FIS phenotype. However, these findings do not necessarily rule out the *TBX6* gene as a contributing factor to the etiology of FIS. Causative variants could potentially be located in a region of the gene that was not sequenced in this study, such as an intron or a regulatory element. As mentioned, rs8060511, in an intron of *TBX6*, is located within a putative enhancer element that is present in normal human osteoblasts. In addition, the most significant SNP in the fine-mapping results, rs6565183, is located 279,550 base pairs upstream of the *TBX6* gene. This variant is located in a large haplotype block (**Supplemental Figure 1**), and could be in LD with a causative variant that affects *TBX6* gene expression. Additionally, this region contains other potential candidate genes, including *CDH11* (cadherin) and *STX1B* (syntaxin 1B). The CDH11 protein functions as a regulator of postnatal skeletal growth and bone mass maintenance [46]. STX1B is a protein

associated with intracellular nerve synaptic complexes. This gene is of particular interest given recent work that relates genes within axonal guidance pathways as potentially related to the etiology of idiopathic scoliosis [8].

As opposed to 'idiopathic' scoliosis, 'congenital' scoliosis (CS) is defined by gross segmentation abnormalities resulting in vertebral malformations that can be identified on spinal radiographs. Studies support the concept that the process of vertebral segmentation is controlled in somitogenesis by a molecular oscillator, known as the segmentation clock, where coordinated pulses of highly phylogenetically conserved genes result in the precursors of the axial skeleton [12,47,48]. To date, mutations associated with CS are in genes that are related to the pathways of the segmentation clock. IS, also identified as a 'heritable' disease, is present at a higher than expected rate within family members of probands with CS [12]. This observation introduces the possibility that the genetic etiology of IS represents a subtle deficiency of the synchronized pathway of somitogenesis, which remains latent until exacerbated by the structural and biochemical environment introduced at the onset of pubertal growth.

The T-box gene family is a group of genes that encode transcription factors involved in the developmental process of somatogenesis. Zebrafish and mouse animal models have shown that the T-box gene family is essential to the specification and separation of paraxial mesoderm structures for somite formation. Mutant phenotypes have linked T-box genes to the Wnt and Notch signaling pathways through various target genes and feedback loops implicated in the early determination of embryonic laterality [47-53]. In humans, the T-box gene family has been associated with congenital vertebral malformations (CVMs). Ghebranious et al. sequenced the *T* (brachyury) gene and the *TBX6* gene in a cohort of 50 patients with CVMs [54]. A *T* gene missense variant was associated with an increase in CVM risk within the study population, but this allele alone was not sufficient to cause CVM.

Fei et al. reported a positive association between two SNPs within the *TBX6* gene and individual susceptibility to congenital scoliosis in a Chinese Han population [16]. One SNP is within an intron between the first and second exons (rs3809624), and the second SNP is in exon 8 (rs2289292). The authors suggest that within their select population the *TBX6* gene acts as a predisposition gene for congenital scoliosis. rs2289292 was also identified in our sequencing, but there was no association with disease status in our study population. Most recently, a mutation in *TBX6*, which results in loss of a stop codon and haploinsufficiency, has been identified as a cause of autosomal dominant spondylocostal dysostosis (SCD), a severe syndromic form of CVMs [17]. In addition, mice with *Tbx6* mutations have skeletal anomalies, such as rib fusions, spinal arch fusions, and vertebral body irregularities [55,56].

In summary, transmission of sequence variants within the *TBX6* gene, located in a candidate region on 16p, was not correlated with the FIS phenotype in this select familial study population. However, rs8060511, in an intron of *TBX6*, was significant in our fine mapping study and is located in a putative enhancer element that is present in normal human osteoblasts, a cell type that is could potentially be relevant in FIS. A variant that contributes to the FIS phenotype in these families could be located in this putative regulatory element and in LD with rs8060511, or could be in another region of the gene that was not sequenced

in this study. Future studies could look at the correlation between rs8060511 alleles and *TBX6* expression levels. In addition, there are several other candidate genes in the region. Further investigation of the *TBX6* gene and additional genes within this region is required in order to delineate their potential roles in the etiology of FIS and axial skeletal development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Our research was supported, in part, by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health. Some of the results were obtained with the program S.A.G.E., which is supported by grant 1 P41 RR03655 from the National Center for Research Resources. Our research used the REDCap database, which is supported by NIH/NCRR Colorado CTSI Grant UL1 TR000154. Additional grant funding was provided by the Institute de France Foundation Yves Cotrel, the LARRK Foundation and National Institutes of Health grant R01-AR048862-04. We would like to thank Keith Smith and Kendra Walton at Stowers Institute for their sequencing efforts, and Jay Hesselberth and Kenneth L. Jones for their assistance with interpreting the data on putative regulatory elements.

