Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Wu N, Ming X, Xiao J, et al. *TBX6* null variants and a common hypomorphic allele in congenital scoliosis. N Engl J Med 2015;372:341-50. DOI: 10.1056/NEJMoa1406829

Supplementary Appendix

Title: TBX6 Null Variants and a Common Hypomorphic Allele in Congenital Scoliosis

Nan Wu, Xuan Ming, Jianqiu Xiao, Zhihong Wu, Xiaoli Chen, Marwan Shinawi, Yiping Shen, Guangju Yu, Jiaqi Liu, Hua Xie, Zoran S. Gucev, Sen Liu, Nan Yang, Hussam Al-Kateb, Jun Chen, Jian Zhang, Natalie Hauser, Ting Zhang, Velibor Tasic, Pengfei Liu, Xinlin Su, Xuedong Pan, Chunyu Liu, Liwen Wang, Joseph Shen, Jianxiong Shen, Yulin Chen, Ting Zhang, Jianguo Zhang, Kwong Wai Choy, Jun Wang, Qiqi Wang, Shugang Li, Weichen Zhou, Jin Guo, Yipeng Wang, Cheng Zhang, Hong Zhao, Yu An, Yu Zhao, Jiucun Wang, Zhenlei Liu, Yuzhi Zuo, Ye Tian, Xisheng Weng, V. Reid Sutton, Hongyan Wang, Yue Ming, Shashikant Kulkarni, Tao P. Zhong, Philip F. Giampietro, Sally L. Dunwoodie, Sau Wai Cheung, Xue Zhang, Li Jin, James R. Lupski, Guixing Qiu, and Feng Zhang

Table of Contents

METHODS

Study Oversight

This work was approved by the institutional review boards of the Peking Union Medical College Hospital, Fudan University, Capital Institute of Pediatrics, and the other participating institutions. We obtained written informed consent from the participants from the participants (those who were ≥18 years of age at the time of enrollment) or their guardians (for participants who were <18 years of age). F.Z. and N.W. designed the study. All the authors were involved in the data collection and vouch for the accuracy and completeness of the data and analyses. F.Z. wrote the first draft of the manuscript, which was reviewed, modified, and approved in its final version by all the contributing co-authors.

Subject Information

The participants reported their ancestries during face-to-face interview. Each participant is required to report his/her ancestry, and the interviewers write the ancestry information down on the questionnaire.

In Series 1 and 2, the patients with known syndromes (including Alagille syndrome, Goldenhar's syndrome, hemifacial microsomia, Klippel-Feil syndrome, spondylocostal dysostosis, spondylothoracic dysostosis, and VACTERL syndrome) were excluded clinically.

For a further replication study, 42 unrelated subjects with 16p11.2 deletion (Series 3) were obtained from multiple centers in both the USA and China. These subjects were initially referred for clinical chromosomal microarray testing due to various medical problems (neurodevelopmental disorders, multiple congenital anomalies, dysmorphic features, and etc.).[1-3]

High-density Oligonucleotide Comparative Genomic Hybridization Microarrays

The genome-wide analyses of copy-number variant (CNV) were conducted using Agilent oligonucleotide-based comparative genomic hybridization (CGH) microarrays. Some experimental details were previously described.[1,2]

The genomic DNA extracted from peripheral blood leukocytes of each subject and the

sex-matched reference DNA (Promega) were respectively fragmented using *Alu*I and *Rsa*I enzyme digestion. DNA labeling was conducted using Agilent SureTag DNA Labeling Kit. Different fluorescence dyes were used for DNA labeling of each subject (Cy5-dUTP) and the reference DNA (Cy3-dUTP). Each labeled subject DNA was hybridized together with the labeled reference DNA onto Agilent human CGH microarrays (including the formats of 1×1M, 1×244k, and 4×180k).

The Agilent 1×1M and 1×244k CGH microarrays are commercially available, which have approximately one million and 244,000 oligonucleotide probes, respectively. These two formats were used for genome-wide CNV analysis in this study. In addition, a custom 4×180k CGH microarray with approximately 180,000 oligonucleotide probes was used for verification of the 16p11.2 deletions suggested by quantitative PCR. This custom microarray has 509 probes in the 16p11.2 deletion region and achieves a much higher probe density in the target region than the Agilent 1×1M and 1×244k CGH microarrays. In addition to 16p11.2, our custom CGH microarray also covers 360 additional neurological genes/loci that are unrelated to this study.

DNA processing, microarray handling, and data analyses were conducted by following the Agilent oligonucleotide CGH protocol (version 6.0). The genome-wide CNV analysis were also conducted in control subjects of Han Chinese using Agilent 1×1M and/or NimbleGen 4.2M (Roche; with approximately 4.2 million probes covering the whole human genome) CGH microarrays.

Quantitative Polymerase Chain Reaction (qPCR) Analysis

Quantitative PCR analysis was conducted to screen for the *TBX6*/16p11.2 deletions. Two test loci (named PA and PB) in the 16p11.2 deletion region and one reference locus (named P1) outside of the deletion region were used. The qPCR primers were provided in the following table. The experiments were conducted using SYBR Green Realtime PCR Master Mix (TOYOBO) and ABI Prism^{TM} 7900HT Sequence Detection System. Three replicates were conducted for each assay. The average Ct value, ΔCt (PA-P1) and ΔCt (PB-P1) were calculated for each sample. In addition, one 16p11.2 deletion and one non-deletion sample previously confirmed by the human

genome CGH microarrays were selected as the positive and negative controls respectively.

- The primers for the quantitative PCR assay.

Primer	Sequence
P1-F	5’-GGGGAAGGAACTTACATGAC-3’
P1-R	5’-TCGTGTTTCCCTGTTGTACC-3’
PA-F	5’-GGTCTAAGCCACACACTAAC-3’
PA-R	5’-TGAGTTTAGGGACCAATCTA-3’
PB-F	5’-GCTGCCAGTATGTGACCGAGA-3’
PB-R	5’-GGGTGGAGGAGAGGATAGGG-3’

DNA Sequencing of the *TBX6* Gene

All of the exons of *TBX6* and the approximately 1-kb upstream region of *TBX6* were investigated in both CS patients and healthy controls.

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

Forward primer: 5'-TAGGGAGAGGGCTCTGTTCTCATGG-3'

Reverse primer: 5'-GCGTCCCAGGGAGGCAACCG-3'

PCR reaction conditions:

98℃	1 min	
98℃	10 sec	35 cycles
60℃	20 sec	
68℃	4 min	
68℃	10 min	
12℃	∞	

PCR system (50ul):

Water	29.5 ul
10×LA PCR Buffer II (Mg^{2+} Plus)	5 ul
dNTP Mixture (each 2.5 mM)	8 ul
Forward primer (10 uM)	2 ul
Reverse primer (10 uM)	2 ul
DMSO	1 ul
Template (50 ng/ul)	2 ul
TaKaRa LA Taq (5U/ul)	0.5 ul

(B) Sequencing primers.

