*Human Molecular Genetics*, 2017, Vol. 26, No. 8 1577–1583

doi: 10.1093/hmg/ddx045
Advance Access Publication Date: 28 February 2017
Association Studies Article

ASSOCIATION STUDIES ARTICLE

# Genome-wide association study identifies novel susceptible loci and highlights *Wnt/beta-catenin* pathway in the development of adolescent idiopathic scoliosis

Zezhang Zhu[1,2,†], Leilei Xu[1,2,†], Nelson Leung-Sang Tang[2,3,4,†], Xiaodong Qin[1,2], Zhenhua Feng[1,2], Weixiang Sun[1,2], Weiguo Zhu[1,2], Benlong Shi[1], Peng Liu[5], Saihu Mao[1,2], Jun Qiao[1], Zhen Liu[1], Xu Sun[1], Fangcai Li[6], Jack Chun-Yiu Cheng[2,7,†] and Yong Qiu[1,2,†,*]

[1]Department of Spine Surgery, The Drum Tower Hospital of Nanjing University Medical School, Nanjing 210008, P.R. China, [2]Joint Scoliosis Research Center of The Chinese University of Hong Kong and Nanjing University, Nanjing 210008 & Hong Kong 999077, P.R. China, [3]Department of Chemical Pathology and School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong 999077, P.R. China, [4]Li Ka Shing Institute of Health Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong 999077, P.R. China, [5]Department of Orthopaedics, China-Japan Union Hospital of Jilin University, Changchun 130116, P.R. China, [6]Department of Orthopaedics, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou 310002, P.R. China and [7]Department of Orthopaedics and Traumatology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong 999077, P.R. China

*To whom correspondence should be addressed at: Department of Spine Surgery, the Drum Tower Hospital of Nanjing University Medical School, Zhongshan Road No. 321, Nanjing 210008, P.R. China. Tel: +86 02568182222; Fax +86 02568183070; Email: scoliosis2002@sina.com



## Abstract

The genetic architecture of adolescent idiopathic scoliosis (AIS) remains poorly understood. Here we present the result of a 4-stage genome-wide association study composed of 5,953 AIS patients and 8,137 controls. Overall, we identified three novel susceptible loci including rs7593846 at 2p14 near *MEIS1* ($P_{combined} = 1.19 \times 10^{-13}$, OR = 1.21, 95% CI = 1.10–1.32), rs7633294 at 3p14.1 near *MAGI1* ($P_{combined} = 1.85 \times 10^{-12}$, OR = 1.20, 95% CI = 1.09–1.32), and rs9810566 at 3q26.2 near *TNIK* ($P_{combined} = 1.14 \times 10^{-11}$, OR = 1.19, 95% CI = 1.08-1.32). We also confirmed a recently reported region associated with AIS at 20p11.22 ($P_{combined} = 1.61 \times 10^{-15}$, OR = 1.22, 95% CI = 1.12–1.34). Furthermore, we observed significantly asymmetric expression of *Wnt/beta-catenin* pathway in the bilateral paraspinal muscle of AIS patients, including *beta-catenin, TNIK,* and *LBX1*. This is the first study that unveils the potential role of *Wnt/beta-catenin* pathway in the development of AIS, and our findings may shed new light on the etiopathogenesis of AIS.

[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors and that the last two authors should be regarded as joint Last Authors.

**Received:** October 14, 2016. **Revised:** January 31, 2017. **Accepted:** February 1, 2017

## Introduction

Adolescent idiopathic scoliosis (AIS) is a 3-dimensional spinal deformity that develops during the puberty (1). To date, the genetic etiology of AIS has been extensively investigated in different populations (2–5). Genome-wide association studies (GWASs) conducted in the Japanese and the Caucasian population have revealed several susceptible genes of AIS, including *LBX1* (6), *GPR126* (7), *BNC2* (8), *PAX1* (9), and *CHL1* (10). We previously performed a GWAS investigating the susceptibility of AIS in the Chinese girls, identifying three novel susceptibility loci including *BCL2*, *PAX3* and *AJAP1* (11). It is noteworthy, however, that these reported variants all confer a small relative risk of AIS, which are far insufficient to provide a full understanding of the genetic background of AIS.

To illustrate the genetic architecture of AIS, we expanded our previous GWAS by including 1,503 patients and 2,123 controls in the discovery stage, the results of which were further replicated in 4,450 patients and 6,014 controls. Overall, we identified three novel susceptible loci at 2p14 near *MEIS1* (rs7593846, $P_{combined} = 1.19 \times 10^{-13}$), at 3p14.1 near *MAGI1* (rs7633294, $P_{combined} = 1.85 \times 10^{-12}$), and at 3q26.2 near *TNIK* (rs9810566, $P_{combined} = 1.14 \times 10^{-11}$). We also confirmed a recently reported region associated with AIS at 20p11.22 (rs6047663, $P_{combined} = 1.61 \times 10^{-15}$). Furthermore, we observed significantly asymmetric expression of *Wnt/beta-catenin* pathway in the bilateral paraspinal muscle of AIS patients, including *beta-catenin, TNIK,* and *LBX1*. For the first time we highlight the potential role of *Wnt/beta-catenin* pathway in the development of AIS, and our findings may shed new light on the etiology of AIS.

## Results

### Association analyses in the discovery and replication stages

The current 4-stage GWAS recruited 5,953 female AIS patients and 8,137 controls from the Chinese population. We routinely performed the quality control procedures for the sample and the genotyping data as previously mentioned (11). The principal-component analysis (PCA) showed that almost all subjects were clustered in the Chinese population (Supplementary Material, Fig. S1). A genomic inflation factor of 1.13 was observed through quantile-quantile (Q-Q) plot, indicating a low possibility of population stratification (Supplementary Material, Fig. S2). As shown by the Manhattan plot (Fig. 1), 56 SNPs were found to surpass the threshold of $P < 1.0 \times 10^{-5}$ that was indicative of the suggestive whole-genome significance.

