

ARTICLE

Received 28 Apr 2015 | Accepted 13 Aug 2015 | Published 22 Sep 2015

DOI: 10.1038/ncomms9355

OPEN

# Genome-wide association study identifies new susceptibility loci for adolescent idiopathic scoliosis in Chinese girls

Ze Zhang Zhu<sup>1,2,\*</sup>, Nelson Leung-Sang Tang<sup>2,3,4,5,\*</sup>, Leilei Xu<sup>1,2,\*</sup>, Xiaodong Qin<sup>1,2</sup>, Saihu Mao<sup>1</sup>, Yueming Song<sup>6</sup>, Limin Liu<sup>6</sup>, Fangcai Li<sup>7</sup>, Peng Liu<sup>8</sup>, Long Yi<sup>9</sup>, Jiang Chang<sup>10</sup>, Long Jiang<sup>11</sup>, Bobby Kin-Wah Ng<sup>12</sup>, Benlong Shi<sup>1</sup>, Wen Zhang<sup>1</sup>, Jun Qiao<sup>1,2</sup>, Xu Sun<sup>1,2</sup>, Xusheng Qiu<sup>1,2</sup>, Zhou Wang<sup>1</sup>, Fei Wang<sup>1</sup>, Dingding Xie<sup>1</sup>, Ling Chen<sup>1</sup>, Zhonghui Chen<sup>1</sup>, Mengran Jin<sup>1</sup>, Xiao Han<sup>1</sup>, Zongshan Hu<sup>1</sup>, Zhen Zhang<sup>1</sup>, Zhen Liu<sup>1</sup>, Feng Zhu<sup>1</sup>, Bang-ping Qian<sup>1,2</sup>, Yang Yu<sup>1,2</sup>, Bing Wang<sup>1,2</sup>, K.M. Lee<sup>5</sup>, Wayne Y.W. Lee<sup>12</sup>, T.P. Lam<sup>12</sup>, Yong Qiu<sup>1,2,\*\*</sup> & Jack Chun-Yiu Cheng<sup>2,5,12,\*\*</sup>

Adolescent idiopathic scoliosis (AIS) is a structural deformity of the spine affecting millions of children. As a complex disease, the genetic aetiology of AIS remains obscure. Here we report the results of a four-stage genome-wide association study (GWAS) conducted in a sample of 4,317 AIS patients and 6,016 controls. Overall, we identify three new susceptibility loci at 1p36.32 near *AJAP1* ( $rs241215$ ,  $P_{\text{combined}} = 2.95 \times 10^{-9}$ ), 2q36.1 between *PAX3* and *EPHA4* ( $rs13398147$ ,  $P_{\text{combined}} = 7.59 \times 10^{-13}$ ) and 18q21.33 near *BCL-2* ( $rs4940576$ ,  $P_{\text{combined}} = 2.22 \times 10^{-12}$ ). In addition, we refine a previously reported region associated with AIS at 10q24.32 ( $rs678741$ ,  $P_{\text{combined}} = 9.68 \times 10^{-37}$ ), which suggests *LBX1AS1*, encoding an antisense transcript of *LBX1*, might be a functional variant of AIS. This is the first GWAS investigating genetic variants associated with AIS in Chinese population, and the findings provide new insight into the multiple aetiological mechanisms of AIS.

<sup>1</sup>Department of Spine Surgery, the Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing 210008, China. <sup>2</sup>Joint Scoliosis Research Center of The Chinese University of Hong Kong and Nanjing University, Nanjing 210008, China. <sup>3</sup>Department of Chemical Pathology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China. <sup>4</sup>School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China. <sup>5</sup>Li Ka Shing Institute of Health Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China. <sup>6</sup>Department of Orthopaedic, The West China Hospital, Sichuan University, Chengdu 610000, China. <sup>7</sup>Department of Orthopaedic, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou 310000, China. <sup>8</sup>Department of Orthopaedic, China-Japan Union Hospital of Jilin University, Changchun 130022, China. <sup>9</sup>Jiangsu Key Laboratory for Molecular Medicine, Nanjing University Medical School, Nanjing 210008, China. <sup>10</sup>State Key Laboratory of Environment Health (Incubation), MOE (Ministry of Education) Key Laboratory of Environment & Health, Ministry of Environmental Protection Key Laboratory of Environment and Health (Wuhan), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China. <sup>11</sup>Department of Orthopaedic, Yixing People Hospital, Wuxi 214200, China. <sup>12</sup>Department of Orthopaedics and Traumatology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China. \* These authors contributed equally to this work. \*\* These authors jointly supervised this work. Correspondence and requests for materials should be addressed to Y.Q. (email: scoliosis2002@sina.com) or to J.C.-Y.C. (email: jackcheng@cuhk.edu.hk).

