Title

Genome-wide meta-analysis and replication studies in multiple ethnicities identify novel adolescent idiopathic scoliosis susceptibility loci.

Journal

Human Molecular Genetics, 27(22)

ISSN

0964-6906

Authors

Khanshour, Anas M
Kou, Ikuyo
Fan, Yanhui
et al.

Publication Date

2018-11-15

DOI

10.1093/hmg/ddy306

Peer reviewed

Human Molecular Genetics, 2018, Vol. 27, No. 22 3986–3998

doi: 10.1093/hmg/ddy306
Advance Access Publication Date: 24 August 2018
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Genome-wide meta-analysis and replication studies in multiple ethnicities identify novel adolescent idiopathic scoliosis susceptibility loci

Anas M. Khanshour[1], Ikuyo Kou[2], Yanhui Fan[3], Elisabet Einarsdottir[4,5,6], Nadja Makki[7,8], Yared H. Kidane[1], Juha Kere[4,5,9,10], Anna Grauers[10,11], Todd A. Johnson[2], Nandina Paria[1], Chandreshkumar Patel[12], Richa Singhania[1,†], Nobuhiro Kamiya[13], Kazuki Takeda[2,14], Nao Otomo[2,14], Kota Watanabe[14], Keith D. K. Luk[15], Kenneth M.C. Cheung[15], John A. Herring[1,16], Jonathan J. Rios[1,12,16], Nadav Ahituv[7,8], Paul Gerdhem[17,10], Christina A. Gurnett[18], You-Qiang Song[3], Shiro Ikegawa[2,*] and Carol A. Wise[1,12,16,*]

[1]Sarah M. & Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, TX 75219, USA, [2]Laboratory of Bone & Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo, 108-8639 Japan, [3]School of Biomedical Sciences, The University of Hong Kong, Hong Kong 10000, China, [4]Folkhälsan Institute of Genetics, University of Helsinki, 00014 University of Helsinki, Finland, [5]Molecular Neurology Research Program, University of Helsinki, 00014 University of Helsinki, Finland, [6]Department of Biosciences & Nutrition, Karolinska Institutet, Huddinge, 14183, Sweden, [7]Department of Bioengineering & Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94131, USA, [8]Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94131, USA, [9]Department of Medical & Molecular Genetics, King's College London, Guy's Hospital, London SE1 9RT, UK, [10]Department of Clinical Science, Intervention & Technology (CLINTEC), Karolinska Institutet, K54 Huddinge, Stockholm 141 86, Sweden, [11]Department of Orthopedics, Sundsvall and Härnösand County Hospital, Sundsvall 856 43, Sweden, [12]McDermott Center for Human Growth & Development, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA, [13]Sports Medicine, Tenri University, Nara Prefecture 632-0032, Japan, [14]Department of Orthopaedic Surgery, Keio University School of Medicine, Tokyo, 160-8582 Japan, [15]Department of Orthopaedics & Traumatology, The University of Hong Kong, Hong Kong 10000, China, [16]Department of Orthopaedic Surgery, Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA, [17]Department of Orthopedics, Karolinska University Hospital, K54 Huddinge, Stockholm, 141 86, Sweden and [18]Department of Neurology, School of Medicine, Washington University, St. Louis, MO 63130, USA

*To whom correspondence should be addressed at: Shiro Ikegawa, Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo 108-8639, Japan. Email: sikegawa@ims.u-tokyo.ac.jp and Carlol A. Wise, 2222 Welborn St, Dallas, TX 75219, USA. Tel: +1 2145597861; Fax: +1 2145597872; Email: Carol.Wise@tsrh.org

[†]Present address: Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10065, USA.