For each haplo blocks, we selected one representative SNP with the lowest *P* value for further replication. After excluding the variants reported by our previous GWAS, a total of 24 representative SNPs (Supplementary Material, Table S2) were then replicated in 1,885 cases and 1,989 controls. Significant associations were found in 4 SNPs, which were further genotyped in 1,813 cases and 2,675 controls. Finally, these 4 SNPs were replicated in a cohort of 752 cases and 1,350 controls. Combining the results of the discovery stage and three replication stages, we observed 4 SNPs that surpassed the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$, including rs7593846 at 2p14, rs7633294 at 3p14.1, rs9810566 at 3q26.2, and rs6047663 at 20p11.22. To illustrate the spectrum of disease-related variants in the susceptible regions, we performed imputation analysis covering a 800 kb region around each SNP. The logistic regression model showed a much stronger signal (rs966604) at 20p11.22 than rs6047663. As for the other three susceptible loci, we observed similar associations with those of the discovery stage (Supplementary Material, Table S4). The strongest evidence of association was attained with rs6047663 ($P_{combined} = 1.61 \times 10^{-15}$, OR = 1.22, 95% CI = 1.12–1.34, $P_{het} = 0.35$) located 90 kb downstream of *PAX1* (Table 1, Fig. 2A, and Supplementary Material, Fig. S3A). SNP rs9810566 ($P_{combined} = 1.14 \times 10^{-11}$, OR = 1.19, 95% CI = 1.08-1.32, $P_{het} = 0.19$) was located in the intron of the *TNIK* (Table 1, Fig. 2B, and Supplementary Material, Fig. S3B). SNP rs7593846 ($P_{combined} = 1.19 \times 10^{-13}$, OR = 1.21, 95% CI = 1.10–1.32, $P_{het} = 0.85$) maps 140kb downstream of *MEIS1* (Table 1, Fig. 2C, and Supplementary Material, Fig. S3C). SNP rs7633294 ($P_{combined} = 1.85 \times 10^{-12}$, OR = 1.20, 95% CI = 1.09–1.32, $P_{het} = 0.59$) was located in the intronic region of *MAGI1* (Table 1, Fig. 2D, and Supplementary Material, Fig. S3D). To further validate the association between the these susceptible region and AIS, we selected other significant SNPs in the LD block of each novel locus for replication, including rs10490189 of *MEIS1*, rs952209 of *TNIK*, rs11131034 of *MAGI1* and rs966604 of *PAX1*. All the 4 variants were significantly associated with AIS (Supplementary Material, Table S5).

### Analysis of gene expression in the paraspinal muscle

Significantly decreased expression of *beta-catenin, TNIK, TCF4, PAX3, LBX1, MyoD, MEIS1,* and *PAX1* was observed in the concave side of the patients ($n = 124$) as compared with that in the convex side (Supplementary Material, Table S10). The expression level of these genes was remarkably correlated with each other (Supplementary Materials, Tables S11 and S12).

### Fiber-type composition of the paraspinal muscle in AIS

The distribution of type I and type II fiber in the paraspinal muscle of 11 AIS patients was summarized in Supplementary Material, Table S14. As shown in Supplementary Material, Fig. S4, the rate of Type I fiber was significantly lower in the concave side than in the convex side (49.5% ± 8.0% vs. 61.5% ± 7.7%, $P = 0.002$). In addition, the concave/convex ratio of type I fiber was significantly correlated with the mRNA expression of *MyoD* in the paraspinal muscle of AIS patients ($r = 0.67$, $P = 0.02$) (Supplementary Material, Fig. S5).

### Functional annotation

The overlap of the novel associated SNPs with the Encyclopedia of DNA Elements (ENCODE)-annotated genomic elements (12) were summarized in Supplementary Materials, Tables S6–S9. Regulatory activities, such as histone modifications, DNase hypersensitivity sites, binding proteins and altered motifs, were observed for the 4 associated SNPs or their proxies.

## Discussion

In the current 4-stage GWAS, we identified three novel SNPs associated with the susceptibility of AIS in the Chinese girls. The most compelling candidate for functional analysis was attained with rs9810566 located in the intron of *TNIK*. Previous studies have identified *TNIK* as a regulatory component of the *beta-catenin* transcriptional complex (13). Besides, *TNIK* was reported to regulate the *Wnt/beta-catenin* pathway in the downstream part (14–16). It is noteworthy that *LBX1* and *PAX3*, two previously reported susceptible genes of AIS, are both involved in the downstream of *Wnt/beta-catenin* pathway (17,18).

