A twin-pair analysis indicates congenital scoliosis is associated with allele-specific methylation in the *SVIL* gene

ZHIFA ZHANG[1*], YONGJUN CHEN[2*], YUEZHOU WU[2*], YONGYU HAO[1], XUELIN ZHAO[1], XIANGYU WANG[1], YAN WANG[1], YANHAI XI[3] and XUESONG ZHANG[1]

[1]Department of Orthopedic Surgery, Surgery Building, The People's Liberation Army General Hospital, Beijing 100000; [2]Department of Spine Surgery, Zhongshan Hospital Xiamen University, School of Medicine, Xiamen University, Xiamen, Fujian 361004; [3]Department of Orthopedics, Spine Surgery, Changzheng Hospital, Second Military Medical University, Shanghai 200003, P.R. China

Received June 10, 2019; Accepted February 21, 2020

DOI: 10.3892/mmr.2020.11273

Abstract. Congenital scoliosis (CS) is a congenital disease resulting in abnormal vertebral development. Several studies have indicated that both genetic and environmental factors during pregnancy increase the risk of CS development. However, the exact mechanisms underlying CS pathogenesis remain unknown. To address this issue, both genetic (by whole-exome sequencing) and epigenetic (by methylated DNA immunoprecipitation sequencing) maps from CS disease-discordant monozygotic twins were generated in the present study. The differences in the presence of common and rare single nucleotide polymorphisms and in methylation patterns between the twins were investigated. The results indicated that rare mutations were more likely to underlie CS development compared with common mutations. Furthermore, differences in the allele-specific methylation pattern in the supervillin (*SVIL*) gene between the twins were identified. It has been reported that *SVIL* exerts a number of functions associated with CS, indicating its role as a novel mechanism promoting CS pathogenesis.

Introduction

Congenital scoliosis (CS) is a rare but debilitating disease with an estimated prevalence of 1 per 1,000 live births worldwide (1). CS is characterized by a wide variety of spine deformities resulting from the abnormal vertebral development during the first 4 weeks of gestation (2). Typical clinical symptoms of CS include malformations of the vertebra and ribs. Congenital vertebral malformations more often include hemivertebrae (half of a vertebrae), additional vertebrae, vertebral bar (a defect of vertebral separation during development), butterfly and wedge-shaped vertebrae (3-5). Previous studies have indicated that genetic and environmental factors are likely to serve a role in the occurrence of CS by interfering with the development of the medial portion of the mesoderm (1,6,7). Several studies supported the association between the induction of genetic mutations with vertebral anomalies in CS. For example, it has been reported that Delta like canonical notch ligand 3 (*DLL3*) mutations were involved in the developmental mechanisms of CS (8). Another study focusing on a form of spondylocostal dysostosis (SCD) suggested that mutations in the *DLL3* gene, located on chromosome 19, are associated with the disease phenotype (9). However, other studies have supported the association between environmental factors and CS development. For example, studies in mice indicated that CS was highly associated with maternal exposure to toxins, such as carbon monoxide (10,11). Other studies indicate that additional potential environmental factors are involved in CS development, including maternal diabetes and ingestion of antiepileptic drugs during pregnancy (12,13).

Given the rare nature of CS, as well as other rare diseases, this disease is difficult to study using conventional methods, such as genome-wide association studies (GWAS). The rationale for GWAS is that common diseases are associated with common variants (14), which is not relevant to rare diseases. In addition, the low occurance rate of rare variants (<1%) requires a large population sample in order for the variants to be distinguished, which is not feasible for rare diseases. Therefore, in the present study, an alternative approach was adopted that allowed the comprehensive evaluation of a congenital rare disease through the use of well-phenotyped, disease-discordant monozygotic twins. Thus, a combination of whole-exome sequencing (WES) for genetic analysis and methylated DNA immunoprecipitation sequencing (MeDIP-seq) for epigenetic analysis, was performed.

Correspondence to: Dr Xuesong Zhang, Department of Orthopedic Surgery, Surgery Building, The People's Liberation Army General Hospital, 28 Fuxing Road, Beijing 100000, P.R. China
E-mail: zhangxuesong301@outlook.com

Dr Yanhai Xi, Department of Orthopedics, Spine Surgery, Changzheng Hospital, Second Military Medical University, 415 Fengyang Road, Shanghai 200003, P.R. China
E-mail: xiyanhai1778@qq.com

*Contributed equally

Key words: congenital scoliosis, DNA methylation, supervillin gene, sequencing, single nucleotide polymorphism

The sequencing results revealed 367 twin-shared, 86 affected-individual specific and 60 healthy-individual specific non-synonymous rare variants located in coding sequence (CDS) regions. In addition, several major variances in the differentially methylated regions between the monozygotic twins were identified via MeDIP analysis. Several non-synonymous single nucleotide polymorphisms (SNPs) in the CDS region of CS-associated candidate genes were also identified; however, none of these genes exhibited differential methylation profiles. Allele specific methylation (ASM) analysis revealed that the supervillin (*SVIL*) gene, associated with cytoskeletal and skeletal developmental functions, exhibited opposite methylation pattern in the twins. Overall, the present study indicated that rare mutations and differential methylation profiles may be involved in CS pathogenesis; however, ASM may serve as a novel mechanism promoting CS development.