LIST OF ABBREVIATIONS

IS	Idiopathic Scoliosis
CS	Congenital Scoliosis
FIS	Familial Idiopathic Scoliosis
SCD	spondylocostal disostosis
AD	autosomal dominant
SNP	single nucleotide polymorphism
Mb	megabases
CVMs	congenital vertebral malformations
LD	linkage disequilibrium

REFERENCES

1. Bonaiti C, Feingold J, Briard ML, Lapeyre F, Rigault P, Guivarch J. [genetics of idiopathic scoliosis]. *Helv Paediatr Acta*. 1976; 31:229–40. [PubMed: 977374]
2. Cowell HR, Hall JN, MacEwen GD. Genetic aspects of idiopathic scoliosis. A nicholas andry award essay, 1970. *Clinical orthopaedics and related research*. 1972; 86:121–31. [PubMed: 5047777]
3. Czeizel A, Bellyei A, Barta O, Magda T, Molnar L. Genetics of adolescent idiopathic scoliosis. *J Med Genet*. 1978; 15:424–7. [PubMed: 745214]
4. Riseborough EJ, Wynne-Davies R. A genetic survey of idiopathic scoliosis in boston, massachusetts. *J Bone Joint Surg Am*. 1973; 55:974–82. [PubMed: 4760104]
5. Miller NH, Marosy B, Justice CM, Novak SM, Tang EY, Boyce P, et al. Linkage analysis of genetic loci for kyphoscoliosis on chromosomes 5p13, 13q13.3, and 13q32. *Am J Med Genet A*. 2006; 140:1059–68. [PubMed: 16596674]
6. Gao X, Gordon D, Zhang D, Browne R, Helms C, Gillum J, et al. Chd7 gene polymorphisms are associated with susceptibility to idiopathic scoliosis. *American journal of human genetics*. 2007; 80:957–65. [PubMed: 17436250]