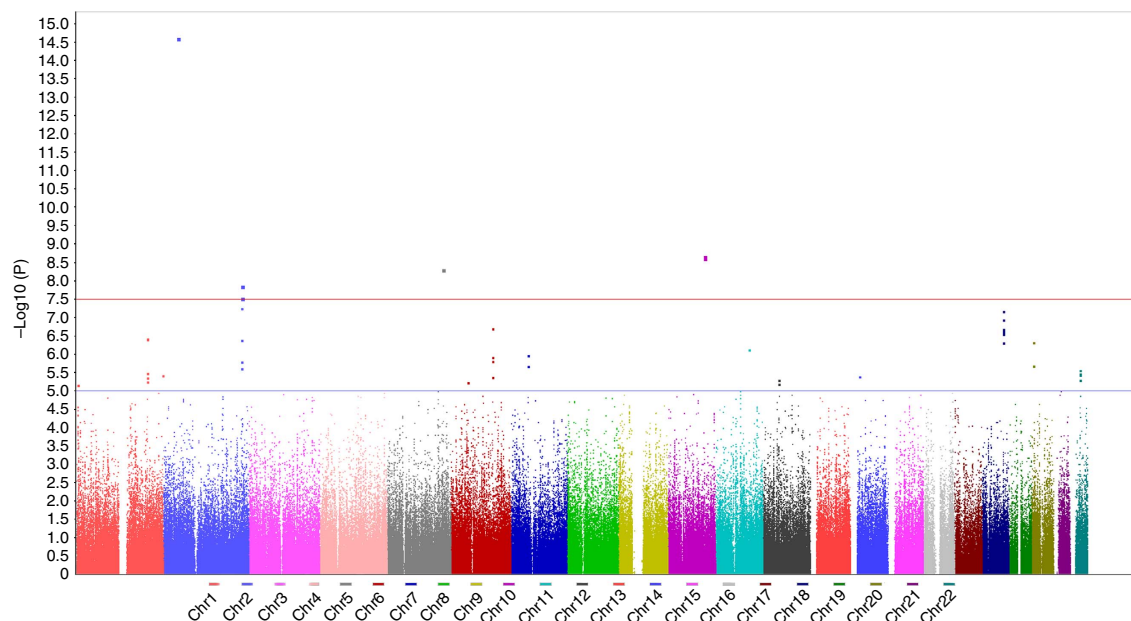
Adolescent idiopathic scoliosis (AIS) is a structural deformity of the spine that is estimated to affect millions of children, with a prevalence of 2–4% (refs 1,2). Commonly understood as a complex trait (polygenic disease), the genetic aetiology of AIS has been widely investigated by linkage analysis and candidate gene association analysis<sup>3–5</sup>. During the era of candidate study, several genes were reported to be associated with AIS, such as *ERα* (ref. 6), *MATN1* (ref. 7), *MTNR1B* (ref. 8) and *TGFβ1* (ref. 9). However, few of them have been successfully replicated in different ethnicities<sup>10–13</sup>. The first genome-wide association study (GWAS) in a Caucasian population reported that rs1400180 of *CHL1* was associated with AIS<sup>14</sup>, which however could not be replicated in a Chinese Han population<sup>15</sup>. Subsequently, a GWAS performed in a Japanese population revealed that *LBX1* and *GPR126* could play a role in the etiopathogenesis of AIS<sup>16,17</sup>. In a recent study that combined the GWAS data of AIS in Caucasian and Japanese populations, Sharma *et al.*<sup>18</sup> reported an enhancer locus of *PAX1* was associated with the susceptibility of female AIS. As estimated by Kou *et al.*<sup>17</sup>, variants of *LBX1* and *GPR126* can only explain ~1% of the trait variance in AIS. Herein additional genetic risk factors need to be discovered for a better understanding of the genetic susceptibility of AIS.

To map novel risk loci for AIS, here we report the results of a four-stage GWAS in a Chinese population. Overall, we identify three new susceptibility loci: 1p36.32 near *AJAPI*, encoding adherens junction associated protein 1 (rs241215,  $P_{\text{combined}} = 2.95 \times 10^{-9}$ ); 2q36.1 between *PAX3* and *EPHA4*, encoding paired box 3 and EPH receptor A4, respectively (rs13398147,  $P_{\text{combined}} = 7.59 \times 10^{-13}$ ); and 18q21.33 near *BCL-2*, encoding B-cell CLL/lymphoma 2 (rs4940576,  $P_{\text{combined}} = 2.22 \times 10^{-12}$ ). In addition, we refine a previously reported region associated with AIS at 10q24.32 (rs678741,  $P_{\text{combined}} = 9.68 \times 10^{-37}$ ), which suggests *LBX1AS1*, encoding an antisense transcript of *LBX1* (lady-bird like homeobox), might be the functional variant of AIS. This is the first GWAS to our knowledge investigating the genetic variants associated with AIS in a Chinese population, and the findings provide new insight into the multiple aetiological mechanisms of AIS.

## Results

**Association analyses in the discovery and replication.** We performed a four-stage GWAS in cohorts of 4,317 patients and 6,016 controls, all of the Chinese Han population. In the discovery stage, 906,703 single nucleotide polymorphisms (SNPs) were genotyped in 1,001 patients presenting a main thoracic curve (major curve located in thoracic region) and 1,500 healthy controls. After quality control filtering of the genotyping data, 516,220 autosomal SNPs in 960 cases and 1,499 controls were retained for the subsequent Cochran–Armitage trend test. Analysis of population stratification with principal-component analysis (PCA) showed that all subjects of our study were clustered in Chinese population (Supplementary Fig. 1). The genomic inflation factor ( $\lambda$ ) as shown by the quantile–quantile plot was 1.031, indicating that there was only a minimal amount of population stratification (Supplementary Fig. 2). As shown by the Manhattan plot (Fig. 1), 43 SNPs were observed to surpass the threshold of suggestive whole-genome significance ( $P < 1.0 \times 10^{-5}$ ) (Supplementary Table 2 and 3).