Materials and methods

Patients inclusion and exclusion criteria. A pair of twins (females, 15 years old) were enrolled into the present study from the Department of Orthopedic Surgery, The People's Liberation Army General Hospital in June 2017. The twin-patient (T-P) individual suffered from CS, and the twin-healthy (T-H) individual did not (Fig. S1). The inclusion criteria were: i) Spinal scoliosis diagnosed by radiological imaging and clinical symptoms; ii) T1-L5 vertebral formation defects including hemivertebra, butterfly vertebra and wedge vertebra; iii) T1-L5 vertebral defects of segmentation including vertebral bar and block vertebrae; and iv) with or without rib abnormalities. The exclusion criteria were: i) Presence of frequently encountered syndromes associated with congenital vertebral malformation (CVM), including Alagille syndrome, Goldenhar's syndrome, Jarcho-Levin syndrome, Klippel-Feil syndrome, Sotos syndrome and VACTERL association conditions; ii) vertebral malformations in association with other renal, cardiac or spinal cord malformations; and iii) cervical CVM. Informed consent was obtained from the patients and their parents. The study (approval no. 3012017114) was approved by the Ethics Committee of The People's Liberation Army General Hospital and written informed consent was obtained from all participants for their DNA samples to be used in the experimental procedures and their images to be published. Any personally identifiable information of the participants were removed. The study was conducted in accordance with the Declaration of Helsinki.

DNA sample preparation. DNA was extracted from peripheral blood of monozygotic twins who were discordant for CS. To confirm that both twins were monozygotic, microsatellite analysis (Excel Microsatellite Toolkit; Version 3.1.1., http://courses.washington.edu/ fish543/Software.htm) was performed as previously described (15). The common SNPs were filtered out from the total SNPs using the SNPinProbe database (dbSNP v129, https://skydrive.live.com/?uc=2&cid=fb2a64e541add2be) and 1,000 Genomes Project (16).

Exome capture followed by sequencing. The genomic DNA samples that met all the inclusion criteria (from the twins) were randomly fragmented and the fragments of 150-200 bp length were purified as described in a previous study (17). Adaptors were ligated to both ends of the fragments, followed by ligation-mediated polymerase chain reaction (PCR), purification, and hybridization using a SureSelect Biotinylated RNA Library with baits (Thermo Fisher Scientific, Inc.). The DNA fragments unlinked to streptavidin beads (Thermo Fisher Scientific, Inc.) were washed out following 24 h hybridization. Each captured library was sequenced using the Illumina Hiseq 2000 sequencing system (Illumina, Inc.).

Raw image files were processed by Illumina Basecaller Software 1.7 for base-calling with default parameters (Illumina, Inc.). The exome sequences of each individual were generated as 90-bp paired-end reads. The map of the reads onto the Genome Reference Consortium Human Build 37 (GRCh37; University of California Santa Cruz Genome Reference Consortium; https://genome.ucsc.edu/cite.html) was performed using SOAPaligner/SOAP2 (Release 2.21, http://soap.genomics.org.cn/down/soap2.21release.tar.gz), with 3 mismatches allowed per read. The minimum and maximum insert sizes allowed were 90 and 600 bp, respectively. Coverage and depth calculation were based on all mappable reads in the exome regions. The mean, 2.1 Gb of mappable sequences were generated per individual and ~60% of the reads were mapped to the exome. Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/), with $P<0.01$ was determined to confer statistical significance.

MeDIP-seq. Genomic DNA was fragmented to 100-500 bp by sonication (frequency, 18-20 KHz). Adaptors were ligated to both ends of the fragments. The end repair, base addition and adaptor ligation steps were performed using the Illumina Paired-End DNA Sample Prep kit (Illumina, Inc.) according to the manufacturer's protocol. The adaptor-ligated DNA was immunoprecipitated with anti-5mC antibody (Abcam; cat. no. ab73938; 5MC-CD, 1:200) at room temperature for 1 h, as described previously (18), captured by protein A/G agarose beads (Applied Biosystems; Thermo Fisher Scientific, Inc.), followed by the MeDIP products were validated by qPCR. qPCR was performed using an ABI7500 qPCR instrument (Applied Biosystems; Thermo Fisher Scientific, Inc.). In total, 5 μl SYBR Premix Ex Taq II (Takara Biotechnology, Co., Ltd.), 0.4 μl forward primer (10 μM), 0.4 μl reverse primer (10 μM), 0.2 μl ROX Reference Dye (Takara Biotechnology, Co., Ltd.), 1.0 μl cDNA template and 3.0 μl ddH_2O were mixed in the reaction solution. qPCR was performed as follows: Initial denaturation for 10 sec at 95˚C, followed by 45 cycles of 5 sec at 95˚C and 34 sec at 60˚C. Relative gene expression was calculated using the comparative quantification cycle (Cq) method. Fold changes were calculated using the $2^{-\Delta\Delta Cq}$ method (19). Primers were designed as follows: RAB3C forward, 5'-GATTCCACGCTTTCCCTCCAG-3' and reverse, 5'-AAATTCTCAAGTGCCCTCGCTACA-3'; and β-actin forward, 5'-GATCATTGCTCCTCCTGAGC-3' and reverse, 5'-ACTCCTGCTTGCTGATCCAC-3'. For the selection of the genomic DNA libraries that were suitable for the Illumina Genome Analyser II, DNA was amplified, purified and size-selected (200-300 bp including the adaptor sequence) using an agarose gel extraction kit (Tiangen Biotech Co., Ltd.), according to the manufacturer's instructions.