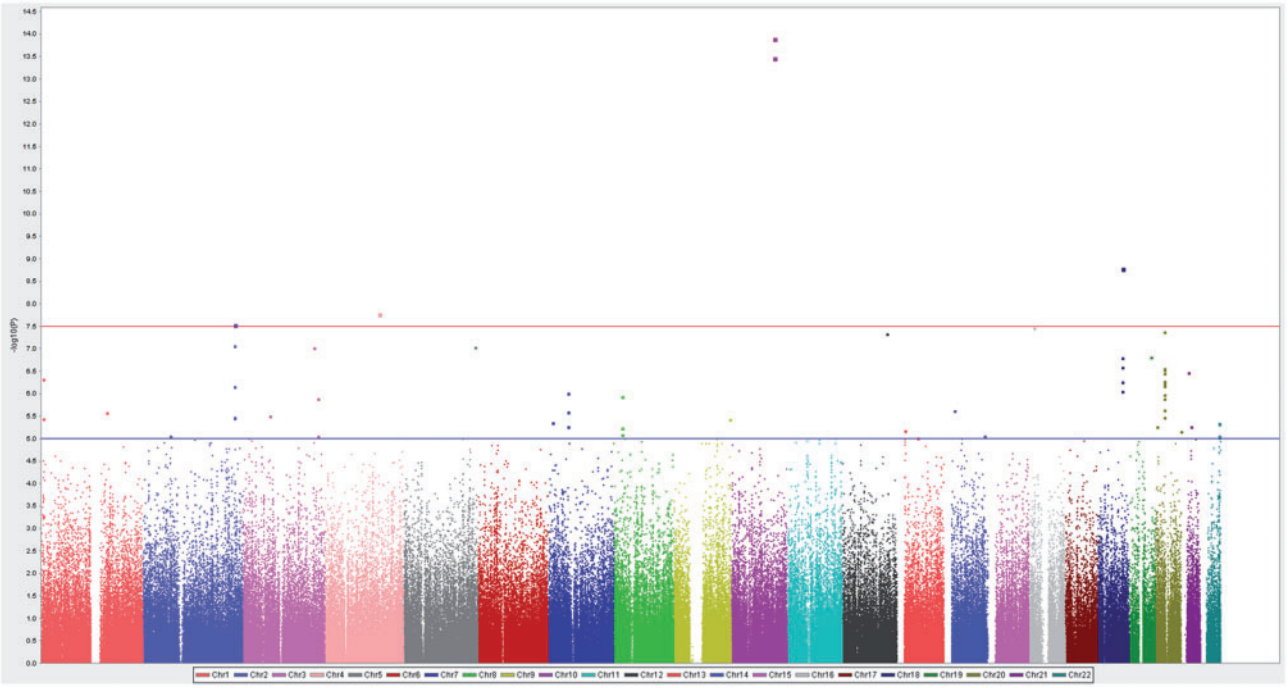

**Figure 1.** The Manhattan plot showing negative log10-transformed *P* values in the discovery stage (1456 cases and 2104 controls). The blue line represents the threshold of the suggestive whole-genome significance ($P < 1.0 \times 10^{-5}$). The red line represents the threshold of genome-wide significance threshold ($P < 1.0 \times 10^{-8}$).

Speculating that *Wnt/beta-catenin* pathway could play a role in the development of AIS, we analyzed the expression of its key components in the bilateral paraspinal muscle of the patients. We observed that the expression level of these genes was highly correlated with each other, which echoed the regulatory mechanism between the *Wnt/beta-catenin* pathway and its downstream elements as reported by previous studies (17–19). Previous studies showed that *MyoD* could be implicated in the development of skeletal muscle (20,21). Also, it was reported to play a role in the fiber-type composition of skeletal muscle (22). Interestingly, an abnormality of the paraspinal muscle has been widely investigated in AIS patients (23–25). Asymmetric distribution of type I and type II fiber has been observed between the concave and the convex side of the curve (25,26). In this study, we confirmed that the proportion of type I fiber was significantly lower in the concave side than in the convex side through immunohistochemical staining. In addition, we observed that the expression level of *MyoD* was remarkably correlated with the proportion of type I fiber. Therefore, it is possible that *MyoD* may be a biologically functional element involved in the mediation of a disproportional muscle fiber between the bilateral paraspinal muscle of the patients. Further studies are necessary to clarify the role of *Wnt/beta-catenin* pathway and its downstream components in the pathogenesis of AIS.

SNP rs7593846 maps 140kb downstream of *MEIS1* gene. Imputation analysis showed that the linkage disequilibrium (LD) block of rs7593846 at 2p14 contains a wide range of risk variant associated with AIS. *MEIS1* functions as a DNA-binding cofactor of *Hox* proteins (27). It was reported that *MEIS1* can be involved in the activation of *Wnt/beta-catenin* signaling in the retina (28). To date, the physiological function of *MEIS1* remains unclear, although it was widely reported to be associated with leukemogenesis (29). We observed that the expression of *MEIS1* was significantly correlated with that of *beta-catenin*. Taken together, these findings suggest that *MEIS1* is a promising candidate gene to explain the relationship between the *Wnt/beta-catenin* pathway and AIS.

SNP rs7633294 was located in the intron of *MAGI1* gene. Data from ENCODE suggest that rs7633294 resides within a region containing different motifs. *MAGI1* is an important molecule for the stabilization of cadherin-mediated cell interactions (30). It associates with a variety of binding molecules such as *beta-catenin* (31). *MAGI1* may be a physiological regulator of *Wnt/beta-catenin* signaling in colon cancer cells (32). In this study, however, we failed to detect the expression of *MAGI1* in the paraspinal muscle of AIS patients. Further studies are warranted to investigate the relationship between *MAGI1* and the development of AIS.

SNP rs6047663 was located 90 kb downstream of *PAX1*. Previous study showed that the PAX1-encoding region could be associated with spinal development (33). Genetic variation rs6137473 of *PAX1* has been identified as a risk locus for AIS in a previous GWAS (9). The functional analysis showed that rs6137473 could be located in an enhancer regulating *PAX1* expression in specific spinal cells (9). Data of our study validated the role of *PAX1* in the development of AIS at the genome-wide level.