From each putative haplotype block defined by linkage disequilibrium ( $r^2 > 0.8$ ), we chose one representative SNP with the lowest  $P$  value for replication. Sixteen representative SNPs (Supplementary Table 2) were then included in replication stage 1, composed of 1,916 cases and 2,001 controls. Significant associations were found in four SNPs, which were further genotyped in 1,400 cases and 2,515 controls. Combining the results from the discovery stage and three replication stages, four SNPs were found to reach a genome-wide significance threshold of  $P < 5 \times 10^{-8}$ , including rs241215 at 1p36, rs13398147 at 2q36, rs678741 at 10q24 and rs4940576 at 18p21 (Table 1). We performed imputation analysis around the four significant regions, and the logistic regression model showed similar association with those of the discovery stage (Supplementary Table 4). The strongest evidence of association was attained with rs678741 ( $P_{\text{combined}} = 9.68 \times 10^{-37}$ , odds ratio (OR) = 1.44, 95% confidential interval (CI): 1.37–1.52,  $P_{\text{het}} = 0.95$ ) located in the intron of the *LBX1AS1* (Table 1, Fig. 2a and Supplementary Fig. 3). SNP rs4940576 ( $P_{\text{combined}} = 2.22 \times 10^{-12}$ , OR = 1.23, 95% CI: 1.14–1.29,  $P_{\text{het}} = 0.13$ ) was located in the intronic region

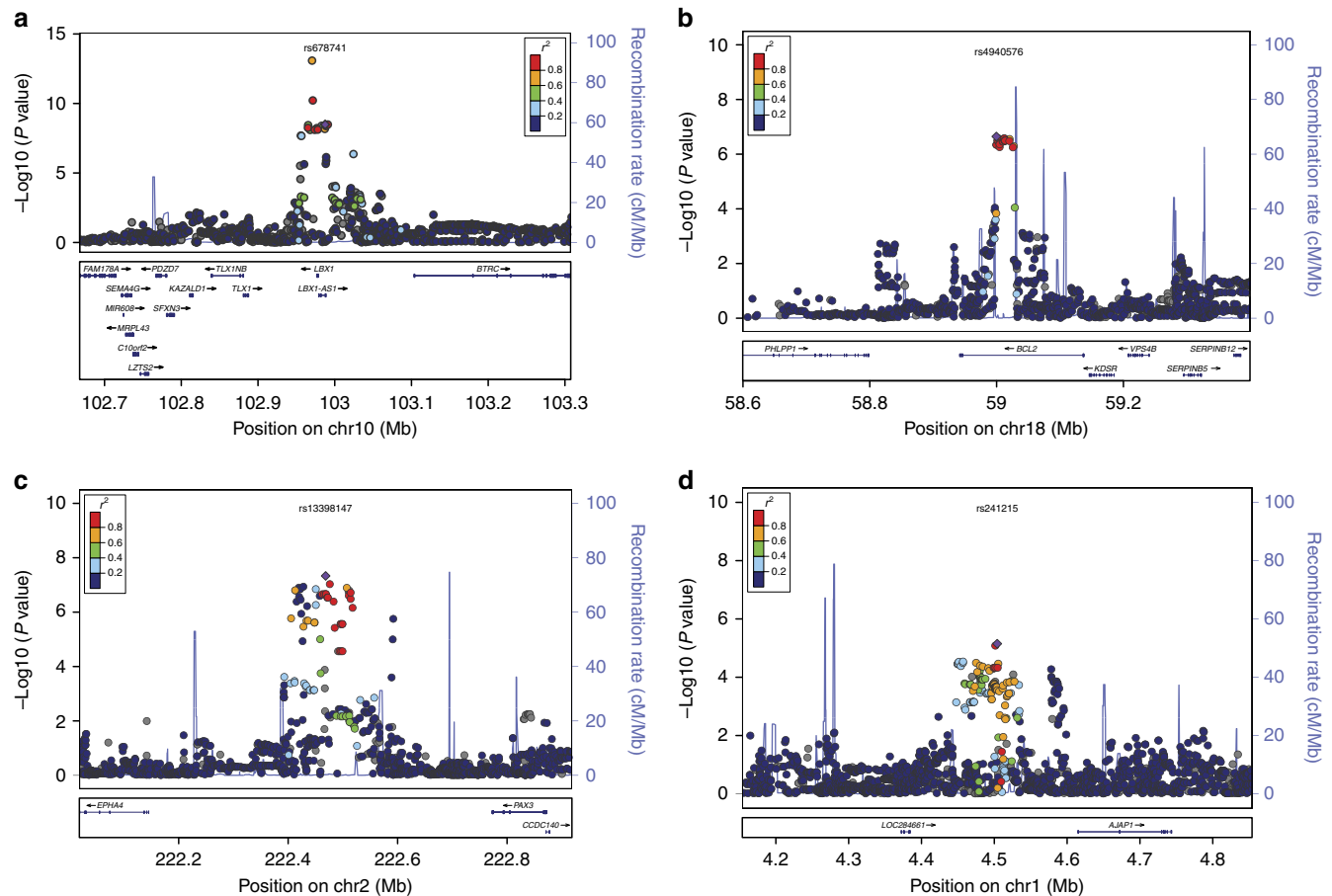


**Figure 1 | Genome-wide association results for AIS in Han Chinese girls.** Here we present Manhattan plot of  $P$  values on the  $-\log_{10}$  scale in the discovery stage (960 cases and 1,499 controls). The x axis shows Chromosomes as delineated by different colours. The y axis shows  $-\log_{10} P$  values of each SNP. The red line represents  $P = 5 \times 10^{-8}$ , and the blue dashed line represents  $P = 1 \times 10^{-5}$ .