Primer for *TBX6* fragment	Sequence
Upstream region	5'-CTCGAAGGGGTCCGAGAGG-3'
Upstream region and Exon 1	5'-CTCCTTCCATAGCTCCCGGT-3'
Exon 2	5'-GTTGCATACTGATCCCGAAT-3'
Exon 3a	5'-CTGCCCGAACTAGGTGTATG-3'
Exons 3b, 4 and 5	5'-AATGGCTTCCTAACAGATGAC-3'
Exons 6, 7 and 8a	5'-GAGCGGGAGGTTTGTGATG-3'
Exon 8b and 3'-UTR	5'-GGCAGCTGGAAACACAGGT-3'

Haplotyping of Common *TBX6* Variants

Haplotyping of common *TBX6* variants was conducted using ClonExpress One Step Cloning Kit (Vazyme). We amplified both the vectors and the insert DNA fragment to perform recombination cloning. The pGEM-T vector (Promega) was used as template for amplifying the vectors. Both vector and insert DNA fragments were amplified using Q5 polymerase (NEB). Their PCR products were dealt with gel exaction kit (GENERAY). After ligation, the products were transfected to *E. coli* competent cells. The bacterial colonies were collected and verified by Sanger DNA sequencing.

- The primers for the haplotyping analysis of common *TBX6* variants.

Primer	Sequence
T7-reverse	5'-TCGCCCTATAGTGAGTCGTATTACA-3'
SP6-reverse	5'-GTATTCTATAGTGTCACCTAAATAG-3'
CS-F	5'-GACTCACTATAGGGCGAGGGGAAGGGAGCGGGAGGTTTGTG-3'
CS-R	5'-GGTGACACTATAGAATACGCGCTGAGCCTGCCGGGAAGTGTAGT-3'

Cell Culture and Induced Differentiation

The cells of HEK293, HepG2 and HeLa were cultured in DMEM medium (Gibco) with 10% fetal bovine serum (FBS, Gibco). After digestion by trypsin, 1×10^5 cells were seeded to a

24-well plate with 500 ul culture medium.

The P19CL6 cells were cultured and induced to differentiate into cardiomyocytes as described previously.[4] The P19CL6 cells were seeded in a 96-well plate at 5×10^3 cells in a-MEM (Gibco) containing 10% FBS (Gibco), and induced to cardiac cells with addition of 1% DMSO (Sigma).[5]

We used a more efficient method to generate integration-free human induced pluripotent stem cells (iPSCs).[6] Okita et al. found that, using p53 suppression and nontransforming L-Myc, enhance the generation of hiPSCs with episomal plasmid vectors.[6] The episomal vectors express OCT4, SOX2, KLF4, LIN28, L-MYC and shRNA against P53 were purchased from Addgene. Human dermal fibroblast (HDF) cells were cultured in DMEM supplement with 10% FBS. The episomal plasmid mixtures were electroporated into 6×10^5 HDF cells. The cells were trypsinized 7d after transduction, and 1×10^5 cells were re-plated onto 10-cm dishes covered with MEF feeder layer. The HDF medium was replaced with hES medium containing bFGF the next day. By days 26–32d after plating, hES cell-like colonies start to appear.

Cardiac differentiation of human iPS cells (hiPSCs) was previously described.[7,8] Briefly, hiPSCs were seeded in Matrigel-coated six-well plates at 5×10^5 cells in E8 medium (Gibco). Next day, the medium was aspirated and replaced with 2 ml of fresh E8 medium. The medium change was repeated until cells reach 80–90% confluence, corresponding to day 0 of differentiation. E8 medium was then removed and replaced with 2 ml of RPMI/B-27 without insulin (Gibco) but containing CHIR99021 (Selleckchem). After 24 hours, the medium was again removed and replaced with 2 ml of RPMI/B-27 without insulin. The cells were harvested at 24h and 48h after differentiation.

RNA Extraction and RT-PCR

After being transferred from the cell incubator and discarding the culture medium, the cells were washed once by PBS. RNA extraction was performed by using RNAiso Plus (TaKaRa) according to the manufacturer. The RNA was reverse-transcribed into cDNA by using PrimeScript™ RT-PCR Kit (TaKaRa). The cDNAs were used for RT-PCR by using the

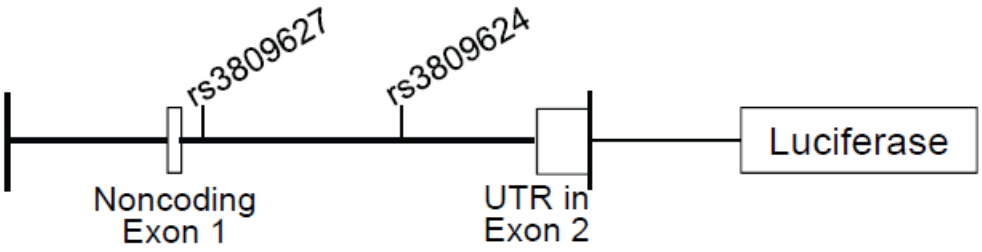
corresponding primers in the following table.

- The primers for the RT-PCR analyses of *TBX6*.

Primer	Sequence
Mouse actin beta-L	5’-ATATCGCTGCGCTGGTCGTC-3’
Mouse actin beta-R	5’-AGGATGGCGTGAGGGAGAGC-3’
Mouse-Tbx6-L	5’-CAGTGCTGAGGCCTACCTTC-3’
Mouse-Tbx6-R	5’-CCAGGCTGTAGGTCCAGAAA-3’
Human Brachyury-L	5’-AATTGGTCCAGCCTTGGAAT-3’
Human Brachyury-R	5’-GATGGGTGAGGGGTGTGTAG-3’
Human MYH6-L	5’-ATGAGCTGGATGAGGCAGAG-3’
Human MYH6-R	5’-GGTTGGCAAGAGTGAGGTTC-3’
Human TBX6-L	5’-TGTACCATCCACGAGAATTGTACCC-3’
Human TBX6-R	5’-GGTTCTGGTAGGCTGTCACGGAGAT-3’
Human GAPDH-L	5’-GCGCTGAGTACGTCGTGGAG-3’
Human GAPDH-R	5’-CTGTAGCCAAATTCGTTGTCATACC-3’

Luciferase Assay

The primers for the reporter plasmids in the luciferase assay are shown in the following table. Luciferase-primer-F and Luciferase-primer-R were used to amplify the wildtype fragment of the 5’ noncoding region in *TBX6*, while the other four primers were used for site-directed mutagenesis through the instruction of KOD-Plus-Mutagenesis Kit (TOYOBO). To construct reporter plasmids, we amplified the 1120-bp DNA fragments that involve the potential upstream regulatory elements of *TBX6* (shown in the following figure). The fragments containing four different variant combinations of *TBX6* (i.e. each differing from the rest by virtue of SNPs at rs3809624 and rs3809627) were constructed to PGL3-Basic vector (Promega).