Several limitations of our study need to be addressed. To our knowledge, this is the first study that investigated the correlation between *MyoD* and fiber-type in AIS patients. Fiber-type composition of skeletal muscle can be regulated by multiple genes. It cannot be ruled out that other regulatory genes may be abnormally expressed in AIS patients, thus leading to the asymmetric distribution of muscle fiber as we observed. Further investigations are warranted to clarify the interactions between *MyoD* and those abnormally expressed genes. Second, consistent with previous studies, we also found that a proportion of type I fiber was significantly lower in the concave side than in the convex side of AIS patients. However, it remains obscure

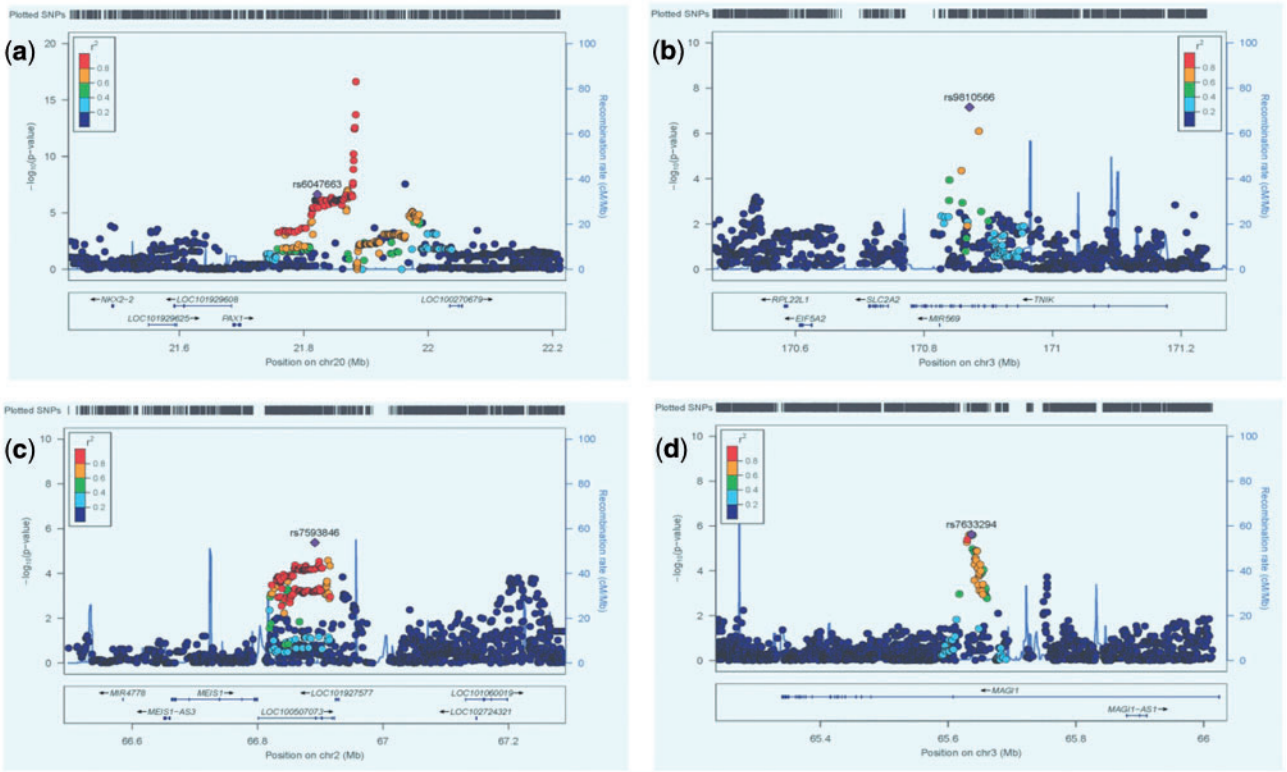


**Figure 2.** Regional association plots of the 4 susceptible loci associated with AIS. Each plot represents −log 10 (*P* value) against the chromosome positions for SNPs in the specific region, together with the positions and orientations of genes. The strongest association signal observed in the discovery stage is indicated by a purple diamond, and the other SNPs are colored according to LD with the proxy SNP. The blue line represents the recombination rate. (**A**) 20p11.22,(**B**) 3q26.2, (**C**) 2p14, (**D**) 3p14.1.