**Table 1 | Association results for four SNPs in the GWAS, replication and combined samples.**

SNP	Chr.	Genes	MA	Stage	MAF		P value	OR (95% CI)	P <sub>het</sub>
					Cases	Control			
rs678741	10q24.32	LBX1/AS1	A	GWAS	0.532	0.444	$1.79 \times 10^{-9}$	1.42 (1.27-1.60)	0.95
				Replication 1	0.546	0.455	$5.96 \times 10^{-14}$	1.44 (1.32-1.58)	
				Replication 2	0.549	0.452	$3.83 \times 10^{-12}$	1.48 (1.32-1.66)	
				Replication 3	0.538	0.453	$1.23 \times 10^{-4}$	1.41 (1.18-1.68)	
				Combined	0.543	0.451	$9.68 \times 10^{-37}$	1.44 (1.37-1.52)	
rs4940576	18q21.33	BCL-2	T	GWAS	0.471	0.393	$7.01 \times 10^{-8}$	1.38 (1.23-1.54)	0.13
				Replication 1	0.419	0.384	$2.06 \times 10^{-3}$	1.16 (1.06-1.27)	
				Replication 2	0.430	0.386	$1.68 \times 10^{-3}$	1.20 (1.07-1.35)	
				Replication 3	0.431	0.382	$2.37 \times 10^{-2}$	1.23 (1.03-1.46)	
				Combined	0.435	0.387	$2.22 \times 10^{-12}$	1.23 (1.14-1.29)	
rs13398147	2q36.1	PAX3/EPHA4	T	GWAS	0.267	0.198	$1.47 \times 10^{-8}$	1.48 (1.29-1.70)	0.11
				Replication 1	0.233	0.201	$9.59 \times 10^{-4}$	1.20 (1.08-1.34)	
				Replication 2	0.239	0.202	$1.39 \times 10^{-3}$	1.24 (1.09-1.42)	
				Replication 3	0.241	0.202	$3.38 \times 10^{-2}$	1.25 (1.02-1.54)	
				Combined	0.243	0.201	$7.59 \times 10^{-13}$	1.28 (1.20-1.37)	
rs241215	1p36.32	AJAP1	A	GWAS	0.261	0.321	$7.55 \times 10^{-6}$	0.75 (0.65-0.85)	0.32
				Replication 1	0.299	0.329	$5.49 \times 10^{-3}$	0.87 (0.79-0.96)	
				Replication 2	0.291	0.323	$7.09 \times 10^{-3}$	0.85 (0.75-0.97)	
				Replication 3	0.284	0.327	$3.71 \times 10^{-2}$	0.82 (0.68-0.99)	
				Combined	0.287	0.325	$2.95 \times 10^{-9}$	0.83 (0.78-0.89)	

Chr., chromosome; MA, minor allele; MAF, minor allele frequency; OR, odds ratio for the minor allele; P<sub>het</sub>, P value from the heterogeneity test based on GWAS and the two replication study groups.



**Figure 2 | Regional association plots of four susceptibility loci for AIS.** Each plot shows  $-\log_{10}(P \text{ value})$  for SNPs in the specific region. Imputation was performed by using 1000 Genomes Project CHB (Han Chinese in Beijing, China) data (November 2010 release) as a reference. The genotyped SNP with the strongest association signal in each locus is indicated by a purple diamond, and the other SNPs are coloured according to the  $r^2$  values with the proxy SNP. The genes within the regions of interest are annotated with the direction of transcription represented by arrows. (a) 10q24.32, (b) 18q21.33, (c) 2q36.1 and (d) 1p36.32.

of *BCL-2* (Table 1, Fig. 2b and Supplementary Fig. 4). SNP rs13398147 ( $P_{\text{combined}} = 7.59 \times 10^{-13}$ , OR = 1.28, 95% CI: 1.20–1.37,  $P_{\text{het}} = 0.11$ ) was located 323 kb upstream of *EPHA4* and 304 kb downstream of *PAX3* (Table 1, Fig. 2c and Supplementary Fig. 5). SNP rs241215 ( $P_{\text{combined}} = 2.95 \times 10^{-9}$ , OR = 0.83, 95% CI: 0.78–0.89,  $P_{\text{het}} = 0.32$ ) maps 111 kb upstream of *AJAPI* (Table 1, Fig. 2d and Supplementary Fig. 6).

**Functional annotation.** We evaluated the overlap of the novel associated SNPs with the Encyclopedia of DNA Elements (ENCODE)-annotated genomic elements<sup>19</sup>. Some signs of regulatory activity, including histone modifications at promoter or enhancer, DNase hypersensitivity sites, binding proteins and motifs changed, were observed for the four associated SNPs and their surrogates (Supplementary Tables 5–8).

## Discussion

In this four-stage GWAS, we identified four SNPs that are significantly associated with risk of AIS in the Chinese population. The linkage disequilibrium block of the top signal SNP (rs678741) at 10q24.32 contains a widely validated risk variant of AIS (rs11190870) mapping to the 3' region of *LBX1* (refs 16,20–22). Highly expressed in spinal cord and skeletal muscle, *LBX1* was identified as an AIS susceptibility gene by a previous GWAS<sup>16</sup>. *LBX1* may be implicated in the aetiology of scoliosis through abnormal somatosensory function<sup>16</sup>. In the current study, we further pinpointed a potentially regulatory SNP in this region. Data from ENCODE showed that rs678741 may be located in a strong enhancer region marked by peaks of several active histone methylation modifications. In addition, a search of HaploReg indicated that rs678741 maps to a DNase hypersensitivity site in multiple cell lines (Supplementary Table 5)<sup>23</sup>. These findings indicated that rs678741 has the potential to regulate the transcriptional activity of *LBX1AS1*. Previous studies have shown that antisense long non-coding RNA can regulate the expression of target genes by directly binding to the transcript<sup>24,25</sup>. Therefore, it is possible that *LBX1AS1* can be a biologically functional element that contributes to the risk of AIS. Further studies are necessary to clarify the role of *LBX1AS1* in the pathogenesis of AIS.

SNP rs4940576 is located in the intron of *BCL-2* gene that encodes an integral outer mitochondrial membrane protein playing a key role in apoptosis<sup>26,27</sup>. Amling *et al.*<sup>28</sup> observed that *BCL-2* is expressed at highest levels in late proliferative and maturing chondrocytes. In *BCL-2* knockout mice, endochondral ossification was accelerated owing to the premature loss of terminal hypertrophic chondrocytes in the growth plate<sup>29</sup>. Thus, the product of *BCL-2* appears to have an important role in controlling osteoblast activity and regulating the endochondral-ossification process. Interestingly, overgrowth of anterior spinal endplate is a classic hypothesis in the etiopathogenesis of AIS<sup>30</sup>. Proliferative and hypertrophic chondrocytes in the anterior column of AIS patients seemed more active than that of the posterior column<sup>31</sup>. Moreover, AIS patients were found to have different growth kinetics in bilateral growth plate of the vertebrae as indicated by differences in cellular activity between the convex and concave side of the curve<sup>32</sup>.