7. Kou I, Takahashi Y, Johnson TA, Takahashi A, Guo L, Dai J, et al. Genetic variants in *gpr126* are associated with adolescent idiopathic scoliosis. *Nature genetics*. 2013; 45:676–9. [PubMed: 23666238]
8. Sharma S, Gao X, Londono D, Devroy SE, Mauldin KN, Frankel JT, et al. Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. *Hum Mol Genet*. 2011; 20:1456–66. [PubMed: 21216876]
9. Takahashi Y, Kou I, Takahashi A, Johnson TA, Kono K, Kawakami N, et al. A genome-wide association study identifies common variants near *lhx1* associated with adolescent idiopathic scoliosis. *Nature genetics*. 2011; 43:1237–40. [PubMed: 22019779]
10. Wardle FC, Papaioannou VE. Teasing out t-box targets in early mesoderm. *Current opinion in genetics & development*. 2008; 18:418–25. [PubMed: 18778771]
11. Giampietro PF, Blank RD, Raggio CL, Merchant S, Jacobsen FS, Faciszewski T, et al. Congenital and idiopathic scoliosis: Clinical and genetic aspects. *Clin Med Res*. 2003; 1:125–36. [PubMed: 15931299]
12. Purkiss SB, Driscoll B, Cole WG, Alman B. Idiopathic scoliosis in families of children with congenital scoliosis. *Clinical orthopaedics and related research*. 2002;27–31. [PubMed: 12151879]
13. McKinley LM, Leatherman KD. Idiopathic and congenital scoliosis in twins. *Spine*. 1978; 3:227–9. [PubMed: 568825]
14. Miller NH, Justice CM, Marosy B, Doheny KF, Pugh E, Zhang J, et al. Identification of candidate regions for familial idiopathic scoliosis. *Spine*. 2005; 30:1181–7. [PubMed: 15897833]
15. Axenovich TI, Zaidman AM, Zorkoltseva IV, Tregubova IL, Borodin PM. Segregation analysis of idiopathic scoliosis: Demonstration of a major gene effect. *Am J Med Genet*. 1999; 86:389–94. [PubMed: 10494097]
16. Fei Q, Wu Z, Wang H, Zhou X, Wang N, Ding Y, et al. The association analysis of *tbx6* polymorphism with susceptibility to congenital scoliosis in a chinese han population. *Spine (Phila Pa 1976)*. 2010; 35:983–8. [PubMed: 20228709]
17. Sparrow DB, McInerney-Leo A, Gucev ZS, Gardiner B, Marshall M, Leo PJ, et al. Autosomal dominant spondylocostal dysostosis is caused by mutation in *tbx6*. *Hum Mol Genet*. 2013; 22:1625–31. [PubMed: 23335591]
18. Miller NH, Justice CM, Marosy B, Swindle K, Kim Y, Roy-Gagnon MH, et al. 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. [PubMed: 23154503]
19. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nature genetics*. 2009; 41:1223–7. [PubMed: 19855392]
20. Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L, et al. A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J Med Genet*. 2012; 49:660–8. [PubMed: 23054248]
21. Al-Kateb H, Khanna G, Filges I, Hauser N, Grange DK, Shen J, et al. Scoliosis and vertebral anomalies: Additional abnormal phenotypes associated with chromosome 16p11.2 rearrangement. *Am J Med Genet A*. 2014
22. Shimojima K, Inoue T, Fujii Y, Ohno K, Yamamoto T. A familial 593-kb microdeletion of 16p11.2 associated with mental retardation and hemivertebrae. *European journal of medical genetics*. 2009; 52:433–5. [PubMed: 19770079]
23. Justice CM, Miller NH, Marosy B, Zhang J, Wilson AF. Familial idiopathic scoliosis: Evidence of an x-linked susceptibility locus. *Spine*. 2003; 28:589–94. [PubMed: 12642767]
24. Armstrong GW, Livermore NB 3rd, Suzuki N, Armstrong JG. Nonstandard vertebral rotation in scoliosis screening patients. Its prevalence and relation to the clinical deformity. *Spine*. 1982; 7:50–4. [PubMed: 7071661]
25. Kane WJ. Scoliosis prevalence: A call for a statement of terms. *Clinical orthopaedics and related research*. 1977:43–6. [PubMed: 598138]
26. Shands AR Jr, Eisberg HB. 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 Joint Surg Am*. 1955; 37-A: 1243–9. [PubMed: 13271471]