- The schematics of the luciferase reporter construct with two potential regulative variants of *TBX6* (rs3809624 and rs3809627).

As for the three cell lines (HEK293, HepG2, and HeLa) used for preliminary functional assays, these cells were cultured respectively for 24 hours before being co-transfected with 500 ng *TBX6* reporter plasmids and 10 ng pRL-TK plasmid as a normalizing control using Lipofectamine 2000 (Invitrogen). After 24 hours of culture, the cells were lysed, and 20 μl of supernatant was used to assay the luciferase activity using the Dual-Luciferase Reporter Assay System (Promega). The relative reporter activity was normalized by the firefly activity to Renilla activity. Each assay was performed with at least three replicates.

For the further functional studies in the induced P19CL6 cells showing significant *TBX6* expression, 3×10^3 cells were seeded in 96-well culture plate in α-MEM (Gibco) that was supplemented with 10% FBS (Gibco) and 1% DMSO (Sigma). After 24 hours, the cells were transfected using Lipofectamine 3000 (Invitrogen) with 100 ng plasmids of each mutant promoter fused with Firefly luciferase reporter plasmid and 10 ng of pRL-TK plasmid as normalizing controls. After 48 hours, Firefly and Renilla luciferase activities were measured by Dual-Glo Luciferase Assay System (Promega). The relative reporter activity was normalized to the Renilla activity.

For the further functional studies during mesodermal and cardiac differentiation in hiPSCs, the hiPSCs were seeded in Matrigel-coated 96-well plates at 5×10^3 cells in E8 medium. After 48 hours of incubation, the cells were transfected with 150 ng of each mutant promoter reporter plasmid and 15 ng of pRL-TK plasmid as normalizing controls. After 24 hours, medium was aspirated and replaced with 100 ul RPMI/B-27 without insulin containing CHIR99021 in each

well. Firefly and Renilla luciferase activities were measured by Dual-Glo Luciferase Assay System after 24 hour incubation. The relative reporter activity was normalized to the Renilla activity.

- The primers for the reporter plasmids in the luciferase assay.

Primer	**Sequence**
Luciferase-primer-F	5’-CCGCTCGAGGCTCAGGCTCGCGTGTAATC-3’
Luciferase-primer-R	5’-CCCAAGCTTGTTGTAGTTCCGTCTGGCCTCA-3’
Mutation 624-F	5’-GTCAAGCCGAGGAGTTAATTATGTA-3’
Mutation 624-R	5’-TGATCCCCAGGGCTCGGACA-3’
Mutation 627-F	5’-TGATGGAGGTTTTTTTAGCATGGGGA-3’
Mutation 627-R	5’-CGAGCCAAGCCTCCAAAGTTCGAA-3’

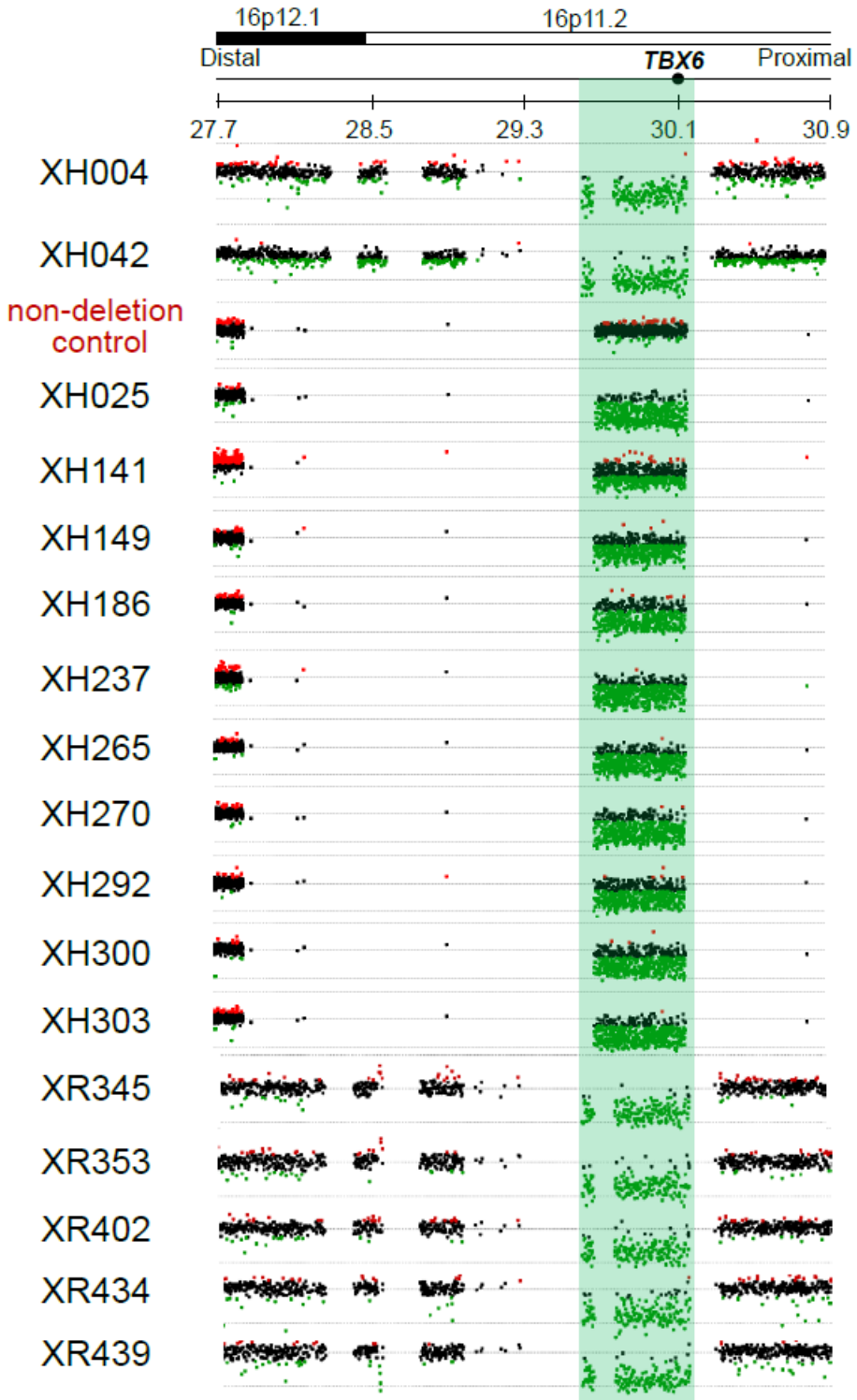
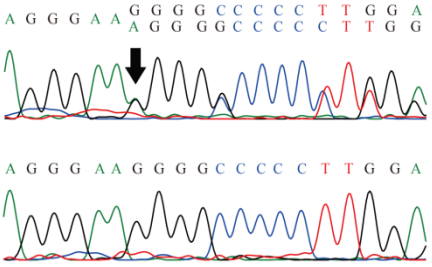
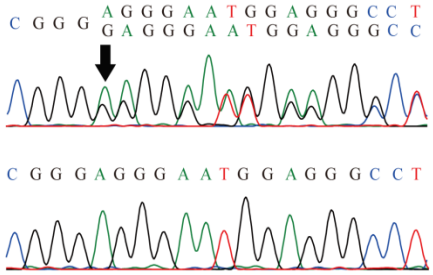