**Table 1.** Association results for four novel loci in the discovery, replication and combined samples

| SNP | Chr. | Genes | MA | Stage | Sample Size | | MAF | | *P* | OR (95%CI) | $P_{het}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Cases | Controls | Cases | Control | | | |
| rs6047663 | 20p11.22 | *PAX1* | G | Discovery | 1433 | 2080 | 0.444 | 0.379 | $4.23 \times 10^{-8}$ | 1.31 (1.18–1.44) | |
| | | | | Replication 1 | 1849 | 1978 | 0.403 | 0.365 | $5.93 \times 10^{-4}$ | 1.17 (1.07–1.29) | |
| | | | | Replication 2 | 1793 | 2628 | 0.416 | 0.370 | $1.24 \times 10^{-5}$ | 1.22 (1.11–1.32) | |
| | | | | Replication 3 | 737 | 1334 | 0.401 | 0.364 | $1.95 \times 10^{-2}$ | 1.17 (1.03–1.33) | |
| | | | | **Combined** | **5812** | **8020** | **0.417** | **0.370** | $\mathbf{1.61 \times 10^{-15}}$ | **1.22 (1.12–1.34)** | **0.35** |
| rs9810566 | 3q26.2 | *TNIK* | A | Discovery | 1446 | 2040 | 0.421 | 0.358 | $9.58 \times 10^{-8}$ | 1.30 (1.18–1.44) | |
| | | | | Replication 1 | 1865 | 1937 | 0.383 | 0.347 | $1.14 \times 10^{-3}$ | 1.17 (1.07–1.28) | |
| | | | | Replication 2 | 1781 | 2665 | 0.380 | 0.352 | $6.47 \times 10^{-3}$ | 1.13 (1.03–1.23) | |
| | | | | Replication 3 | 734 | 1339 | 0.381 | 0.346 | $2.41 \times 10^{-2}$ | 1.16 (1.02–1.33) | |
| | | | | **Combined** | **5826** | **7981** | **0.391** | **0.351** | $\mathbf{1.14 \times 10^{-11}}$ | **1.19 (1.08–1.32)** | **0.19** |
| rs7633294 | 3p14.1 | *MAGI1* | G | Discovery | 1446 | 2059 | 0.387 | 0.333 | $3.18 \times 10^{-6}$ | 1.27 (1.15–1.40) | |
| | | | | Replication 1 | 1852 | 1976 | 0.350 | 0.312 | $4.25 \times 10^{-4}$ | 1.19 (1.08–1.31) | |
| | | | | Replication 2 | 1778 | 2635 | 0.359 | 0.326 | $1.16 \times 10^{-3}$ | 1.16 (1.06–1.27) | |
| | | | | Replication 3 | 738 | 1329 | 0.347 | 0.312 | $2.13 \times 10^{-2}$ | 1.18 (1.02–1.34) | |
| | | | | **Combined** | **5814** | **7999** | **0.362** | **0.322** | $\mathbf{1.85 \times 10^{-12}}$ | **1.20 (1.09–1.32)** | **0.59** |
| rs7593846 | 2p14 | *MEIS1* | G | Discovery | 1430 | 2091 | 0.364 | 0.314 | $8.88 \times 10^{-6}$ | 1.26 (1.14–1.39) | |
| | | | | Replication 1 | 1858 | 1942 | 0.351 | 0.311 | $1.93 \times 10^{-4}$ | 1.20 (1.09–1.32) | |
| | | | | Replication 2 | 1792 | 2643 | 0.352 | 0.315 | $2.03 \times 10^{-4}$ | 1.19 (1.09–1.30) | |
| | | | | Replication 3 | 739 | 1330 | 0.355 | 0.314 | $6.82 \times 10^{-3}$ | 1.20 (1.05–1.38) | |
| | | | | **Combined** | **5819** | **8006** | **0.355** | **0.313** | $\mathbf{1.19 \times 10^{-13}}$ | **1.21 (1.10–1.32)** | **0.85** |

Chr., chromosome; MA, minor allele; MAF, minor allele frequency; OR, odds ratio for the minor allele; 95% CI, 95% confidence intervals; $P_{het}$, *P* value from the heterogeneity test based on the discovery stage and the two replication stage groups.

whether the asymmetric distribution of muscle fiber is primary or secondary to the spinal deformity. In future study, more paraspinal muscles from patients with congenital scoliosis need to be collected as control group so as to solve this puzzle. Third, although we have observed asymmetrical gene expression in AIS patients, we do not have a definitive mechanism to explain this finding. More studies are warranted to clarify the role of asymmetrical gene expression in the development of AIS.

In summary, our GWAS identified three new susceptible genes of AIS at 2p14, 3p14.1, and 3q26.2, and validated a previously reported susceptible region associated with AIS at 20p11.22. We noted that several susceptible genes of AIS may be involved in the *Wnt/beta-catenin* pathway, including *LBX1*, *PAX3*, *TNIK*, *MAGI1* and *MEIS1*. Our experiments indicated that the asymmetric expression of *Wnt/beta-catenin* pathway can be associated with the disproportional paraspinal muscle fiber. The discovery of these novel susceptible genes and the *Wnt/beta-catenin* pathway should shed new light on the pathogenesis of AIS.

## Materials and Methods

### Subjects

The current study was approved by the Ethical Committee of each participating center. Patients who came to our joint scoliosis research center during 2000 April and 2016 October were evaluated for the eligibility to be included in the current study. Female AIS patients with right major thoracic curve were included in this study. Patients were excluded from the study if having scoliosis secondary to known etiology, such as neuromuscular scoliosis, congenital scoliosis, scoliosis associated with skeletal dysplasia and connective tissue abnormalities. The normal controls were recruited from the healthy subjects who came to our hospital for routine physical examinations. Overall, 5,953 AIS patients and 8,137 controls were finally included in our study, all from Chinese Han population. In the discovery stage, 1,230 cases and 2,123 controls were collected in Nanjing, and the other 273 cases were collected in Hong Kong. In replication stage I, all the 1,885 cases and 1,989 controls were collected in Nanjing. In replication stage II, 1180 cases and 2,225 controls were collected in Nanjing, 580 cases and 150 controls were collected in Hong Kong, 25 cases were collected in Hangzhou, and 28 cases and 300 controls were collected in Changchun. In replication stage III, all the 752 cases and 1,350 controls were collected in Nanjing. Informed consent was obtained from all the participants or from the guardians of the children. Supplementary Material, Table S1 summarized clinical characteristics of the subjects in the discovery stage.

### Genotyping and quality control for data in the discovery stage

We extended our previous GWAS by including additional 502 patients and 623 controls in the discovery stage. The Affymetrix Genome-Wide Human SNP Array 6.0 was used for the genotyping analysis of these subjects, and the data were merged with that of the previous GWAS. Quality control (QC) on the raw genotyping data was performed using PLINK (v1.90) as previously described (11). The following criteria were used to exclude unqualified SNPs: 1. with call rates $< 95\%$; 2. with minor allele frequency (MAF) $< 0.01$; 3. with Hardy-Weinberg equilibrium deviation ($P < 1.0 \times 10^{-5}$); 4. on non-autosomal Chromosomes. For sample QC, subjects with more than 10% missing genotypes ($n = 17$) or showing second-degrees relatedness ($n = 49$) were removed. Population stratification was detected by PCA implemented in the software package EIGENSTRAT (34). A total of 3,560 subjects were analyzed together with 210 HapMap subjects, including 60 African (YRI), 60 European (CEU), 45 Japanese (JPT) and 45 Han Chinese (CHB) individuals. Using the top two associated principal components, we identified that all cases and controls were of Chinese ancestry with minimal evidence of population stratification (Supplementary Material, Fig. S1). After the QC process, a total of 1456 cases and 2104 controls, and 435,183 SNPs were used in the final analysis.