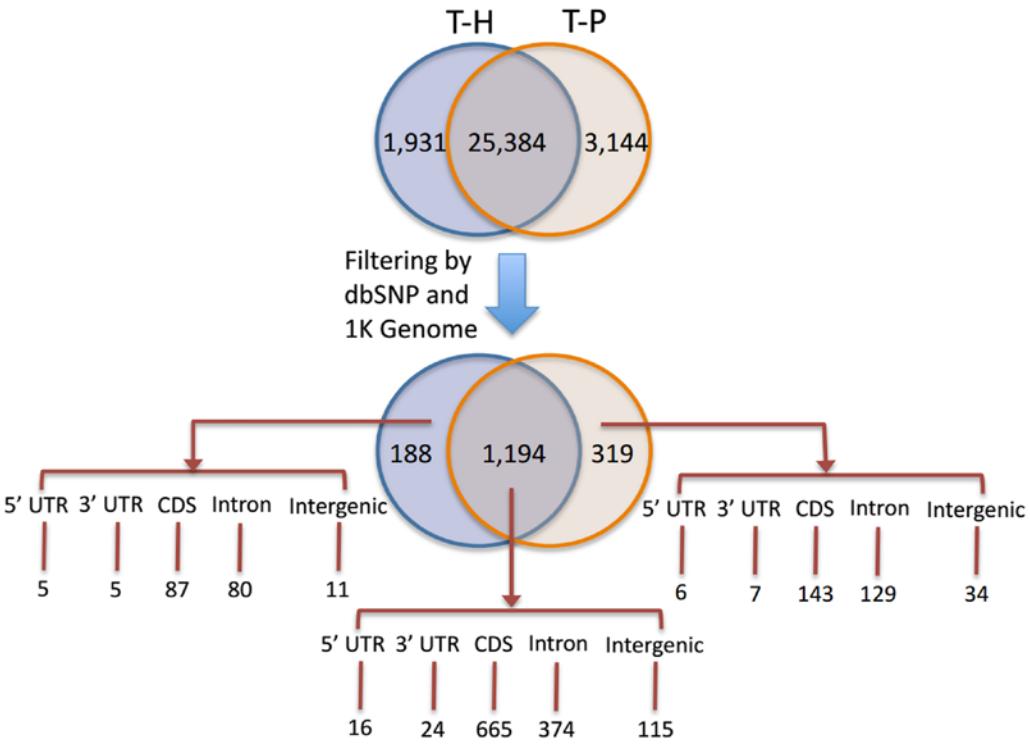

Figure 1. Identification method of rare mutations. T-H, twin-healthy; T-P, twin-patient; UTR, untranslated region; CDS, coding sequence.

The sequences of each individual were generated as 44-bp paired-end reads.

The sequencing results were mapped to the human reference genome GRCh37. The mean, 4.4 Gb of mappable sequences were generated per individual and ~89% of the reads could be uniquely mapped to the reference genome. The peak calling of the mappable reads was performed using the MACS software (version 1.4, http://miltenyi.ebiomall.com/18-1.html). The MACS parameters for peak calling of each individual were-g hs-nolambda-nomodel-shiftsize 150, whereas the parameters for differential peak calling were-gsize hs-nomodel-shiftsize 150. In addition, the parameters for T-H specific peak calling were-t T-H.bed-c T-P.bed, and for T-P specific peak calling-t T-P.bed-c T-H.bed. As the sequencing depth of the MeDIP-seq experiment was high (~95 million reads per individual), $P<10^{-20}$ was used as the cutoff for peak calling. The allelic bias in the MeDIP-seq read coverage for each individual was determined by χ^2 test ($P<0.01$). SPSS software (version 18.0; IBM Corp.) was used for statistical analysis.

Results

SNPs detection in twins. To determine the DNA sequence variation in the disease-discordant twins, WES was performed using the Illumina Hiseq 2000 sequencing system. A total of ~2.6 (with 78.4% mappable reads) and 2.5 Gb (83.6% mappable reads) sequences for T-H and T-P were generated, respectively, of which 60% were annotated as exons. The mean sequencing depth of the target region of the exome was 34.09- and 36.01-fold, respectively, which was considered sufficient for accurate SNP detection. Subsequently, 25,384 SNPs distinct from the reference human genome GRCh37, but shared by the twins, were identified using SOAPsnp (Release 1.03, http://soap.genomics.org.cn/soapsnp.html). Among them, 1,931 and 3,144 SNPs were T-H and T-P specific, respectively (Table SI). In addition, 1,194 (shared), 188 (T-H specific) and 319 (T-P specific) SNPs were located in CDS and annotated as non-synonymous mutations (Fig. 1).