Figure S1. High-density CGH microarrays identified twelve 16p11.2 deletions in Series 1 (XH) and five deletions in Series 2 (XR). These deletions are recurrent. Seven of them were identified by genome-wide CGH microarrays. The remaining deletions initially suggested by qPCR were confirmed by our custom CGH microarrays with 509 oligonucleotide probes located in the deletion region (green shadow).

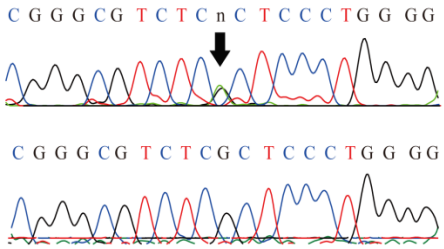
A) XH101 (*TBX6*, c.1250_1251insT)



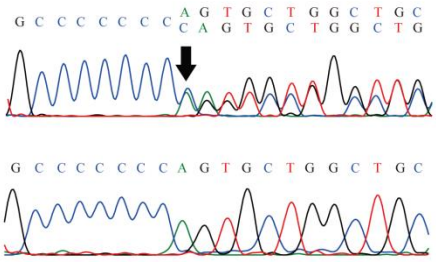
B) XH122 (*TBX6*, c.266_267insC)



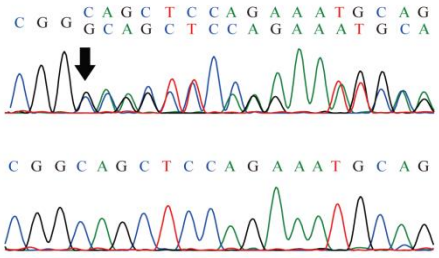
C) XH148 (*TBX6*, c.844C>T, p.R282X)



D) XH170 (*TBX6*, c.704_705insG)



E) XH286 (*TBX6*, c.1169_1170insC)



F) XR341 (*TBX6*, c.1179_1180delAG)

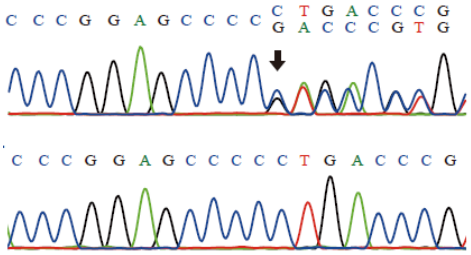


Figure S2. Sanger sequencing identified five *TBX6* frameshift mutations and one nonsense mutation in Series 1 and 2. The black arrows indicate the insertion/mutation positions. The DNA sequencing plots show the + strand of the human genome, whereas the *TBX6* gene is located in the - strand of the human genome. The homozygous reference alleles were shown in the bottom of each column. (A) A heterozygous insertion of c.1250_1251insT in XH101. (B) A heterozygous insertion of c.266_267insC in XH122. (C) A heterozygous nonsense mutation of c.844C>T (p.R282X) in XH148. (D) A heterozygous insertion of c.704_705insG in XH170. (E) A heterozygous insertion of c.1169_1170insC in XH286. (F) A heterozygous small deletion of c.1179_1180delAG in XR341. GenBank accession numbers for these *TBX6* null mutations: KP216200-KP216205.