### Selection of SNPs into the replication stage

56 SNPs were observed to surpass the threshold of suggestive genome-wide significance ($P < 1.0 \times 10^{-5}$) in the discovery stage. 14 variants were excluded since they had been replicated in our previous GWAS. We selected 24 representative SNPs for replication stage I and removed the other 18 SNPs due to high LD with at least 1 of the 24 representative SNPs (Supplementary Materials, Table S2 and S3). In replication stage I, 4 SNPs were significantly associated with AIS and further included in replication stage II and III (Table 1, Fig. 1). The genotyping analysis was performed using Sequenom MassARRAY system (Sequenom Inc.) in replication stage I, and using TaqMan SNP Genotyping Assay (Applied Bio systems, Foster City, CA) in replication stage II and III. For Sequenom MassARRAY analysis, DNA samples with $>10\%$ missing genotyping were excluded. For TaqMan SNP Genotyping Assay, 10% of the samples were randomly selected for blind duplication, yielding a reproducibility rate of 100%.

### Imputation

MaCH-Admix software was used to perform genotype imputation, with the LD and haplotype information from the 1000 Genomes Project (phased CHB and CHS data; March 2012 release) used as the reference (35). We prepared the input files after excluding SNPs with a low imputation quality ($R^2 \leq 0.30$) or with MAF $\leq 0.10$. The association of the SNPs with the AIS was calculated by Logistic regression analysis. The online tool Locus Zoom was used to plot the genomic regions that covered 800kb around the 4 susceptible loci (36). According to the results of imputation analysis, the most other significant SNPs in the LD block of each novel locus was selected for replication in 1,885 cases and 1,989 controls, including rs10490189 of *MEIS1*, rs952209 of *TNIK*, rs11131034 of *MAGI1* and rs966604 of *PAX1*.

### Statistical analysis

PLINK 1.90 was used for general statistical analysis (37). In the discovery stage, the association of each SNP with AIS was calculated by Cochran-Armitage trend test as previously described. Haploview (v2.6) was used to generate the Manhattan plot presenting the -log10 *P* value of each SNP (38). Data of the discovery stage and the two replication stages were combined using the Mantel-Haenszel method, with Q-statistic *P* value used to indicate the heterogeneity among the three stages (39). Quantile-quantile (Q-Q) plot was constructed using the observed *P* values against expected *P* values. The inflation factor value was calculated to evaluate the population stratification (Supplementary Material, Fig. S2).

### Functional annotation

Chromatin state segmentation in LCL data generated by the ENCODE project was explored to annotate the regulatory properties of the significant signals (12). We utilized HaploReg to examine whether the most associated SNPs or their proxies can annotate enhancer elements or putative transcription factor binding sites (Supplementary Materials, Tables S6–S9) (40).

### Collection of paraspinal muscle sample

Paraspinal muscle was collected from 124 female AIS patients (mean age, 16.4 ± 3.1 years). All patients or their guardians gave the informed consent for the collection of paraspinal muscle. Deep paraspinal muscle biopsies of 1.5 × 1.5 × 1.5 $cm^3$ were taken at two different sites including the concave side and the convex side at the apex of the curve. All the muscle samples were collected in Nanjing of Jiangsu Province.

### Quantitative real-time PCR

Total RNA was extracted from each paraspinal muscle tissue using TRIzol reagent (Invitrogen). After digestion with DNase, cDNA was synthesized with a random primer from 1 μl RNA using a reverse transcription kit (Takara, Japan) according to the manufacture's instruction. Quantification of mRNA was performed with the SYBR Premix ExTaq II (Takara, Japan) by Roche Light Cycler 480 II instrument (Roche Diagnostics, Mannheim, Germany). GAPDH expression was used as an internal reference and detailed sequences for the primers can be obtained in Supplementary Material, Table S13. All reactions were carried out in triplicate to reduce variation. The Student t test was used to compare the gene expression between the concave side and the convex side of the spinal curve.

### Immunohistochemical staining

Muscle cross-sections (8 μm) were prepared from unfixed OCT embedded samples of 11 AIS patients, which were quickly frozen in isopentane cooled in liquid nitrogen. Skeletal muscle fiber typing was performed using metachromatic dye-ATPase methods as previously described (41). Sections were pre-incubated at alkaline (pH9.4) and at acid (pH4.6), respectively. Muscle fibers were classified as Type I or II by analyzing at least 500 fibers in each specimen using ImageJ software (National Institutes of Health, Bethesda, Maryland). The proportion of type I fiber and type II fiber were summarized in Supplementary Material, Table S14 for each patient.

## Sites

1000 Genomes Project, http://www.1000genomes.org/; date last accessed February 15, 2017
LocusZoom, http://csg.sph.umich.edu/locuszoom/; date last accessed February 15, 2017
R v2.13.0, http://www.r-project.org/; date last accessed February 15, 2017
UCSC Genome Browser (ENCODE data), http://genome.ucsc.edu/ENCODE/; date last accessed February 15, 2017
HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php; date last accessed February 15, 2017
Mach-Admix, http://www.unc.edu/~yunmli/MaCH-Admix/download.php; date last accessed February 15, 2017
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; date last accessed February 15, 2017
EIGENSTRAT, http://genetics.med.harvard.edu/reich/Reich_Lab/Software.html; date last accessed February 15, 2017

## Supplementary Material

Supplementary Material is available at *HMG* online.