27. Sambrook, J.; Fritsch, EF.; Maniatis, T. Molecular cloning : A laboratory manual. ed 2nd.. Cold Spring Harbor Laboratory Press; 1989.
28. Moore D, Dowhan D. Purification and concentration of DNA from aqueous solutions. Curr Protoc Mol Biol. 2002 Chapter 2:Unit 2 1A.
29. Comings DE, Gade-Andavolu R, Gonzalez N, Wu S, Muhleman D, Blake H, et al. Multivariate analysis of associations of 42 genes in adhd, odd and conduct disorder. Clin Genet. 2000; 58:31–40. [PubMed: 10945659]
30. Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered. 2000; 50:211–23. [PubMed: 10782012]
31. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: Application to asthma genetics. Genet Epidemiol. 2004; 26:61–9. [PubMed: 14691957]
32. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. Genet Epidemiol. 2000; 19(Suppl 1):S36–42. [PubMed: 11055368]
33. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of ld and haplotype maps. Bioinformatics. 2005; 21:263–5. [PubMed: 15297300]
34. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at ucsc. Genome Res. 2002; 12:996–1006. [PubMed: 12045153]
35. Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, et al. The ucsc genome browser database: 2014 update. Nucleic acids research. 2014; 42:D764–70. [PubMed: 24270787]
36. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The ucsc table browser data retrieval tool. Nucleic acids research. 2004; 32:D493–6. [PubMed: 14681465]
37. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
38. Zhou X, Maricque B, Xie M, Li D, Sundaram V, Martin EA, et al. The human epigenome browser at washington university. Nature methods. 2011; 8:989–90. [PubMed: 22127213]
39. Zhou X, Wang T. Using the wash u epigenome browser to examine genome-wide sequencing data. Current protocols in bioinformatics / editorial board, Andreas D Baxeavanis [et al]. 2012 Chapter 10:Unit10.
40. Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA, et al. 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. [PubMed: 19914906]
41. Bijlsma EK, Gijbbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, Franssen van de Putte DE, Anderlid BM, et al. Extending the phenotype of recurrent rearrangements of 16p11.2: Deletions in mentally retarded patients without autism and in normal individuals. European journal of medical genetics. 2009; 52:77–87. [PubMed: 19306953]
42. Fernandez BA, Roberts W, Chung B, Weksberg R, Meyn S, Szatmari P, et al. Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. J Med Genet. 2010; 47:195–203. [PubMed: 19755429]
43. Schaaf CP, Goin-Kochel RP, Nowell KP, Hunter JV, Aleck KA, Cox S, et al. Expanding the clinical spectrum of the 16p11.2 chromosomal rearrangements: Three patients with syringomyelia. Eur J Hum Genet. 2011; 19:152–6. [PubMed: 20959866]
44. Shen Y, Chen X, Wang L, Guo J, Shen J, An Y, et al. Intra-family phenotypic heterogeneity of 16p11.2 deletion carriers in a three-generation chinese family. Am J Med Genet B Neuropsychiatr Genet. 2011; 156:225–32. [PubMed: 21302351]
45. Smit A, Hubley R, Green P. Repeatmasker open-3.0. 1996-2010
46. Di Benedetto A, Watkins M, Grimston S, Salazar V, Donsante C, Mbalaviele G, et al. N-cadherin and cadherin 11 modulate postnatal bone growth and osteoblast differentiation by distinct mechanisms. Journal of cell science. 2010; 123:2640–8. [PubMed: 20605916]
47. Griffin KJ, Kimelman D. One-eyed pinhead and spadetail are essential for heart and somite formation. Nat Cell Biol. 2002; 4:821–5. [PubMed: 12360294]

48. Nakaya MA, Biris K, Tsukiyama T, Jaime S, Rawls JA, Yamaguchi TP. Wnt3a links left-right determination with segmentation and anteroposterior axis elongation. *Development*. 2005; 132:5425–36. [PubMed: 16291790]
49. Yamaguchi TP, Takada S, Yoshikawa Y, Wu N, McMahon AP. T (brachyury) is a direct target of wnt3a during paraxial mesoderm specification. *Genes & development*. 1999; 13:3185–90. [PubMed: 10617567]
50. Showell C, Binder O, Conlon FL. T-box genes in early embryogenesis. *Developmental dynamics : an official publication of the American Association of Anatomists*. 2004; 229:201–18. [PubMed: 14699590]
51. Naiche LA, Harrelson Z, Kelly RG, Papaioannou VE. T-box genes in vertebrate development. *Annu Rev Genet*. 2005; 39:219–39. [PubMed: 16285859]
52. Martin BL, Kimelman D. Regulation of canonical wnt signaling by brachyury is essential for posterior mesoderm formation. *Developmental cell*. 2008; 15:121–33. [PubMed: 18606146]
53. Chapman DL, Cooper-Morgan A, Harrelson Z, Papaioannou VE. Critical role for tbx6 in mesoderm specification in the mouse embryo. *Mech Dev*. 2003; 120:837–47. [PubMed: 12915233]
54. Ghebranious N, Blank RD, Raggio CL, Staubli J, McPherson E, Ivacic L, et al. A missense t (brachyury) mutation contributes to vertebral malformations. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2008; 23:1576–83.
55. Chapman DL, Papaioannou VE. Three neural tubes in mouse embryos with mutations in the t-box gene tbx6. *Nature*. 1998; 391:695–7. [PubMed: 9490412]
56. Watabe-Rudolph M, Schlautmann N, Papaioannou VE, Gossler A. The mouse ribvertebrae mutation is a hypomorphic tbx6 allele. *Mech Dev*. 2002; 119:251–6. [PubMed: 12464437]

Figure 1. Region Surrounding the *TBX6* Gene

Figure generated using the UCSC Genome Browser (<http://genome.ucsc.edu>) [34].

A: Locations of SNPs of interest surrounding the *TBX6* gene, in relation to known genes (“SNPs of Interest” track and “UCSC Genes” track).