The third SNP rs13398147 is located between *EPHA4* and *PAX3*. *EPHA4* belongs to the EPH receptor subfamily of the protein–tyrosine kinase family<sup>33</sup>. EPH receptors are implicated in mediating developmental events, particularly in the nervous system<sup>34</sup>. The *PAX3* subfamily regulates both myogenesis and neurogenesis in the neural tube<sup>35–37</sup>. The neural tube/notochord complex has a critical function during the development of vertebral muscle<sup>38</sup>. In addition, *PAX3* mutation can lead to

muscular and neural tube defects, as well as malformation of the vertebral column<sup>39,40</sup>. Abnormality of the paravertebral muscles has been proposed as the cause of AIS<sup>30</sup>. Many diseases that affect muscle function can present a secondary scoliosis<sup>41,42</sup>. The fourth SNP rs241215 maps near *AJAPI*, encoding a transmembrane protein that interacts with E-cadherin/ $\beta$ -catenin complexes and the adaptor protein complex AP-1B in polarized epithelial cells<sup>43</sup>. Recent findings suggested a potential role for *AJAPI* in cell–cell and cell–extracellular matrix interactions that could be involved in cell adhesion, migration and invasion<sup>44,45</sup>. Regulation of cell adhesion has been found to be essential in the bone growth and osteoblast differentiation<sup>46–48</sup>.

To determine if there was any association between the novel associated SNPs and curve severity, we investigated a subgroup of 632 patients with a mean Cobb angle of  $37.2 \pm 9.4^\circ$  (range  $27–49^\circ$ ). Of these, 214 patients received fusion surgery, and all other 418 who did not require surgical correction had been observed until skeletal maturity. As shown in Supplementary Table 9, rs241215 might be associated with the curve severity. Patients with genotype AA had a higher curve magnitude ( $39.6 \pm 7.3^\circ$ ) than those with genotype TA ( $36.9 \pm 7.7^\circ$ ) or TT ( $36.7 \pm 8.1^\circ$ ). As for the other three SNPs, no difference regarding the curve severity was found among the genotypes.

The primary limitation of our study lies in the lack of experiments supporting the functional role of the newly identified susceptibility genes. We believe that future functional analysis concerning the expression of these genes in tissues of AIS patients and related regulatory pathways can shed light on their relationships with the development of AIS. Moreover, future fine-mapping studies based on data from the resequencing of these regions may provide a clear understanding of the association at these loci.

In summary, our GWAS conducted in a Chinese Han population identifies three new susceptibility loci of AIS at 1p36.32, 2q36.1 and 18q21.33. This study also refines a region previously associated with AIS at 10q24.32, suggesting that *LBX1AS1*, an antisense transcript of *LBX1*, might be the functional variant of AIS. These findings provide additional insights into the genetic architecture of AIS. Further dissection of these loci is likely to expand our understanding of the aetiology of AIS.

## Methods

**Subjects.** The current case-control analysis was composed of four stages, including the initial genome-wide discovery stage followed by three replication stages. The study has been approved by the following local Institutional Review Boards (IRB): IRB of Nanjing Drum Tower Hospital, IRB of The Chinese University of Hong Kong, IRB of The West China Hospital, IRB of The Second Affiliated Hospital of Zhejiang University and IRB of China-Japan Union Hospital of Jilin University. Patients who came to our joint scoliosis research clinics between 2000 and 2015 were reviewed for the eligibility to be included in this study. The inclusion criteria were as follows: (1) female; (2) diagnosed as AIS through clinical and radiologic examinations; (3) with Cobb angle  $> 20^\circ$ ; and (4) having right major thoracic curve. Patients with scoliosis secondary to known aetiology, such as congenital scoliosis, neuromuscular scoliosis, scoliosis secondary to skeletal dysplasia and connective tissue abnormalities, were excluded from the study. Healthy subjects who came to our hospital for routine physical examinations were recruited as controls. Overall, there were 4,317 patients and 6,016 controls included in our study, all from ethnic Han Chinese population and collected in Nanjing of Jiangsu Province, Hangzhou of Zhejiang Province, Chengdu of Sichuan province, Changchun of Jilin Province and Hong Kong. The detailed profile of the subjects was as follows: in the discovery stage, 728 cases and 1,500 controls were collected in Nanjing, and the other 273 cases were collected in Hong Kong; in replication stage 1, all the 1,916 cases and 2,001 controls were collected in Nanjing; in replication stage 2, 307 cases and 1,735 controls were collected in Nanjing, and 600 cases and 180 controls were collected in Hong Kong, 15 cases were collected in Hangzhou, 10 cases were collected in Chengdu and 18 cases were collected in Changchun; in replications stage 3, all the 450 cases and 600 controls were collected in Nanjing. Under the approval of each participating centre's Ethical Committee, informed



consent was obtained from all the participants and from the guardians of the adolescents. Clinical characteristics of the participants in the discovery stage were summarized in Supplementary Table 1.