## Acknowledgements

We gratefully thank Prof. Long Yi and Prof. Yaping Wang from the Nanjing University Medical School for useful advice. We also sincerely thank Dr. Bin Wang, Dr. Zhou Wang, Dr. Fei Wang and Dr. Jun Jiang for their contributions to the collection of samples.

*Conflict of Interest statement*. None declared.

## Funding

Science and Technology Project of Jiangsu Province's Clinical Medicine (BL2012002), Jiangsu Province's Key Medical Center (ZX201107), the Natural Science Foundation of China (No. 81501849, No.81661168013, No.81171672, & No. 81501932), Natural Science Foundation of Jiangsu Province (BK20150099) and Nanjing Medical Science and technology development Foundation (Grant No. YKK15060).

## References

1. Grivas, T.B., Vasiliadis, E., Savvidou, O., Mouzakis, V. and Koufopoulos, G. (2006) Geographic latitude and prevalence of adolescent idiopathic scoliosis. *Stud. Health. Technol. Inform.*, **123**, 84–89.
2. Buchan, J.G., Alvarado, D.M., Haller, G.E., Cruchaga, C., Harms, M.B., Zhang, T., Willing, M.C., Grange, D.K., Braverman, A.C., Miller, N.H., *et al.* (2014) Rare variants in FBN1 and FBN2 are associated with severe adolescent idiopathic scoliosis. *Hum. Mol. Genet.*, **23**, 5271–5282.
3. Gao, X., Gordon, D., Zhang, D., Browne, R., Helms, C., Gillum, J., Weber, S., Devroy, S., Swaney, S., Dobbs, M., *et al.* (2007) CHD7 gene polymorphisms are associated with susceptibility to idiopathic scoliosis. *Am. J. Hum. Genet.*, **80**, 957–965.
4. Mao, S., Xu, L., Zhu, Z., Qian, B., Qiao, J., Yi, L. and Qiu, Y. (2013) Association between genetic determinants of peak height velocity during puberty and predisposition to adolescent idiopathic scoliosis. *Spine (Phila Pa 1976)*, **38**, 1034–1039.
5. Xu, L., Huang, S., Qin, X., Mao, S., Qiao, J., Qian, B.P., Qiu, Y. and Zhu, Z. (2015) Investigation of the 53 Markers in a DNA-Based Prognostic Test Revealing New Predisposition Genes for Adolescent Idiopathic Scoliosis. *Spine (Phila Pa 1976*, **40**, 1086–1091.
6. Takahashi, Y., Kou, I., Takahashi, A., Johnson, T.A., Kono, K., Kawakami, N., Uno, K., Ito, M., Minami, S., Yanagida, H., *et al.* (2011) A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. *Nat. Genet.*, **43**, 1237–1240.
7. Kou, I., Takahashi, Y., Johnson, T.A., Takahashi, A., Guo, L., Dai, J., Qiu, X., Sharma, S., Takimoto, A., Ogura, Y., *et al.* (2013) Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nat. Genet.*, **45**, 676–679.
8. Ogura, Y., Kou, I., Miura, S., Takahashi, A., Xu, L., Takeda, K., Takahashi, Y., Kono, K., Kawakami, N., Uno, K., *et al.* (2015) A Functional SNP in BNC2 Is Associated with Adolescent Idiopathic Scoliosis. *Am. J. Hum. Genet.*, **97**, 337–342.
9. Sharma, S., Londono, D., Eckalbar, W.L., Gao, X., Zhang, D., Mauldin, K., Kou, I., Takahashi, A., Matsumoto, M., Kamiya, N., *et al.* (2015) A PAX1 enhancer locus is associated with