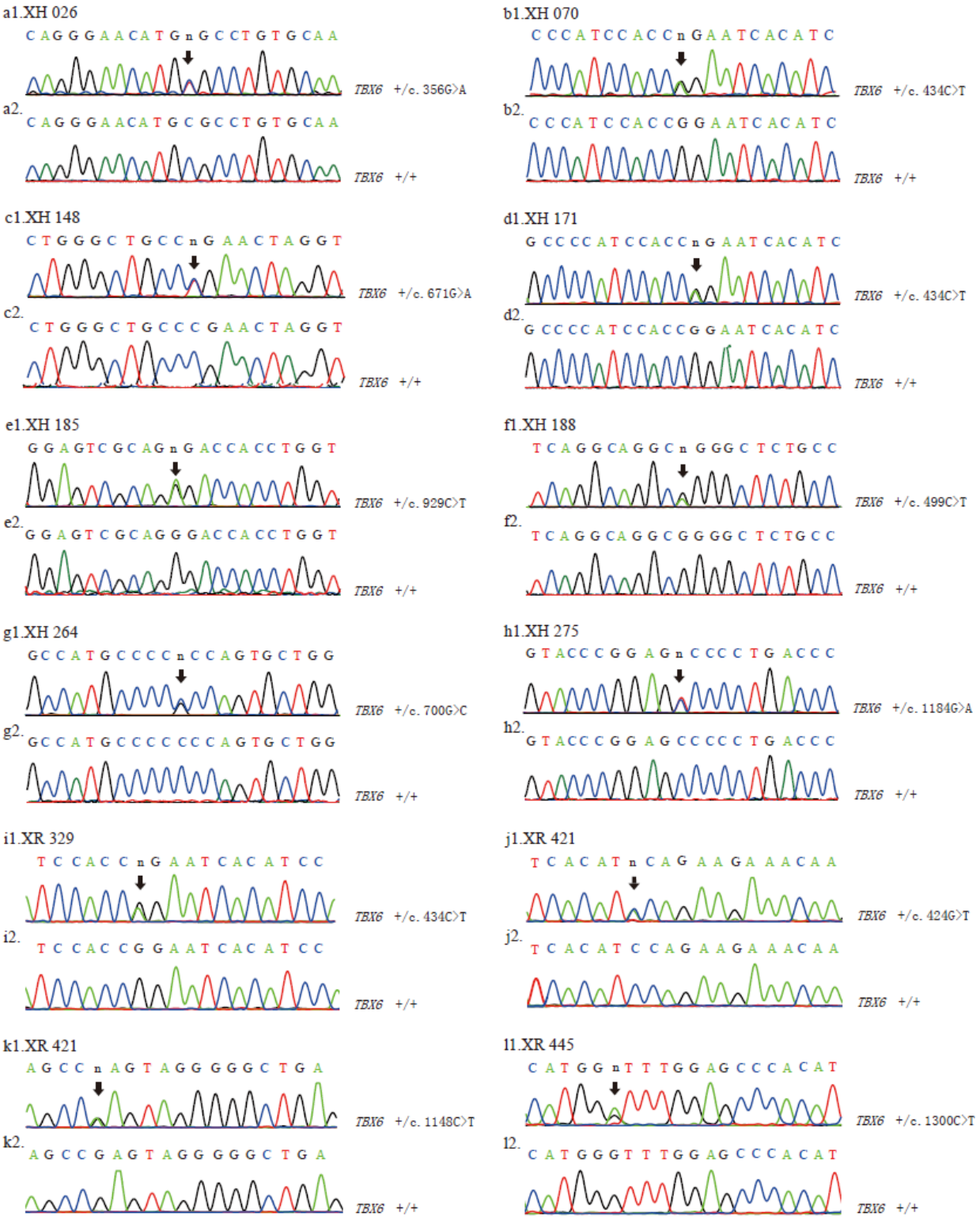
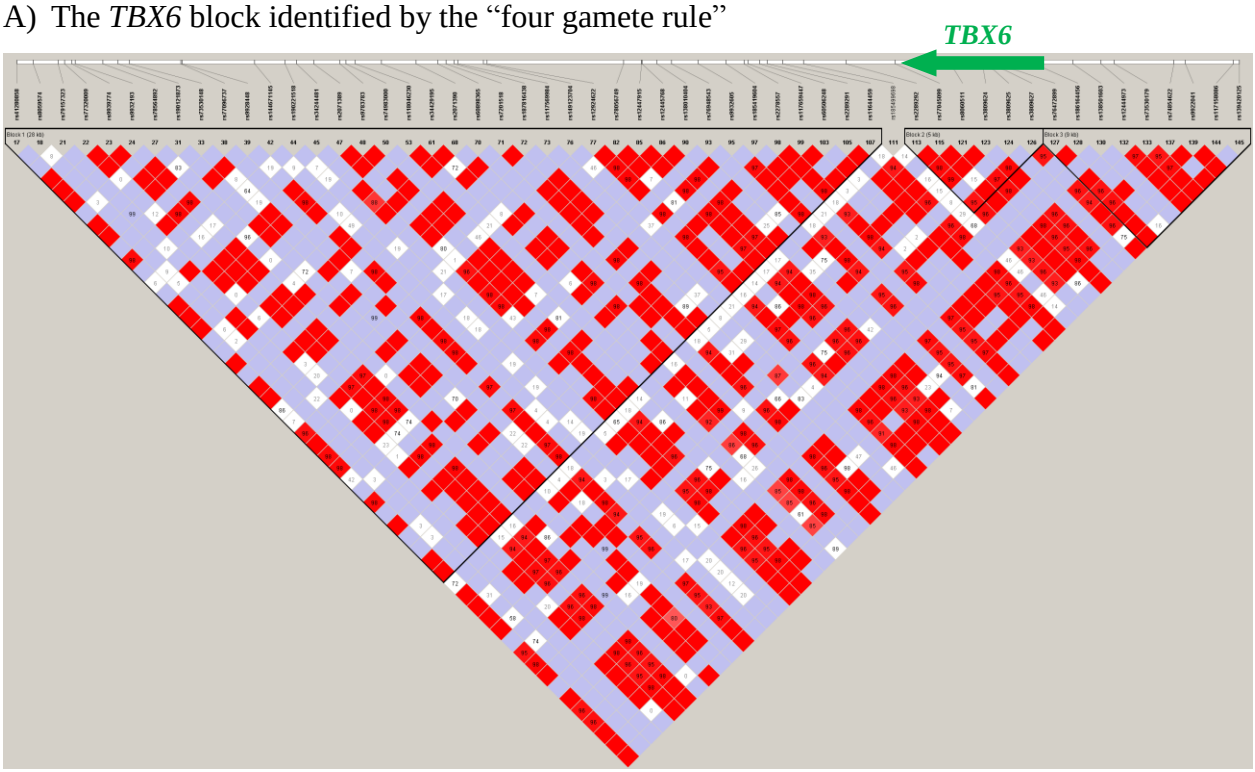


Figure S3. *TBX6* missense mutations in Series 1 & 2. Black arrows indicate nucleotide substitution positions. Sequencing plots show the + strand of the human genome, whereas *TBX6* is located in the - strand. The homozygous reference alleles were shown in the bottom of each column. The missense mutation in XH148 was located together with a *TBX6* null mutation; therefore, it has not been further analyzed in this study.

A) The *TBX6* block identified by the "four gamete rule"



B) The *TBX6* block identified by the "solid spine of LD"

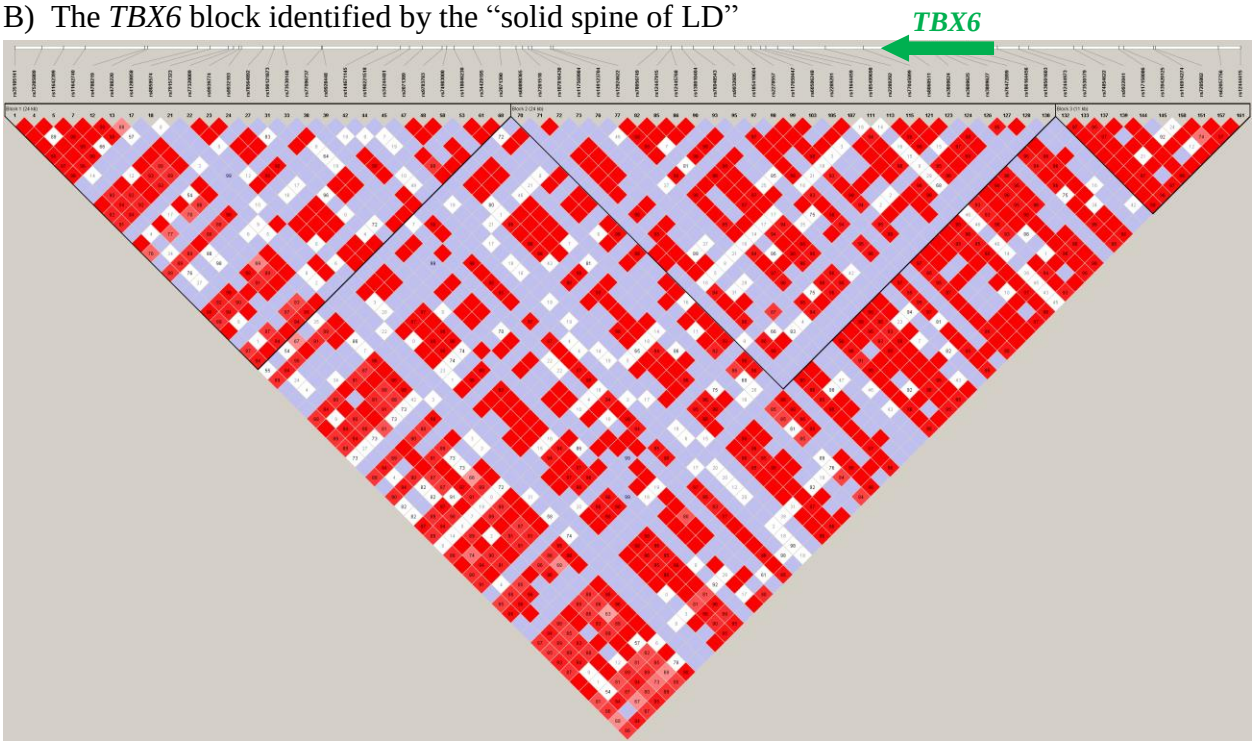


Figure S4. The Haploview tool shows the LD blocks of *TBX6* in the CHB+CHS populations.[9,10] (A) *TBX6* is located in a 5-kb block identified by the "four gamete rule". (B) *TBX6* is located in a 24-kb block identified by the "solid spine of LD".