Although common variants, defined as those present in >5% of the population according to the 1,000 Genomes Project (http://www.1000genomes.org/page.php) (16), may be involved in common diseases or traits, rare variants are considered more likely to be involved in rare diseases (20). Therefore, in the present study, minor allele frequency variants were identified (≤1%). In brief, the common SNPs were filtered out from the total SNPs using the SNPinProbe database (dbSNP v129, https://skydrive.live.com/?uc=2&cid=fb2a64e541add2be) and 1,000 Genomes Project (16). The analysis revealed 1,194 shared rare SNPs in the twins, 188 specific to the T-H and 319 specific to the T-P. Among these rare variants, 665 (shared), 87 (T-H specific) and 143 (T-P specific) were located in CDS regions, of which 367, 60 and 86, respectively, were non-synonymous (Fig. 1 and Table SI).

Subsequently, functional analysis of the genes with non-synonymous changes in the CDS was performed. Given that the basic traits of CS are developed during gestation and that cytoskeletal mechanisms are considered important for proper spinal development, the present study specifically focused on these categories. However, although >3,000 genes with non-synonymous CDS changes were identified using the total number of SNPs, none of these were demonstrated to be enriched during the embryonic development or cytoskeleton classifications. In contrast, when candidate genes (n=329) with rare non-synonymous CDS changes were considered, the cytoskeleton and development functional terms were significantly enriched ($P<0.00032$; Table I). Taken together, these results

Table I. Functional annotation for groups of SNPs.

A, All shared non-synonymous SNPs in CDS regions (3466 DAVID IDs)

Category	Term	P-value[a]
GOTERM_MF_FAT	Metallopeptidase activity	$2.3x10^{-7}$
GOTERM_MF_FAT	Serine-type endopeptidase inhibitor activity	$5.9x10^{-7}$
GOTERM_BP_FAT	Biological adhesion	$1.6x10^{-6}$

B, Rare variants of shared non-synonymous SNPs in CDS regions (329 DAVID IDs)

Category	Term	P-value[a]
GOTERM_BP_FAT	Cytoskeleton organization	$3.2x10^{-4}$
GOTERM_BP_FAT	Microtubule-based process	$2.0x10^{-3}$
GOTERM_BP_FAT	Blood vessel development	$4.6x10^{-3}$

C, All T-P–specific non-synonymous SNPs in CDS regions (361 DAVID IDs)

Category	Term	P-value[a]
GOTERM_BP_FAT	Ectoderm development	$1.1x10^{-4}$
GOTERM_BP_FAT	Epidermal cell differentiation	$4.1x10^{-4}$
GOTERM_BP_FAT	Epidermis development	$7.6x10^{-4}$

[a]For each mutation group, only the top 3 annotation terms with P<0.01 are listed.

supported the hypothesis that rare mutations were more likely to induce the development of rare diseases. Furthermore, the T-P-specific mutations, according to functional enrichment terms, appeared to be more likely to affect CS development than T-H-specific mutations.

Disease-discordant twins exhibit few differential DNA methylation regions. Several studies have proposed that the environment may contribute to CS pathogenesis by effecting DNA methylation (5,7,21,22). Therefore, the methylation profiles of the twins were investigated using MeDIP-seq. Following fragmentation, immunoprecipitation and high-throuhgput sequencing, ~95 million reads (2.6 Gb) per individual were generated, with 89% of the reads mapped to the genome. Overall, the DNA methylation profiles of T-H and T-P were identical at both the genomic level and in gene-associated regions (promoter and gene body), indicating that the twins did not exhibit any significant epigentic differences (Fig. S2). As the sequencing depth was sufficiently high, ~95 million reads per individual, stringent criteria were employed in order the MeDIP-seq peaks to be indetified ($P<10^{-20}$) by using MACS (23). The analysis identified 132,130 and 125,907 peaks for T-H and T-P, respectively. In addition, the same peak calling criteria were used in order the individual-specific peaks to be identified. The differential DNA methylation regions in the twins were 0 and 7 in T-H and T-P specific peaks, respectively. Among the 7 T-P specific peaks, 3 were aligned to RAB3C, member RAS oncogene family, ATPase phospholipid transporting 10B (putative) and leucine zipper protein 2 genes. Although the possibility that methylation changes in ≥1 of these genes may be associated with the pathogenicity of CS cannot be excluded, the small number of the differential regions indicated that this probability is low.

Genetic and epigenetic assessment of known CS-associated genes. Previous studies in mice and humans have identified 57 genes that may be involved in CS pathogenesis, the majority of which have been demonstrated to be crucial for normal somite formation (6,10,24-32). The known CS candidate genes are summarized in Table SII. Therefore, analyses on these genes were performed so that differences in their genetic or epigenetic features could be identified.

The SNPs identified in T-H or T-P human Refseq genes were aligned. Among the 57 identified SNPs in these genes, 35 were present in either the T-P or the T-H individuals (Table II). In addition, 29 shared, and 4 T-H- and 2 T-P-specific SNPs were detected that were annotated as non-synonymous. None of these SNPs were considered rare variants. Finally, the methylation analysis did not reveal any significant differential methylation peaks ($P<10^{-20}$) in the promoter or gene body of any of the 57 known CS-associated candidate genes.

ASM may contribute to CS pathogenesis. ASM may directly affect the pattern of gene expression. Several DNA regions acquire ASM during development and often in a tissue-specific manner, whereas others are methylated during gametogenesis, which is stably maintained throughout

Table II. Information on the 35 SNPs present in coding sequence regions.