Received: May 17, 2018. **Revised:** July 6, 2018. **Accepted:** August 20, 2018

Abstract

Adolescent idiopathic scoliosis (AIS) is the most common musculoskeletal disorder of childhood development. The genetic architecture of AIS is complex, and the great majority of risk factors are undiscovered. To identify new AIS susceptibility loci, we conducted the first genome-wide meta-analysis of AIS genome-wide association studies, including 7956 cases and 88 459 controls from 3 ancestral groups. Three novel loci that surpassed genome-wide significance were uncovered in intragenic regions of the *CDH13* (*P*-value_rs4513093 = 1.7E-15), *ABO* (*P*-value_ rs687621 = 7.3E-10) and *SOX6* (*P*-value_ rs1455114 = 2.98E-08) genes. Restricting the analysis to females improved the associations at multiple loci, most notably with variants within *CDH13* despite the reduction in sample size. Genome-wide gene-functional enrichment analysis identified significant perturbation of pathways involving cartilage and connective tissue development. Expression of both *SOX6* and *CDH13* was detected in cartilage chondrocytes and chromatin immunoprecipitation sequencing experiments in that tissue revealed multiple HeK27ac-positive peaks overlapping associated loci. Our results further define the genetic architecture of AIS and highlight the importance of vertebral cartilage development in its pathogenesis.

Introduction

Scoliosis is defined as a rotational deformity of the spine that measures >10 degrees in the coronal plane. Scoliosis can occur with other comorbidities, particularly in heritable disorders of neuromuscular or connective tissue development. However, in more than 80% of patients, the origins of scoliosis are idiopathic. This form of scoliosis affects ∼3% of children worldwide, typically coinciding with the adolescent growth spurt in individuals who are otherwise healthy and bear no obvious structural deficiencies in the vertebral column and associated soft tissues (1,2). Individuals affected with adolescent idiopathic scoliosis (AIS) risk increasing deformity until the cessation of growth at the time of skeletal maturity. Individuals affected with large curves (>50 degrees) may continue to worsen throughout adulthood, albeit at a slower rate (1,3). Progression of the deformity can be rapid, requiring comprehensive screening for early diagnosis. AIS patients with progressive disease require active treatment, such as bracing and spinal fusion surgery, modalities that are associated with psychological and physical ramifications and may be complicated by neurologic damage, infection and long-term loss of spinal motion (1). Hospital costs of operative treatment alone exceed 1 billion USD annually in the US, a number that is rising sharply due to the increasing price of rod instrumentation systems (4). The natural history of untreated progressive AIS is the potential for chest wall compromise with concomitant lung restriction, pain and spinal osteoarthritis (1).

AIS is notable for its striking sexual dimorphism, with girls having a more than 5-fold greater risk of progressive deformity than boys and nearly 10 times more likely to require operative treatment (5). Consequently, hospital-based AIS populations are mostly female. The reasons for the dichotomy in female/male disease expression are not known. One model to explain sexual dimorphism is differences in genetic loading between males and females, with the least affected sex (males) requiring a stronger genetic load to acquire disease, a so-called Carter effect. One study of a cohort of multiplex families has provided epidemiological evidence for this genetic model (6).

Genetic factors in human AIS have been recognized for almost a century. Population studies have revealed increased concordance in monozygotic versus dizygotic twins and ∼5-fold increased risk to siblings of individuals with AIS (7–9). While rare AIS families with Mendelian-like inheritance patterns exist, heritability studies support a more complex model in which multiple genetic susceptibility factors comprise the total mutational burden (10–13). Population-based genome-wide association studies (GWASs) have defined the majority of AIS susceptibility loci thus far, mostly in cohorts that are wholly or predominantly female. A study of 1033 East Asian cases and 1473 matched controls identified the first major AIS susceptibility locus on chromosome 10q24.31 near the gene encoding *LBX1* (14). Subsequent GWASs identified AIS susceptibility loci on chromosome 6q24.1 within the *GPR126* (*ADGRG6*) gene (15) and within putative enhancers of the *PAX1* (16) and *BNC2* genes (17). GWASs of severely affected AIS subjects (spinal deformity above 40 degrees) have identified two loci, one on chromosome 17q24.3 in a gene desert between the *SOX9* and *KCNJ2* genes and a second within *MIR4300HG*.