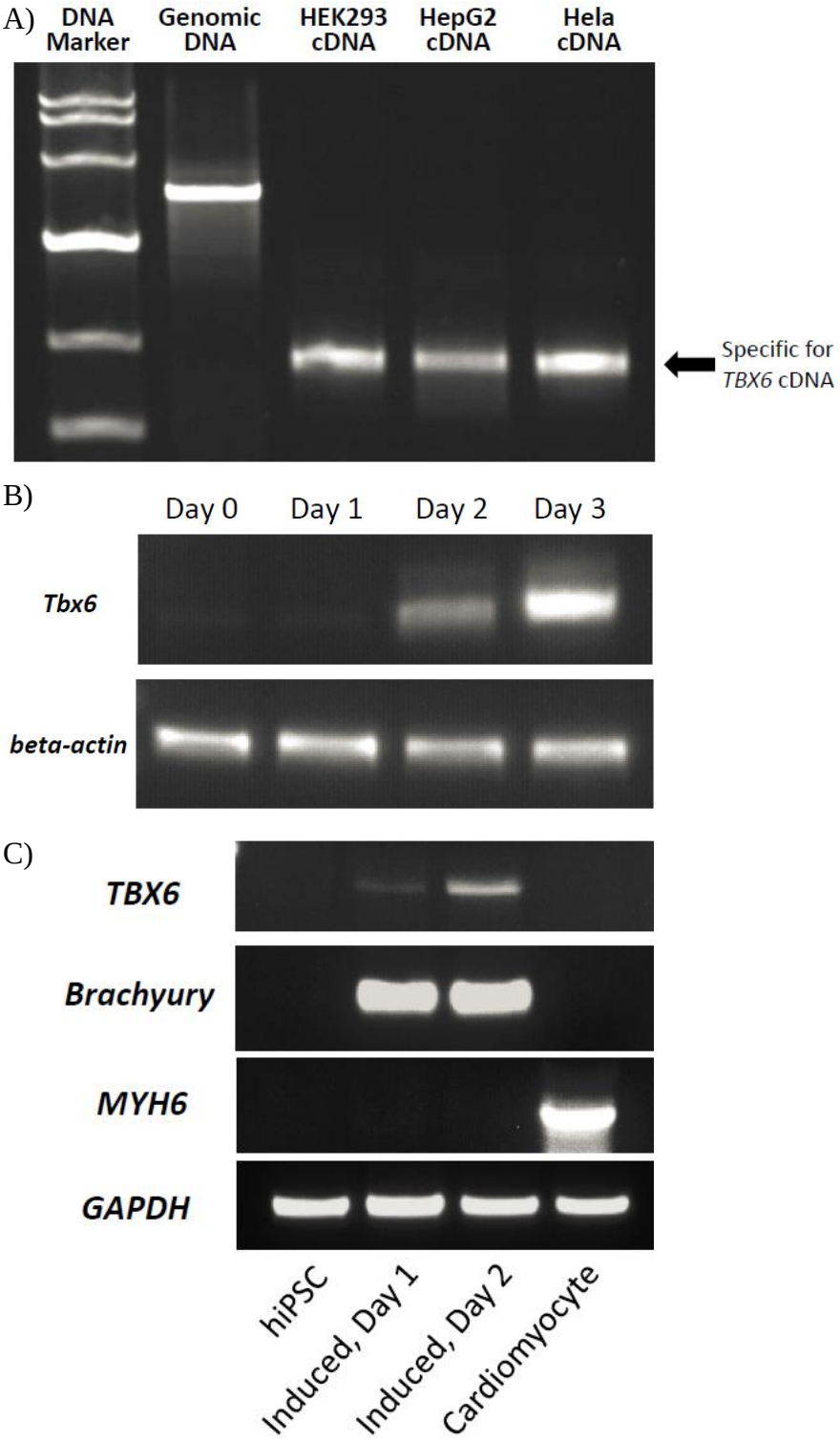


Figure S5. RT-PCR showed *TBX6* expression in the cell models used in this study. (A) *TBX6* is expressed in the cell lines of HEK293, HepG2 and HeLa. (B) The induced *Tbx6* expression by DMSO in the P19CL6 cells. (C) *TBX6* expression is observed when hiPSCs differentiate into mesodermal cells but its expression is absent in cardiomyocytes. *Brachyury* is a mesodermal cell marker, whereas *MYH6* is a cardiomyocyte marker.