A, Shared by identical twins

SNP Position	Sequence change	Amino acid change	Synonymous or non-synonymous	Relevant gene
chr12:54367061	TCT to TCG	Ser to Ser	Synonymous	HOXC11:NM_014212
chr17:46607021	GAA to GGA	Glu to Gly	Non-synonymous	HOXB1:NM_002144
chr17:46670520	CAA to CAG	Gln to Gln	Synonymous	HOXB5:NM_002147
chr17:46688135	GCT to GTT	Ala to Val	Non-synonymous	HOXB7:NM_004502
chr17:46688256	ACT to GCT	Thr to Ala	Non-synonymous	HOXB7:NM_004502
chr7:27135314	CGC to CAC	Arg to His	Non-synonymous	HOXA1:NM_153620
chr7:27169093	CGA to CGC	Arg to Arg	Synonymous	HOXA4:NM_002141
chr7:27196069	GCT to GCG	Ala to Ala	Synonymous	HOXA7:NM_006896
chr19:39994711	CTG to CCG	Leu to Pro	Non-synonymous	DLL3:NM_203486
chr9:139391636	GAC to GAT	Asp to Asp	Synonymous	NOTCH1:NM_017617
chr9:139397707	GAC to GAT	Asp to Asp	Synonymous	NOTCH1:NM_017617
chr9:139402760	TGC to TGT	Cys to Cys	Synonymous	NOTCH1:NM_017617
chr9:139407932	AAT to AAC	Asn to Asn	Synonymous	NOTCH1:NM_017617
chr9:139418260	AAT to AAC	Asn to Asn	Synonymous	NOTCH1:NM_017617
chr21:38117040	TCG to TCA	Ser to Ser	Synonymous	SIM2:NM_009586
chr3:123357037	AAT to AAC	Asn to Asn	Synonymous	MYLK:NM_053028
chr3:123368013	GAT to GAC	Asp to Asp	Synonymous	MYLK:NM_053028
chr3:123418913	AAC to AAT	Asn to Asn	Synonymous	MYLK:NM_053028
chr3:123419288	GAG to GAA	Glu to Glu	Synonymous	MYLK:NM_053028
chr3:123419573	GAC to GAA	Asp to Glu	Non-synonymous	MYLK:NM_053028
chr3:123419733	CTA to CCA	Leu to Pro	Non-synonymous	MYLK:NM_053028
chr3:123452838	ACC to ACT	Thr to Thr	Synonymous	MYLK:NM_053028
chr3:123453061	GTG to GCG	Val to Ala	Non-synonymous	MYLK:NM_053028
chr3:123457893	CCA to TCA	Pro to Ser	Non-synonymous	MYLK:NM_053028
chr1:165173216	GAT to GAC	Asp to Asp	Synonymous	LMX1A:NM_177398
chr1:165218792	CTG to TTG	Leu to Leu	Synonymous	LMX1A:NM_177398
chr5:127614472	AGT to AGC	Ser to Ser	Synonymous	FBN2:NM_001999
chr5:127622491	ATG to GTG	Met to Val	Non-synonymous	FBN2:NM_001999
chr5:127685135	GTT to ATT	Val to Ile	Non-synonymous	FBN2:NM_001999

B, T-H-specific

SNP Position	Sequence change	Amino acid change	Synonymous or non-synonymous	Relevant gene
chr12:54394497	GCC to GCT:	Ala to Ala	Synonymous	HOXC9:NM_006897
chr17:46629593	CCC to ACC	Pro to Thr	Non-synonymous	HOXB3:NM_002146
chr20:590542	GTG to CTG	Val to Leu	Non-synonymous	TCF15:NM_004609
chr20:590543	ACC to ACG	Thr to Thr	Synonymous	TCF15:NM_004609

C, T-P-specific

SNP Position	Sequence change	Amino acid change	Synonymous or non-synonymous	Relevant gene
chr17:46805443	TCT to TCC	Ser to Ser	Synonymous	HOXB13:NM_006361
chr7:27196113	GCT to ACT	Ala to Thr	Non-synonymous	HOXA7:NM_006896

SNPs, single nucleotide polymorphisms.

development (3,22,33-36). To determine the role of ASM in CS pathogenesis, the possible differences in ASM in the twins were investigated by assessing the methylation status of different alleles.

Table III. Mutations showing allele-specific methylation in the T-P and T-H individuals.

			T-H		T-P			
Chr	Position	Ref.	Allelic methylation	P-value	Allelic methylation	P-value	Relevant gene	Consistent
Chr1	248801897	G	G>A	0.0068	G>A	0.0001	*OR2T35*	Yes
Chr10	29839864	A	C>A	0.005	A>C	0.0077	*SVIL*	No
Chr10	95095667	A	G>A	0.0004	G>A	0.0036	*FER1L3*	Yes
Chr11	86267633	C	T>C	0.0052	T>C	0.0065	*ME3*	Yes
Chr11	128782002	T	C>T	0.0002	C>T	0.0019	*KCNJ5*	Yes
Chr16	71008417	G	C>G	0.0004	C>G	0	*HYDIN*	Yes
Chr20	60588049	G	C>G	0.0087	C>G	0.0009	*TAF4*	Yes
Chr3	38350519	G	A>G	<0.001	A>G	0.0014	*SLC22A14*	Yes
Chr6	90482397	T	C>T	0.0005	>T	0.0045	*MDN1*	Yes
Chr9	93640009	G	G>A	0.0003	G>A	0.0001	*SYK*	Yes
Chr9	132569553	C	C>T	0.0091	C>T	0.0019	*TOR1B*	Yes

Chr, chromosome; Ref., reference base.