**Genotyping and quality control for GWAS.** Genomic DNA were extracted from peripheral blood leukocytes using a commercial DNA extraction kit (QIAGEN). Genotype analysis in the discovery stage was conducted using the Affymetrix Genome-Wide Human SNP Array 6.0 according to the manufacturer's protocol. Systemic quality control on the raw genotyping data in the discovery stage was carried out using PLINK (v1.90). The exclusion of unqualified SNPs was performed on the basis of following criteria: (1) with call rates <95%; (2) with minor allele frequency <0.05; (3) showing deviation from Hardy–Weinberg equilibrium ( $P < 1 \times 10^{-5}$ ); and (4) mapping on non-autosomal Chromosomes. For sample quality control, cryptic relatedness for each sample was evaluated with an identity-by-state method, and subjects showing second-degree relatedness or closer were removed ( $n = 35$ ). In addition, seven individuals with more than 10% missing genotypes were removed. There was no individual showing gender discrepancies based on their X chromosome genotypes. To detect population outliers and stratification, we used PCA implemented in the software package EIGENSTRAT<sup>49</sup>. A total of 2,459 participants remaining after sample quality control were analysed together with 210 HapMap subjects (60 YRI, 60 CEU, 45 JPT and 45 CHB individuals). Using the top two associated principal components, we identified that all individuals were of Chinese ancestry. Subsequently, we performed PCA using only the genotype information of the case and control subjects, and the scatterplot indicated that all cases and controls were genetically matched with minimal evidence of population stratification (Supplementary Fig. 1). After the quality control process, a total of 960 cases and 1,499 controls within Chinese Han population and 516,220 SNPs were used in the final analysis.

**Selection of SNPs for the replication stage.** SNPs surpassing a threshold of suggestive genome-wide significance ( $P < 1 \times 10^{-5}$ ) in the discovery stage were potential candidates for replication. Among these SNPs, those with ambiguous genotype scatter plots were excluded from the replication stage. A total of 43 SNPs remained after filtering, from which we chose 16 representative SNPs for replication stage 1 and removed the other 27 SNPs due to high linkage disequilibrium with at least 1 of the 16 representative SNPs (Supplementary Table 2 and 3). After replication stage 1, 4 SNPs were included in replication stage 2 and 3, including rs678741, rs241215, rs13398147 and rs4940576 (Table 1 and Fig. 1). The SNP genotyping in replication stage 1 was performed with Sequenom MassARRAY system (Sequenom Inc.). Sixteen SNPs were multiplexed in one assay, and the primers were designed by custom software to give unique mass ranges for each SNP. Samples of replication stage 2 and 3 were genotyped using TaqMan SNP Genotyping Assay, which was read with an ABI Step-One-Plus sequence detection system (Applied Biosystems, Foster City, CA). Approximately 20 ng of genomic DNA was used to genotype each sample. Several quality control measures were applied as listed below. Laboratory technicians performing the genotyping assay were kept blind of case or control samples. In all, 5% of the samples were randomly selected as blind duplicates, with a reproducibility rate of 100%. DNA samples with >10% missing genotyping were excluded.

**Statistical analysis.** In the discovery stage, a Cochran–Armitage trend test was used to calculate the association of each SNP with the disease. We calculated OR values and 95% CIs from a  $2 \times 2$  allele frequency table. Data from the discovery stage and three replication stages were combined using the Mantel–Haenszel method. The heterogeneity between studies was examined using the Q-statistic  $P$  value<sup>50</sup>. The inflation factor  $\lambda$  was calculated by dividing the mean of the lower 90% of the test statistics by the mean of the lower 90% of the expected values from a  $\chi^2$  distribution with 1 degree of freedom. We used PLINK 1.90 for general statistical analysis<sup>51</sup>. The Manhattan plot of  $-\log_{10} P$  was generated using Haploview (v4.2)<sup>52</sup>. A quantile–quantile plot generated with R (v2.6) was used to evaluate the overall significance of the GWAS results and the potential impact of population stratification.

**Imputation.** We performed genotype imputation within 800 kb around the four significant regions using MaCH-Admix software<sup>53</sup>. We used the linkage disequilibrium and haplotype information from the 1000 Genomes Project (phased CHB and CHS data; March 2012) as the reference set to impute ungenotyped markers (Supplementary Table 4). SNPs with minor allele frequency  $\leq 0.05$  or with a low imputation quality score ( $R^2 \leq 0.30$ ) were excluded to ensure good imputation quality. Logistic regression analysis was used to determine the association of the SNPs with the AIS risk. The linkage disequilibrium pattern around the risk-associated SNPs was analysed to identify susceptibility genes that were covered by the linkage disequilibrium blocks. Linkage disequilibrium structures and haplotype block plots were generated with Haploview (v4.2)<sup>52</sup>. The four significant regions were plotted using the online tool LocusZoom<sup>54</sup>.

**Functional annotation.** To explore the regulatory properties of the association signals, we used chromatin state segmentation in LCL data generated by the

ENCODE project<sup>19</sup>. HaploReg was used to examine whether any of the SNPs or their proxies (that is,  $r^2 > 0.8$  in the 1000 Genomes CHB reference panel) annotate putative transcription factor binding or enhancer elements (Supplementary Table 5–8)<sup>23</sup>.

**Subgroup analysis.** To investigate the relationship between the novel associated SNPs and curve severity, two types of patients were included in the subgroup analysis: (1) being consistently observed until skeletal maturity; and (2) undergoing fusion surgery due to curve progression. Skeletal maturity was determined based on chronological age ( $\geq 16$ -years-old) or Risser sign ( $> 4$ ). The curve severity was recorded at the latest visit for patients being consistently observed, or at the last preoperative visit for those undergoing surgery. One-way analysis of variance test was used to compare the curve severity among different genotypes. Statistical significance was set at a  $P$  value of  $< 0.05$ .