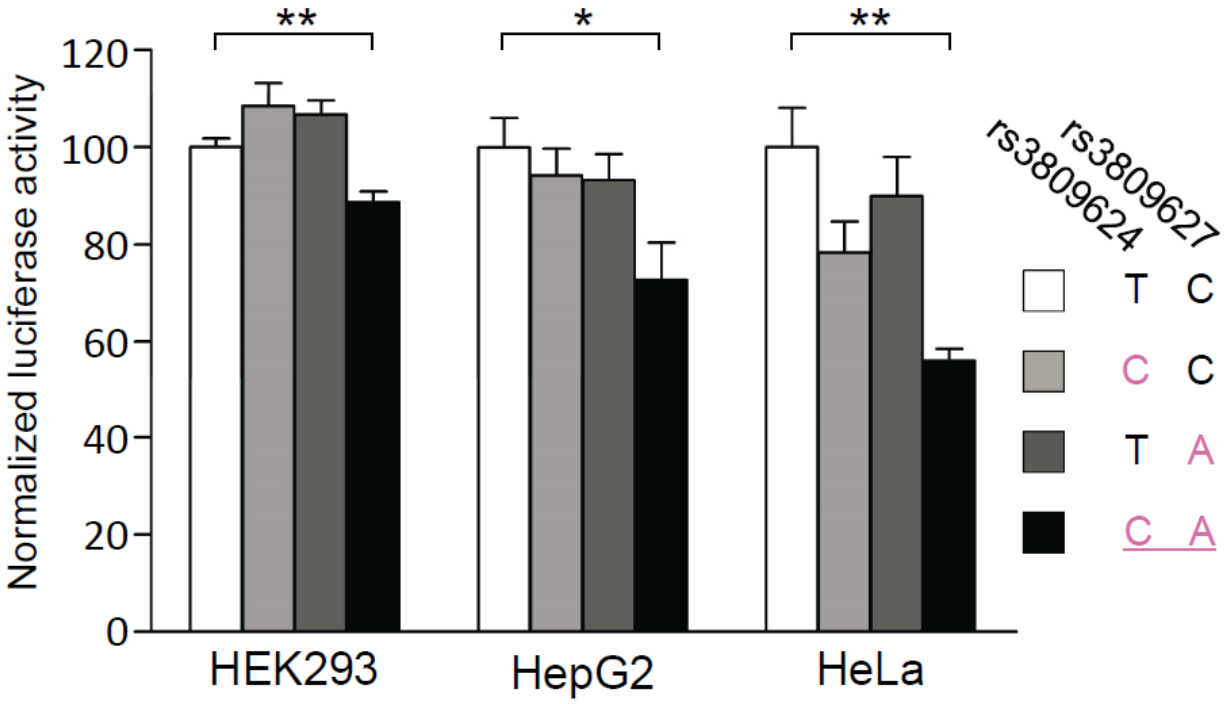


Figure S6. *In vitro* functional assay of two upstream non-coding variants of *TBX6* in three different cell lines (HEK293, HepG2 and HeLa). The results of luciferase reporter assay are shown. The non-reference alleles of rs3809624 and rs3809627 are shown in pink, and the risk haplotype is underlined. The unpaired T test is used for statistical analysis. Each value represents mean ± SEM of at least three independent experiments. *P<0.05, **P<0.01.

Table S1. Genetic and clinical information of the *TBX6* missense mutations identified in Series 1 and 2.

Patient No.	Sex	Age (*yr*)	*TBX6* variant		Vertebrate malformation	Other congenital malformation
			Allele 1	Allele 2		
			Missense mutation*	Common haplotype†		
1) Series 1						
XH026	M	5	**c.356G>A (p.R119H)**	T-C-A	T12 hemivertebra	Missing right 12th rib
XH070	M	7	**c.434C>T (p.P145L)**	T-C-A	Extensive defect of formation and segmentation from C6-T4 (including two hemivertebrae T8 & T11)	Atrial septal bulge; bilateral 13 ribs, abnormal morphology of left 3th and 8th ribs); congenital hernia
XH171	M	7	**c.434C>T (p.P145L)‡**	T-C-A	Extensive defect of formation and segmentation from T6-T12 (T7 & T9 & T10 hemivertebrae)	Missing right 7th, 9th and 10th ribs; fused right 8th, 11th and 12th ribs
XH185	F	15	c.929C>T (p.P310L)	T-C-A	T7 & T8 wedged vertebrae; fused L1 and L2	Mitral valve prolapse; tricuspid valve prolapse
XH188	F	13	c.499C>T (p.R167C)	C-T-C	T9 wedged vertebra	Fused left 9th -10th ribs
XH264	F	2	c.700G>C (p.G234R)	C-T-C	L2 hemivertebra	None
XH275	F	9	c.1184G>A (p.G395D)	T-C-A	T4 & T6 & T8 hemivertebrae; T7-9 segmentation defect	Missing left 1-6th ribs and fused 7-8th ribs; dextrocardia
2) Series 2						
XR329	M	5	**c.434C>T (p.P145L)**	T-C-A	T2 & T5 hemivertebrae T3 & T4 & T11 butterfly vertebrae	None
XR421	M	3	**c.424G>T (p.D142Y)** c.1148C>T (p.S383L)	T-C-A C-T-C	T10 hemivertebra	Missing right 10th rib, abnormal morphology of left 11th rib
XR445	F	11	c.1300C>T (p.P434S)	C-T-C	T12 & L1 unsegmented hemivertebra	Missing right 12th rib; syringomyelia

* All these *TBX6* missense mutations are rare and absent in the human populations studied by the 1000 Genomes Project. The mutations that are suggested to be functional by all or at least three bioinformatic tools (see Table S3) are also shown in bold.

† The haplotype defined by three common *TBX6* SNPs (reference/non-reference): rs2289292 (C/T) - rs3809624 (T/C) - rs3809627 (C/A).

‡ This mutation is *de novo*.

Table S2. Three missense mutations of *TBX6* identified in the control subjects.

Control No.	*TBX6* missense mutation	Allele frequency in populations*
H039	c.1061C>T (p.A354V)	None
H040	c.671G>A (p.R224Q)	None
H070	c.670C>G (p.R224G)	None

* Based on the data of the 1000 Genomes Project.

Table S3. Functional prediction of the *TBX6* missense mutations using bioinformatic tools.*

Subject	*TBX6* missense mutation	Bioinformatic tools[11-14]			
		PolyPhen-2	SIFT	SNAP	Mutationassesor
1) Series 1					
XH026	**p.R119H**	**Probably damaging**	**Damaging**	**Non-neutral**	**High**
XH070	**p.P145L**	**Probably damaging**	**Damaging**	**Non-neutral**	Low
XH171	**p.P145L**	**Probably damaging**	**Damaging**	**Non-neutral**	Low
XH185	p.P310L	**Probably damaging**	Tolerated	Neutral	Neutral
XH188	p.R167C	**Possibly damaging**	Tolerated	Neutral	Neutral
XH264	p.G234R	Benign	Tolerated	**Non-neutral**	Neutral
XH275	p.G395D	Benign	Tolerated	**Non-neutral**	Neutral
2) Series 2					
XR329	**p.P145L**	**Probably damaging**	**Damaging**	**Non-neutral**	Low
XR421	**p.D142Y**	**Probably damaging**	**Damaging**	**Non-neutral**	Medium
	p.S383L	Benign	Tolerated	Neutral	Neutral
XR445	p.P434S	Benign	**Damaging**	**Non-neutral**	Neutral
3) Controls					
H039	p.A354V	**Probably damaging**	Tolerated	**Non-neutral**	Neutral
H040	p.R224Q	Benign	Tolerated	Neutral	Low
H070	p.R224G	Benign	**Damaging**	**Non-neutral**	Medium

* The potential pathogenic roles predicted by the bioinformatic tools are shown in bold. The *TBX6* missense mutations that are suggested to be functional by all or at least three bioinformatic tools are also shown in bold.