susceptibility to idiopathic scoliosis in females. *Nat. Commun.*, **6**, 6452.
10. Sharma, S., Gao, X., Londono, D., Devroy, S.E., Mauldin, K.N., Frankel, J.T., Brandon, J.M., Zhang, D., Li, Q.Z., Dobbs, M.B., *et al.* (2011) Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. *Hum. Mol. Genet.*, **20**, 1456–1466.
11. Zhu, Z., Tang, N.L., Xu, L., Qin, X., Mao, S., Song, Y., Liu, L., Li, F., Liu, P., Yi, L., *et al.* (2015) Genome-wide association study identifies new susceptibility loci for adolescent idiopathic scoliosis in Chinese girls. *Nat. Commun.*, **6**, 8355.
12. Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
13. Masuda, M., Sawa, M. and Yamada, T. (2015) Therapeutic targets in the Wnt signaling pathway: Feasibility of targeting TNIK in colorectal cancer. *Pharmacol. Ther.*, **156**, 1–9.
14. Mahmoudi, T., Li, V.S., Ng, S.S., Taouatas, N., Vries, R.G., Mohammed, S., Heck, A.J. and Clevers, H. (2009) The kinase TNIK is an essential activator of Wnt target genes. *embo. J.*, **28**, 3329–3340.
15. Park, D.S., Seo, J.H., Hong, M. and Choi, S.C. (2013) Role of the Rap2/TNIK kinase pathway in regulation of LRP6 stability for Wnt signaling. *Biochem. Biophys. Res. Commun.*, **436**, 338–343.
16. Satow, R., Shitashige, M., Jigami, T., Honda, K., Ono, M., Hirohashi, S. and Yamada, T. (2010) Traf2- and Nck-interacting kinase is essential for canonical Wnt signaling in Xenopus axis formation. *J. Biol. Chem.*, **285**, 26289–26294.
17. Fan, C.M., Lee, C.S. and Tessier-Lavigne, M. (1997) A role for WNT proteins in induction of dermomyotome. *Dev. Biol.*, **191**, 160–165.
18. Lou, Q., He, J., Hu, L. and Yin, Z. (2012) Role of lbx2 in the non-canonical Wnt signaling pathway for convergence and extension movements and hypaxial myogenesis in zebrafish. *Biochim. Biophys. Acta*, **1823**, 1024–1032.
19. Buckingham, M., Bajard, L., Chang, T., Daubas, P., Hadchouel, J., Meilhac, S., Montarras, D., Rocancourt, D. and Relaix, F. (2003) The formation of skeletal muscle: from somite to limb. *J. Anat.*, **202**, 59–68.
20. Hughes, S.M., Koishi, K., Rudnicki, M. and Maggs, A.M. (1997) MyoD protein is differentially accumulated in fast and slow skeletal muscle fibres and required for normal fibre type balance in rodents. *Mech. Dev.*, **61**, 151–163.
21. Russo, T.L., Peviani, S.M., Durigan, J.L., Gigo-Benato, D., Delfino, G.B. and Salvini, T.F. (2010) Stretching and electrical stimulation reduce the accumulation of MyoD, myostatin and atrogin-1 in denervated rat skeletal muscle. *J. Mus. Res. Cell Motil.*, **31**, 45–57.
22. Hennebry, A., Berry, C., Siriett, V., O'Callaghan, P., Chau, L., Watson, T., Sharma, M. and Kambadur, R. (2009) Myostatin regulates fiber-type composition of skeletal muscle by regulating MEF2 and MyoD gene expression. *Am. J. Phy. Cell Phy.*, **296**, C525–C534.
23. Gaudreault, N., Arsenault, A.B., Lariviere, C., DeSerres, S.J. and Rivard, C.H. (2005) Assessment of the paraspinal muscles of subjects presenting an idiopathic scoliosis: an EMG pilot study. *BMC Mus. Dis.*, **6**, 14.
24. Kwok, G., Yip, J., Cheung, M.C. and Yick, K.L. (2015) Evaluation of Myoelectric Activity of Paraspinal Muscles in Adolescents with Idiopathic Scoliosis during Habitual Standing and Sitting. *Bio. Res. Int.*, **2015**, 958450.
25. Stetkarova, I., Zamecnik, J., Bocek, V., Vasko, P., Brabec, K. and Krbec, M. (2016) Electrophysiological and histological changes of paraspinal muscles in adolescent idiopathic scoliosis. *Eur. Spine J.*, **25**, 1–8.
26. Meier, M.P., Klein, M.P., Krebs, D., Grob, D. and Muntener, M. (1997) Fiber transformations in multifidus muscle of young patients with idiopathic scoliosis. *Spine (Phila Pa 1976)*, **22**, 2357–2364.
27. Garcia-Cuellar, M.P., Steger, J., Fuller, E., Hetzner, K. and Slany, R.K. (2015) Pbx3 and Meis1 cooperate through multiple mechanisms to support Hox-induced murine leukemia. *Haematologica*, **100**, 905–913.
28. Stephens, W.Z., Senecal, M., Nguyen, M. and Piotrowski, T. (2010) Loss of adenomatous polyposis coli (apc) results in an expanded ciliary marginal zone in the zebrafish eye. *Dev. Dyn.*, **239**, 2066–2077.
29. Roychoudhury, J., Clark, J.P., Gracia-Maldonado, G., Unnisa, Z., Wunderlich, M., Link, K.A., Dasgupta, N., Aronow, B., Huang, G., Mulloy, J.C., *et al.* (2015) MEIS1 regulates an HLF-oxidative stress axis in MLL-fusion gene leukemia. *Blood*, **125**, 2544–2552.
30. Mizuhara, E., Nakatani, T., Minaki, Y., Sakamoto, Y., Ono, Y. and Takai, Y. (2005) MAGI1 recruits Dll1 to cadherin-based adherens junctions and stabilizes it on the cell surface. *J. Biol. Chem.*, **280**, 26499–26507.
31. Feng, X., Jia, S., Martin, T.A. and Jiang, W.G. (2014) Regulation and involvement in cancer and pathological conditions of MAGI1, a tight junction protein. *Anticancer Res.*, **34**, 3251–3256.
32. Zaric, J., Joseph, J.M., Tercier, S., Sengstag, T., Ponsonnet, L., Delorenzi, M. and Ruegg, C. (2012) Identification of MAGI1 as a tumor-suppressor protein induced by cyclooxygenase-2 inhibitors in colorectal cancer cells. *Oncogene*, **31**, 48–59.
33. Wallin, J., Wilting, J., Koseki, H., Fritsch, R., Christ, B. and Balling, R. (1994) The role of Pax-1 in axial skeleton development. *Development*, **120**, 1109–1121.
34. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
35. Liu, E.Y., Li, M., Wang, W. and Li, Y. (2013) MaCH-admix: genotype imputation for admixed populations. *Genet. Epidemiol.*, **37**, 25–37.
36. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.
37. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
38. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
39. MANTEL, N. and HAENSZEL, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer. Inst.*, **22**, 719–748.
40. Ward, L.D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, 930–934.
41. Ogilvie, R.W. and Feeback, D.L. (1990) A metachromatic dye-ATPase method for the simultaneous identification of skeletal muscle fiber types I, IIA, IIB and IIC. *Stain Technol.*, **65**, 231–241.