The exon-capture MeDIP-seq reads were used to identify the heterozygous variants between the twins. A total of 14,535 heterozygous variants were identified and subsequently the allelic bias in the MeDIP-seq read coverage for each individual was determined by χ^2 test ($P<0.01$). In the ASM regions, 68 and 54 SNPs were detected in T-H and T-P, respectively. In addition, the twins shared 11 of these SNPs (Table III). Among them, 1 SNP, located on chr10:29839864, was detected in the *SVIL* gene. The *SVIL* gene was of interest, as DAVID analysis (37,38) indicated that it was related to Gene Ontology (GO) terms associated with skeletal tissue development, including cytoskeleton organization (GO:0007010), muscle organ development (GO:0007517) and skeletal muscle tissue development (GO:0007519) (35,37-39). In T-H, one allele containing a C was highly methylated compared with the other allele, whereas in T-P the allele containing an A showed comparatively higher methylation. The aforementioned observation indicates that the pathogenesis of CS is associated with more limited alterations in the methylation status of alleles and not to wider ones. These minor changes in the methylation status lead to differential allelic expression between the twins that affected their development.

Discussion

It has been reported that ~10% of CS cases in Japan are associated with mutations in the T-box transcription factor 6 (*TBX6*) gene. These patients exhibited compound heterozygosity for null mutations and the common hypomorphic risk haplotype defined by 3 SNPs in *TBX6* (2). Several studies have suggested that *TBX6*-associated CS and SCD may represent a spectrum of a disease caused by the compound heterozygosity model (2,26,27). A previous study reported that Dynein cytoplasmic 1 Heavy Chain 1, a mutant gene was identified in a patient with CS and spinal atrophy with lower extremity predominance using WES (40). In addition, several studies have demonstrated that Notch signaling pathway genes, including *DLL3*, mesoderm posterior BHLH transcription factor 2, LFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase (*LFNG*), Hes family BHLH transcription factor 7, Ripply transcription repressor 2 and TBX6, were associated with somitogenesis and their mutations were identified in SCD (9,10,41). A mutation in the *LFNG* gene was also reported in a case of SCD, a case of skeletal dysplasia with severe malformations of vertebra and rib (25). The patient with CS with *LFNG* mutations had multiple vertebral malformations, including hemivertebrae, butterfly and block vertebrae, and rib malformations (25). Therefore, *LFNG* mutations may cause a spectrum of phenotypes including CS and SCD. The current list of known disease-associated genes may only explain a small fraction of the genetic cause of CS (42).

The use of monozygotic twins with different phenotypes may serve as an ideal model for studying the differential genetic and epigenetic/environmental effects in CS and other diseases with complex inheritance pattern (21,41,43-46). In the present study, in order to investigate the potential factors involved in the development of a complex congenital bone disease on a genome scale, monozygotic twins with CS was used and a combination of exon capture sequencing and MeDIP-seq technologies was performed.

The results of the present study demonstrated that the identified rare SNPs, in contrast to the common ones, were enriched in functional categories that were associated with spinal development. These observations may indicate that rare mutations serve an important role in CS development. However, environmental factors which may be associated with the development of CS were not involved in this study. Due to lack of *in silico* analysis or functional validation, more samples and a more precise study design including healthly controls are required for further study.

The investigation of the differentially methylated regions did not reveal any differences in methylation among the known CS-associated genes. However, several non-synonymous SNPs were identified in the CDS regions of genes that were

associated with the pathogenesis of CS. Among these genes, SNPs in homeobox (*HOX)B1*, *HOXB7*, *HOXA1*, *DLL3*, myosin light chain kinase and fibrillin 2 were shared by the twins. These results indicated that both twins exhibited, genetically, a similarly high risk of abnormal vertebrae development and the resulting non-genetic differences between them further increased CS risk in one of the genetically identical pairs. In addition, *HOXA7*, a well-known key gene in vertebrae development, carried a T-P-specific SNP that could further promote CS in this individual.

In the future, PCR sequencing of these genes in a large sample of patients with CS and healthy individuals may be valuable to evaluate the potential effect of these genes on CS pathology.

In addition, a difference in the ASM patterns between the T-H and T-P in the *SVIL* gene was revealed. Notably, the *SVIL* gene serves a potential functional role in cytoskeletal and skeletal development. Therefore, further investigation of the ASM in *SVIL* and other genes, as well as an assessment of the differences in expression profiles, would be worthwhile. Additionally, it is worth investigating whether there could be any effects relevant to parent-of-origin allele-specific methylation, as CS has a higher likelihood of occurrence in females compared with males. Furthermore, imprinting mutations have been identified as causative in several sex-linked multiple-system diseases. Therefore, hormonal and other biochemical differences may affect sex-dependent CS development.

Acknowledgements

Not applicable.