B: *TBX6* gene, with exons denoted with blue boxes and introns with a line (“UCSC Genes” track). Black boxes indicate regions of the *TBX6* gene sequenced in this study (“Regions Sequenced in This Study” track).

Figure 2. Fine-Mapping Results for Chromosome 16p

Fine-mapping results using multipoint linkage analysis as described in the Methods. The negative log of the p-value is plotted for each SNP genotyped in this study [18].

Blue solid line: p-value results in all 95 families (552 individuals)

Red dashed line: p-value results in 11 families (13 trios, 42 individuals) selected for *TBX6* sequencing *TBX6* gene position is denoted with a green triangle. p-value thresholds are marked with green dotted lines.

Figure 3.

Plots of regulatory marks generated in normal human osteoblasts by ENCODE. These plots were generated using the WashU Epigenome Browser. Data for 12 regulatory marks is plotted (red=H3K27me3, green=H3K4me3, blue=H4K20me1, yellow=H3K9me3, pink=H3K4me1, teal=H2A.Z, brown=H3K27ac, olive=P300, light blue=H3K79me2, maroon=H3K36me3, orange=H3K4me2, grey=CTCF).

A: Region surrounding rs8060511, a SNP of interest that is located in an intron of the *TBX6* gene.

B: Region that includes both rs8060511 and rs6565183.

Table 1Summary of *TBX6* variants

Location (bp) ^a	Region in Gene	Nucleotide Change ^a	Amino Acid Change	Affected ^b		Unaffected ^b		dbSNP	Allele Frequencies in Population ^c
				Proband	Parent	Parent	Sibling		
30098190-30098192	Intron	TCA→deletion	N/A	5	6	3	2	rs3833842	deletion: 37.7%; TCA: 62.3%
30097699	Exon 9	T→C	Phe386Phe	1	0	0	0	-	-
30097674	Exon 9	G→C	Gly395Ser	0	0	1	0	-	-
30097630	Exon 9	G→A	Pro409Pro	3	5	3	0	rs2289292	G: 64.4%; A: 35.6%
30097596	Exon 9	T→C	Tyr421His	0	0	1	0	-	-

^aBase positions and allele designations correspond to (-) strand on NCBI GRCh37; reference allele → alternate allele

^bNumber (i.e. 5) is the number of individuals with the alternate allele for each variant (i.e. 5 affected probands had a TCA deletion). Note that an individual might have more than one variant and that variants were not identified in every individual.

^cFrequency in 1000 Genomes dataset (European populations)

Table 2

Analysis of Putative Regulatory Elements (rs8060511 and rs6565183)

Regulatory Mark	Potential Effect on Transcription ^a	rs8060511 peak p-value	rs6565183 peak p-value
CTCF	May function as an insulator, blocking enhancer activity	NS ^b	NS ^b
H2A.Z	"Histone protein variant (H2A.Z) associated with regulatory elements with dynamic chromatin"	7.94×10^{-12}	2.00×10^{-14}
H3K4me1	"Associated with enhancers and other distal elements, but also enriched downstream of transcription starts"	6.31×10^{-16}	1.58×10^{-15}
H3K4me2	"Associated with promoters and enhancers"	1.26×10^{-14}	3.98×10^{-14}
H3K4me3	"Primarily associated with promoters/transcription starts"	0.0158	5.01×10^{-14}
H3K9me3	"Repressive mark associated with constitutive heterochromatin and repetitive elements"	NS ^b	NS ^b
H3K27ac	"Mark of active regulatory elements; may distinguish active enhancers and promoters from their inactive counterparts"	1.58×10^{-7}	3.98×10^{-16}
H3K27me3	"Repressive mark established by polycomb complex activity associated with repressive domains and silent developmental genes"	NS ^b	NS ^b
H3K36me3	"Elongation mark associated with transcribed portions of genes, with preference for 3' regions after intron 1"	NS ^b	NS ^b
H3K79me2	"Transcription-associated mark, with preference for 5' end of genes"	NS ^b	2.51×10^{-13}
H4K20me1	"Preference for 5' end of genes"	NS ^b	NS ^b
P300	"Histone acetyltransferase linked to enhancer activity"	0.0251	3.16×10^{-9}

^aDescriptions of histone modifications are taken from ENCODE [37]^bNS=not significant