Table S4. Eight rare missense mutations of *TBX6* identified in the 1000 Genome Project.*

SNP No.*	*TBX6* missense mutation	Allele frequency in CHB+CHS populations†	Allele frequency in all populations†
rs201241148	c.400G>A (p.E134K)	0.25% (1/394)	0.05% (1/2184)
rs201656580	c.406C>T. (p.R136C)	None	0.05% (1/2184)
rs199605335	c.457T>C (p.W153R)	None	0.05% (1/2184)
rs56098093	c.484G>A (p.G162S)	None	0.05% (1/2184)
rs201799291	c.500G>A (p.R167H)	None	0.05% (1/2184)
rs61738521	c.699G>C (p.W233C)	None	0.05% (1/2184)
rs201231713	c.815G>A (p.R272Q)	None	0.05% (1/2184)
rs149105120	c.1010C>T (p.P337L)	None	0.27% (6/2184)

* The rare mutations have allele frequencies of < 1%.

† Based on the data of the 1000 Genomes Project. Each person is supposed to have two *TBX6* alleles.

Table S5. The haplotype distribution of common *TBX6* SNPs in the human populations studied in the 1000 Genome Project.*

Population	CHB+CHS	CEU	YRI
Individual/allele numbers	197/394	85/170	88/176
TBX6 haplotype (%)†			
C-T-C	170 (43.1%)	93 (54.7%)	111 (63.1%)
T-C-A	**175 (44.4%)**	**56 (32.9%)**	**1 (0.6%)**
C-T-A	45 (11.4%)	17 (10.0%)	64 (36.3%)
T-T-C	2 (0.5%)	4 (2.4%)	0
T-C-C	1 (0.3%)	0	0
C-C-A	1 (0.3%)	0	0

* CHB+CHS, two Han Chinese populations. CEU, a Caucasian population. YRI, an African population

† The haplotype defined by three common *TBX6* SNPs (reference /non-reference): rs2289292 (C/T) - rs3809624 (T/C) - rs3809627 (C/A). The risk haplotype for congenital scoliosis is shown in bold.

Table S6. The *TBX6* haplotype distributions in the multi-center Series 3 of 16p11.2 deletion.

Center	Subject	Scoliosis phenotype*	Hemizygous *TBX6* haplotype	Method/Reference
Baylor College of Medicine	BL01	NO	C-T-C	2
	BL02	NO	C-T-C	2
	BL03	NO	C-T-C	2
	BL05	NO	C-T-C	2
	BL08	NO	C-T-C	2
Washington University School of Medicine	PT01	scoliosis w/o mal.	C-T-C	3
	PT02	scoliosis w/o mal.	C-T-C	3
	PT03	CS w. mal.	C-T-C	3
	PT04	CS w. mal.	T-C-A	3
	PT05	scoliosis w/o mal.	T-C-A	3
	PT07	scoliosis w/o mal.	C-T-C	3
Children's Hospital Central California	PT08	CS w. mal.	T-C-A	3
	PT10	scoliosis w/o mal.	C-T-C	3
Boston Children's Hospital	BS01	CS w. mal.	T-C-A	1
	BS02	NO	T-C-A	1
	BS03	NO	C-T-C	1
	BS04	NO	C-T-C	1
	BS05	NO	C-T-C	1
	BS06	NO	C-T-C	1
	BS07	NO	C-T-C	1
	BS08	NO	C-T-C	1
	BS09	NO	T-C-A	1
	BS10	scoliosis w/o mal.	T-C-A	1
	BS11	NO	C-T-C	1
	BS12	NO	C-T-C	1
	BS13	NO	C-T-C	1
	BS14	NO	T-C-A	1
	BS15	NO	C-T-A	1
	BS16	NO	C-T-C	1
	BS17	NO	C-T-A	1
	BS18	NO	C-T-C	1
	BS19	CS w. mal.	T-C-A	1
	BS20	NO	C-T-C	1
	BS21	NO	T-C-A	1
	BS22	NO	C-T-A	1
	BS23	NO	C-T-C	1
	BS24	NO	C-T-C	1
	BS25	NO	C-T-A	1
	BS26	NO	T-C-A	1
	BS27	NO	C-T-C	1
Shanghai Children's Hospital	ET01	NO	C-T-C	4×180k custom CGH
	ET02	CS w. mal.	T-C-A	Agilent 1×1M CGH

* CS w. mal. (congenital scoliosis with vertebral malformations), scoliosis w/o mal. (scoliosis without vertebral malformations), NO (no evidence for scoliosis).

Table S7. The biased distribution of the hemizygous *TBX6* risk haplotype to congenital scoliosis in the multi-center Series 3 of 16p11.2 deletion.*

Hemizygous *TBX6* haplotype	Vertebral phenotype	
	CS with vertebral malformations	No evidence for scoliosis
T-C-A	5 (83.3%)	5 (16.7%)
Remaining	1 (16.7%)	25 (83.3%)
All	6	30

* The haplotype defined by three common *TBX6* SNPs (reference/non-reference): rs2289292 (C/T) - rs3809624 (T/C) - rs3809627 (C/A).

Table S8. Vertebral malformations in *TBX6*-associated CS patients.

Patient	Number of the vertebral segments with malformations (%)	
	Hemivertebra/hypoplasia	Other malformations
TBX6-associated	28 (82.4%)	6 (17.6%)
No *TBX6* mutation	345 (37.1%)	585 (62.9%)
P	< 0.001	
OR (95% CI)	7.91 (3.24-19.30)	

References

1. Shen Y, Irons M, Miller DT, et al. Development of a focused oligonucleotide-array comparative genomic hybridization chip for clinical diagnosis of genomic imbalance. Clin Chem 2007;53:2051-9.

2. Boone PM, Bacino CA, Shaw CA, et al. Detection of clinically relevant exonic copy-number changes by array CGH. Hum Mutat 2010;31:1326-42.

3. Al-Kateb H, Khanna G, Filges I, et al. Scoliosis and vertebral anomalies: additional abnormal phenotypes associated with chromosome 16p11.2 rearrangement. Am J Med Genet A 2014;164A:1118-26.

4. Li L, Gong H, Yu H, et al. Knockdown of nucleosome assembly protein 1-like 1 promotes dimethyl sulfoxide-induced differentiation of P19CL6 cells into cardiomyocytes. J Cell Biochem 2012;113:3788-96.

5. Gavrilov S, Nuhrenberg TG, Ashton AW, et al. Tbx6 is a determinant of cardiac and neural cell fate decisions in multipotent P19CL6 cells. Differentiation 2012;84:176-84.

6. Okita K, Matsumura Y, Sato Y, et al. A more efficient method to generate integration-free human iPS cells. Nat Methods 2011;8:409-12.

7. Beers J, Gulbranson DR, George N, et al. Passaging and colony expansion of human pluripotent stem cells by enzyme-free dissociation in chemically defined culture conditions. Nat Protoc 2012;7:2029-40.

8. Lian X, Zhang J, Azarin SM, et al. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions. Nat Protoc 2013;8:162-75.

9. The 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.

10. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

11. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.

12. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.

13. Bromberg Y, Yachdav G, Rost B. SNAP predicts effect of mutations on protein function. Bioinformatics 2008;24:2397-8.

14. Li B, Krishnan VG, Mort ME, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 2009;25:2744-50.