Funding

The current study was supported by the National Natural Science Foundation of China (grant no. 81371911).

Availability of data and materials

All data generated or analyzed during this study are included in the present article.

Authors' contributions

XZhan, YC and ZZ conceived and designed the experiments. YH, XZhao and YWu participated in the design of the study. YWa, YX and XW performed the experiments and were involved in acquisition of data, analysis and interpretation of data. XZhan, YH and ZZ also analyzed the data. XZhan, YC and ZZ revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The present study was approved by the People's Liberation Army General Hospital. Written informed consent was obtained from all participants in order their DNA samples to be used in the experimental procedures and their images to be published. The study was conducted in accordance with the Declaration of Helsinki.

Patient consent for publication

Written informed consent was obtained from all participants in order their DNA samples to be used in the experimental procedures and their images to be published. Any personally identifiable information of the participants were not included in the images. The study was conducted in accordance with the Declaration of Helsinki.

Competing interests

The authors declare that they have no conflict of interest.

References

1. Weiss HR and Moramarco M: Congenital scoliosis (Mini-review). Curr Pediatr Rev 12: 43-47, 2016.
2. 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 372: 341-350, 2015.
3. de Baat P, van Biezen FC and de Baat C: Scoliosis: Review of types, aetiology, diagnostics, and treatment 2. Ned Tijdschr Tandheelkd 119: 531-553, 2012 (In Dutch).
4. McMaster MJ and Ohtsuka K: The natural history of congenital scoliosis. A study of two hundred and fifty-one patients. J Bone Joint Surg Am 64: 1128-1147, 1982.
5. Shands AR Jr and Bundens WD: Congenital deformities of the spine; an analysis of the roentgenograms of 700 children. Bull Hosp Joint Dis 17: 110-133, 1956.
6. Bouman A, Waisfisz Q, Admiraal J, van de Loo M, van Rijn RR, Micha D, Oostra RJ and Mathijssen IB: Homozygous DMRT2 variant associates with severe rib malformations in a newborn. Am J Med Genet A 176: 1216-1221, 2018.
7. Li Z, Yu X and Shen J: Environmental aspects of congenital scoliosis. Environ Sci Pollut Res Int 22: 5751-5755, 2015.
8. Chapman G, Sparrow DB, Kremmer E and Dunwoodie SL: Notch inhibition by the ligand DELTA-LIKE 3 defines the mechanism of abnormal vertebral segmentation in spondylocostal dysostosis. Hum Mol Genet 20: 905-916, 2011.
9. Giampietro PF, Raggio CL, Reynolds C, Ghebranious N, Burmester JK, Glurich I, Rasmussen K, McPherson E, Pauli RM, Shukla SK, *et al*: DLL3 as a candidate gene for vertebral malformations. Am J Med Genet A 140: 2447-2453, 2006.
10. Chapman DL: Impaired intermediate formation in mouse embryos expressing reduced levels of Tbx6. Genesis 57: e23270, 2019.
11. Deputy NP, Kim SY, Conrey EJ and Bullard KM: Prevalence and changes in preexisting diabetes and gestational diabetes among women who had a live birth-United States, 2012-2016. MMWR Morb Mortal Wkly Rep 67: 1201-1207, 2018.
12. Weston J, Bromley R, Jackson CF, Adab N, Clayton-Smith J, Greenhalgh J, Hounsome J, McKay AJ, Tudur Smith C and Marson AG: Monotherapy treatment of epilepsy in pregnancy: Congenital malformation outcomes in the child. Cochrane Database Syst Rev 11: CD010224, 2016.
13. Chanson P and Salenave S: Diabetes insipidus and pregnancy. Ann Endocrinol (Paris) 77: 135-138, 2016.
14. Génin E: Missing heritability of complex diseases: Case solved? Hum Genet 139: 103-113, 2020.
15. Kim KS and Sappington TW: Microsatellite data analysis for population genetics. Methods Mol Biol 1006: 271-295, 2013.
16. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA and Abecasis GR: A global reference for human genetic variation. Nature 526: 68-74, 2015.
17. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, *et al*: Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci USA 106: 19096-19101, 2009.
18. Xiong Y, Wang M, Zhao J, Wang L, Li X, Zhang Z, Jia L and Han Y: SIRT3 is correlated with the malignancy of non-small cell lung cancer. Int J Oncol 50: 903-910, 2017.
19. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25: 402-408, 2001.