## References

- Miller, N. H. Cause and natural history of adolescent idiopathic scoliosis. *Orthop. Clin. North Am.* **30**, 343–352 (1999).
- Weinstein, S. L. Adolescent idiopathic scoliosis: prevalence and natural history. *Instr. Course Lect.* **38**, 115–128 (1989).
- Wise, C. A., Gao, X., Shoemaker, S., Gordon, D. & Herring, J. A. Understanding genetic factors in idiopathic scoliosis, a complex disease of childhood. *Curr. Genomics* **9**, 51–59 (2008).
- Riseborough, E. J. & Wynne-Davies, R. A genetic survey of idiopathic scoliosis in Boston, Massachusetts. *J. Bone Joint Surg. Am.* **55**, 974–982 (1973).
- Cheng, J. C., Tang, N. L., Yeung, H. Y. & Miller, N. Genetic association of complex traits: using idiopathic scoliosis as an example. *Clin. Orthop. Relat. Res.* **462**, 38–44 (2007).
- Inoue, M. *et al.* Association between estrogen receptor gene polymorphisms and curve severity of idiopathic scoliosis. *Spine (Phila Pa 1976)* **27**, 2357–2362 (2002).
- Chen, Z. *et al.* Promoter polymorphism of matrilin-1 gene predisposes to adolescent idiopathic scoliosis in a Chinese population. *Eur. J. Hum. Genet.* **17**, 525–532 (2009).
- Qiu, X. S. *et al.* The role of melatonin receptor 1B gene (MTNR1B) in adolescent idiopathic scoliosis—a genetic association study. *Stud. Health Technol. Inform.* **123**, 3–8 (2006).
- Ryzhkov, I. I. *et al.* Transforming growth factor beta 1 is a novel susceptibility gene for adolescent idiopathic scoliosis. *Spine (Phila Pa 1976)* **38**, E699–E704 (2013).
- Bae, J. W., Cho, C. H., Min, W. K. & Kim, U. K. Associations between matrilin-1 gene polymorphisms and adolescent idiopathic scoliosis curve patterns in a Korean population. *Mol. Biol. Rep.* **39**, 5561–5567 (2012).
- Takahashi, Y. *et al.* Lack of association between adolescent idiopathic scoliosis and previously reported single nucleotide polymorphisms in MATN1, MTNR1B, TPH1, and IGF1 in a Japanese population. *J. Orthop. Res.* **29**, 1055–1058 (2011).
- Tang, N. L. *et al.* A relook into the association of the estrogen receptor [alpha] gene (PvuII, XbaI) and adolescent idiopathic scoliosis: a study of 540 Chinese cases. *Spine (Phila Pa 1976)* **31**, 2463–2468 (2006).
- Takahashi, Y. *et al.* Replication study of the association between adolescent idiopathic scoliosis and two estrogen receptor genes. *J. Orthop. Res.* **29**, 834–837 (2011).
- Sharma, S. *et al.* Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. *Hum. Mol. Genet.* **20**, 1456–1466 (2011).
- Qiu, X. S. *et al.* Lack of association between the CHL1 gene and adolescent idiopathic scoliosis susceptibility in Han Chinese: a case-control study. *BMC Musculoskelet Disord.* **15**, 38 (2014).
- Takahashi, Y. *et al.* A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. *Nat. Genet.* **43**, 1237–1240 (2011).
- Kou, I. *et al.* Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nat. Genet.* **45**, 676–679 (2013).
- Sharma, S. *et al.* A PAX1 enhancer locus is associated with susceptibility to idiopathic scoliosis in females. *Nat. Commun.* **6**, 6452 (2015).
- Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
- Jiang, H. *et al.* Association of rs11190870 near LBX1 with adolescent idiopathic scoliosis susceptibility in a Han Chinese population. *Eur. Spine J.* **22**, 282–286 (2013).
- Gao, W. *et al.* Association between common variants near LBX1 and adolescent idiopathic scoliosis replicated in the Chinese Han population. *PLoS One* **8**, e53234 (2013).
- Fan, Y. H. *et al.* SNP rs11190870 near LBX1 is associated with adolescent idiopathic scoliosis in southern Chinese. *J. Hum. Genet.* **57**, 244–246 (2012).

23. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
24. Kerin, T. *et al.* A noncoding RNA antisense to moesin at 5p14.1 in autism. *Sci. Transl. Med.* **4**, 128ra40 (2012).
25. Mizuta, I. *et al.* YY1 binds to alpha-synuclein 3'-flanking region SNP and stimulates antisense noncoding RNA expression. *J. Hum. Genet.* **58**, 711–719 (2013).
26. El-Emshaty, H. M., Saad, E. A., Toson, E. A., Abdel, M. C. & Gadelhak, N. A. Apoptosis and cell proliferation: correlation with BCL-2 and P53 oncoprotein expression in human hepatocellular carcinoma. *Hepatogastroenterology* **61**, 1393–1401 (2014).
27. Siddiqui, W. A., Ahad, A. & Ahsan, H. The mystery of BCL2 family: Bcl-2 proteins and apoptosis: an update. *Arch. Toxicol.* **89**, 289–317 (2015).
28. Amling, M., Posl, M., Hentz, M. W., Priemel, M. & Dellling, G. PTHrP and Bcl-2: essential regulatory molecules in chondrocyte differentiation and chondrogenic tumors. *Verh Dtsch Ges Pathol.* **82**, 160–169 (1998).
29. Boot-Handford, R. P. *et al.* The bcl-2 knockout mouse exhibits marked changes in osteoblast phenotype and collagen deposition in bone as well as a mild growth plate phenotype. *Int. J. Exp. Pathol.* **79**, 329–335 (1998).
30. Kouwenhoven, J. W. & Castelein, R. M. The pathogenesis of adolescent idiopathic scoliosis: review of the literature. *Spine (Phila Pa 1976)* **33**, 2898–2908 (2008).
31. Zhu, F., Qiu, Y., Yeung, H. Y., Lee, K. M. & Cheng, J. C. Histomorphometric study of the spinal growth plates in idiopathic scoliosis and congenital scoliosis. *Pediatr. Int.* **48**, 591–598 (2006).
32. Wang, S. *et al.* Histomorphological study of the spinal growth plates from the convex side and the concave side in adolescent idiopathic scoliosis. *J. Orthop. Surg. Res.* **2**, 19 (2007).
33. Farenc, C., Celie, P. H., Tensen, C. P., de Esch, I. J. & Siegal, G. Crystal structure of the EphA4 protein tyrosine kinase domain in the apo- and dasatinib-bound state. *FEBS Lett.* **585**, 3593–3599 (2011).
34. Frisen, J. & Barbacid, M. Genetic analysis of the role of Eph receptors in the development of the mammalian nervous system. *Cell Tissue Res.* **290**, 209–215 (1997).
35. Schubert, F. R. *et al.* Early mesodermal phenotypes in splotch suggest a role for Pax3 in the formation of epithelial somites. *Dev. Dyn.* **222**, 506–521 (2001).
36. Buckingham, M. & Relaix, F. The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu. Rev. Cell Dev. Biol.* **23**, 645–673 (2007).
37. Young, A. P. & Wagers, A. J. Pax3 induces differentiation of juvenile skeletal muscle stem cells without transcriptional upregulation of canonical myogenic regulatory factors. *J. Cell Sci.* **123**, 2632–2639 (2010).
38. Rong, P. M., Teillet, M. A., Ziller, C. & Le Douarin, N. M. The neural tube/notochord complex is necessary for vertebral but not limb and body wall striated muscle differentiation. *Development* **115**, 657–672 (1992).
39. Xiao, Y. *et al.* Characterization of a novel missense mutation on murine Pax3 through ENU mutagenesis. *J. Genet. Genomics* **38**, 333–339 (2011).
40. Farin, H. F., Mansouri, A., Petry, M. & Kispert, A. T-box protein Tbx18 interacts with the paired box protein Pax3 in the development of the paraxial mesoderm. *J. Biol. Chem.* **283**, 25372–25380 (2008).
41. Hsu, J. D. The natural history of spine curvature progression in the nonambulatory Duchenne muscular dystrophy patient. *Spine (Phila Pa 1976)* **8**, 771–775 (1983).
42. Allam, A. M. & Schwabe, A. L. Neuromuscular scoliosis. *PM&R* **5**, 957–963 (2013).
43. Bharti, S. *et al.* Novel membrane protein shrew-1 targets to cadherin-mediated junctions in polarized epithelial cells. *Mol. Biol. Cell* **15**, 397–406 (2004).
44. McDonald, J. M. *et al.* The SHREW1 gene, frequently deleted in oligodendrogliomas, functions to inhibit cell adhesion and migration. *Cancer Biol. Ther.* **5**, 300–304 (2006).
45. Lin, N. *et al.* Deletion or epigenetic silencing of AJAP1 on 1p36 in glioblastoma. *Mol. Cancer Res.* **10**, 208–217 (2012).
46. Sosa-Garcia, B. *et al.* A role for the retinoblastoma protein as a regulator of mouse osteoblast cell adhesion: implications for osteogenesis and osteosarcoma formation. *PLoS ONE* **5**, e13954 (2010).
47. Hay, E. *et al.* N- and E-cadherin mediate early human calvaria osteoblast differentiation promoted by bone morphogenetic protein-2. *J. Cell Physiol.* **183**, 117–128 (2000).
48. Li, H. *et al.* Role of neural-cadherin in early osteoblastic differentiation of human bone marrow stromal cells cocultured with human umbilical vein endothelial cells. *Am. J. Physiol. Cell Physiol.* **299**, C422–C430 (2010).
49. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
50. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl Cancer Inst.* **22**, 719–748 (1959).
51. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
52. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
53. Liu, E. Y., Li, M., Wang, W. & Li, Y. MaCH-admix: genotype imputation for admixed populations. *Genet. Epidemiol.* **37**, 25–37 (2013).
54. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).

### Acknowledgements

We thank all the research staff and students who participated in this research and devoted their time and effort. We sincerely thank Dr Gorman Kristen of CHU Saint Justine Research Center (Montreal, Canada) for advice on language and style of the manuscript. We especially thank Linlin Zhang, Le Zhang, Jian Wang and Lili Sha for their great help in solving the administration issues. Besides, we thank Wei Liu for his coordination of the sample transportation. This work was funded by the Science and Technology Project of Jiangsu Province's Clinical Medicine (BL2012002), Jiangsu Province's Key Medical Center (ZX201107) and the Natural Science Foundation of China (81171767, and 81171672). J.C.-Y.C. and N.L.-S.T. received grant support from Fondation Yves Cotrel, Institut de France.

### Author contributions

Z.Z., N.L.-S.T., Y.Q., and J.C.-Y.C. directed the study and obtained financial support. L.X., and J.C. performed the statistical analysis. L.X. and N.L.-S.T. processed the data and drafted the manuscript. B.-p.Q., S.M., J.Q., B.S., F.W., X.S., Z.L. and F.Z. were responsible for the collection of blood samples in Nanjing. Z.W., D.X., X.Q., Z.C., L.C., Z.H., M.J., X.H., Z.Z., and L.J. were responsible for the extraction of DNA of Nanjing samples. N.L.-S.T., K.M.L., B.K.-W.N., W.Y.W.L. and T.P.L. were responsible for subject recruitment and DNA preparation for the Hong Kong samples. Y.S., L.L., P.L., and F.L. were responsible for subject recruitment and DNA preparation in Chengdu, Changchun and Hangzhou, respectively. All authors approved the final version of the manuscript.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Zhu, Z. *et al.* Genome-wide association study identifies new susceptibility loci for adolescent idiopathic scoliosis in Chinese girls. *Nat. Commun.* **6**:8355 doi: 10.1038/ncomms9355 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>