20. Fairley S, Lowy-Gallego E, Perry E and Flicek P: The international genome sample resource (IGSR) collection of open human genomic variation resources. Nucleic Acids Res 48 (D1): D941-D947, 2020.
21. Castellani CA, Laufer BI, Melka MG, Diehl EJ, O'Reilly RL and Singh SM: DNA methylation differences in monozygotic twin pairs discordant for schizophrenia identifies psychosis related genes and networks. BMC Med Genomics 8: 17, 2015.
22. Giampietro PF, Raggio CL, Blank RD, McCarty C, Broeckel U and Pickart MA: Clinical, genetic and environmental factors associated with congenital vertebral malformations. Mol Syndromol 4: 94-105, 2013.
23. Liu T: Use model-based analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Methods Mol Biol 1150: 81-95, 2014.
24. Homans JF, Baldew VGM, Brink RC, Kruyt MC, Schlösser TPC, Houben ML, Deeney VFX, Crowley TB, Castelein RM and McDonald-McGinn DM: Scoliosis in association with the 22q11.2 deletion syndrome: An observational study. Arch Dis Child 104: 19-24, 2019.
25. Takeda K, Kou I, Mizumoto S, Yamada S, Kawakami N, Nakajima M, Otomo N, Ogura Y, Miyake N, Matsumoto N, *et al*: Screening of known disease genes in congenital scoliosis. Mol Genet Genomic Med 6: 966-974, 2018.
26. 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 38: 317-323, 2017.
27. Lefebvre M, Duffourd Y, Jouan T, Poe C, Jean-Marçais N, Verloes A, St-Onge J, Riviere JB, Petit F, Pierquin G, *et al*: Autosomal recessive variations of TBX6, from congenital scoliosis to spondylocostal dysostosis. Clin Genet 91: 908-912. 2017.
28. Yabe T and Takada S: Molecular mechanism for cyclic generation of somites: Lessons from mice and zebrafish. Dev Growth Diffe 58: 31-42, 2016.
29. Jing B, Yuan J, Yin Z, Lv C, Lu S, Xiong H, Tang H, Ye G and Shi F: Dynamic properties of the segmentation clock mediated by microRNA. Int J Clin Exp Pathol 8: 196-206, 2015.
30. Alsiddiky AM: An insight into early onset of scoliosis: New update information-a review. Eur Rev Med Pharmacol Sci 19: 2750-2765, 2015.
31. Abe K, Takamatsu N, Ishikawa K, Tsurumi T, Tanimoto S, Sakurai Y, Lisse TS, Imai K, Serikawa T and Mashimo T: Novel ENU-induced mutation in Tbx6 causes dominant spondylocostal dysostosis-like vertebral malformations in the rat. PLoS One 10: 0130231, 2015.
32. Fernández-Jaén A, Suela J, Fernández-Mayoralas DM, Fernández-Perrone AL, Wotton KR, Dietrich S, Castellanos Mdel C, Cigudosa JC, Calleja-Pérez B and López-Martín S: Microduplication 10q24.31 in a Spanish girl with scoliosis and myopathy: The critical role of LBX. Am J Med Genet A 164A: 2074-2078, 2014.
33. Ouellet J and Odent T: Animal models for scoliosis research: State of the art, current concepts and future perspective applications. Eur Spine J 22 (Suppl 2): S81-S95, 2013.
34. McMaster MJ and McMaster ME: Prognosis for congenital scoliosis due to a unilateral failure of vertebral segmentation. J Bone Joint Surg Am 95: 972-979, 2013.
35. Imamichi T, Yang J, Huang da W, Sherman B and Lempicki RA: Interleukin-27 induces interferon-inducible genes: Analysis of gene expression profiles using Affymetrix microarray and DAVID. Methods Mol Biol 820: 25-53, 2012.
36. Giampietro PF: Genetic aspects of congenital and idiopathic scoliosis. Scientifica (Cairo) 2012: 152365, 2012.
37. Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57, 2009.
38. Huang da W, Sherman BT and Lempicki RA: Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1-13, 2009.
39. Fisch KM: Biological interpretation of complex genomic data. Methods Mol Biol 1908: 61-71, 2019.
40. Punetha J, Monges S, Franchi ME, Hoffman EP, Cirak S and Tesi-Rocha C: Exome sequencing identifies DYNC1H1 variant associated with vertebral abnormality and spinal muscular atrophy with lower extremity predominance. Pediatr Neurol 52: 239-244, 2015.
41. McRae AF, Visscher PM, Montgomery GW and Martin NG: Large autosomal copy-number differences within unselected monozygotic twin pairs are rare. Twin Res Hum Genet 18: 13-18, 2015.
42. Otomo N, Mizumoto S, Lu HF, Takeda K, Campos-Xavier B, Mittaz-Crettol L, Guo L, Takikawa K, Nakamura M, Yamada S, *et al*: Identification of novel LFNG mutations in spondylocostal dysostosis. J Hum Genet 64: 261-264, 2019.
43. Stamouli S, Anderlid BM, Willfors C, Thiruvahindrapuram B, Wei J, Berggren S, Nordgren A, Scherer SW, Lichtenstein P, Tammimies K and Bölte S: Copy number variation analysis of 100 twin pairs enriched for neurodevelopmental disorders. Twin Res Hum Genet 21: 1-11, 2018.
44. Wu PW, Mason KE, Durbin-Johnson BP, Salemi M, Phinney BS, Rocke DM, Parker GJ and Rice RH: Proteomic analysis of hair shafts from monozygotic twins: Expression profiles and genetically variant peptides. Proteomics 17, 2017.
45. Tan Q, Li W and Vandin F: Disease-concordant twins empower genetic association studies. Ann Hum Genet 81: 20-26, 2017.
46. Watanabe M, Honda C; Osaka Twin Research Group, Iwatani Y, Yorifuji S, Iso H, Kamide K, Hatazawa J, Kihara S, Sakai N, *et al*: Within-pair differences of DNA methylation levels between monozygotic twins are different between male and female pairs. BMC Med Genomics 9: 55, 2016.