The mechanisms driving AIS susceptibility conferred by these loci are largely undefined. *LBX1* (OMIM 604255) activity is observed during embryogenesis, with expression restricted to the developing central nervous system and muscles. *Lbx1* knockout (KO) mice, which display extensive muscle loss at birth (18–20), defects in heart looping and myocardial hyperplasia (21) and aberrant neuronal populations, do not exhibit scoliosis (22,23). However, both over-expression as well as morpholino knockdown of the three lbx ohnologs in zebrafish produced scoliosis, suggesting that *LBX1* may play a dosage-sensitive role in spine deformity (24). Several candidate genes associated with human AIS participate in forming the intervertebral disc (IVD). *PAX1*, encoding the paired box 1 transcription factor, is an early marker of the developing sclerotome that is required for appropriate formation of vertebral bodies and may regulate the transformation of the notochord into the nucleus pulposus of the IVD (25). Selective removal of Gpr126 protein (also known as adhesion G-protein-coupled receptor G6) from osteochondroprogenitor cells produces mice with a late-onset idiopathic scoliosis phenotype and an apparent failure of midline fusion within the developing annulus fibrosis of the IVDs (26). Whether *SOX9* contributes to the chromosome 17q24.3 association with AIS severity is unclear, but it is an excellent candidate, given its known role in regulating *GPR126* gene expression as well as other factors that control development and homeostasis of the growth plate, articular cartilage and the IVD (27,28).

While these studies have enabled hypothesis-driven investigations of AIS pathogenesis (24,26), the majority of factors contributing genetic risk to the disease remains undefined. Experience with other complex genetic disorders and traits suggests that common variation may contribute appreciably to disease risk; however, continually increasing sample sizes and

other measures are needed to achieve the power necessary for their detection (29,30). Trans-ethnic genetic studies can leverage power from the differences in population-specific genetic architecture to identify novel common variation contributing to complex traits and diseases (31). AIS does not generally cluster within any particular geographic region, population or ethnicity (32). In addition, prior multi-ethnic meta-analyses of single AIS susceptibility loci on chromosomes 10q24.31 (33) and 9p22.2 (34) provided substantially increased power over individual results. We therefore sought to identify and validate novel AIS susceptibility loci by performing the first genome-wide trans-ethnic meta-analysis applied to existing imputation-driven GWASs, an effort we organized through The International Consortium for Spine Genetics, Development, and Disease (ICSGDD) (35) (formerly International Consortium for Scoliosis Genetics). For the first time, we also analyzed the combined effects of all GWASs loci to identify pathways that are significantly perturbed in AIS. Finally, we performed RNA expression analyses and chromatin immunoprecipitation sequencing (ChIPseq) experiments, revealing potential regulatory elements overlapping with AIS-associated loci. Our results highlight pathways of cartilage and connective tissue development in AIS and provide new genetic insights into its pathogenesis.

Results

Meta-analysis of GWASs

Through ICSGDD we organized an international collaboration to discover new genetic risk factors for AIS. We performed a genome-wide meta-analysis that combined data for 3 493 832 single-nucleotide polymorphisms (SNPs) shared among six discovery GWASs from Japan [JP1 (17)], Hong Kong (HK) and the USA (Texas-GWAS1 (36), Texas-GWAS2 (37), Texas-GWAS3 and Missouri-MO1) of a total of 33 476 samples (Supplementary Material, Table S1). All studies were approved by the relevant ethics committees and all subjects were collected with informed consent, as detailed in the Methods section. Cohorts and genotyping platforms are summarized in Supplementary Material, Table S1. After performing standard quality controls, the shared 3 493 832 SNPs (Supplementary Material, Table S2) did not have any obvious allele miscoding issues (Supplementary Material, Fig. S1). Association testing in JP1 was previously computed (17). We computed tests of association with imputed allele dosages for 5 data sets (Texas-GWAS1, Texas-GWAS2, Texas-GWAS3, MO1 and HK) independently using logistic regression with gender and 10 additional principal components as covariates (38). To perform the meta-analysis, summary statistics across all six discovery studies were then combined using the inverse-variance-based method as implemented in METAL (39) (Fig. 1 and Supplementary Material, Fig. S2). From these results we identified 18 loci with $P_{META} < 1.0E-5$, including 4 within or near the *LBX1* [MIM:604255], *GPR126*[MIM: 612243], *PAX1*[MIM: 16741] and *BNC2*[MIM:608669] genes that were previously associated with AIS in multiple studies (14,15,17,37) (Fig. 2A–D and Supplementary Material, Table S3). We sought to validate the association of the novel loci with AIS by genotyping them in 4871 additional cases and 58 871 controls from 2 independent cohorts [JP2 and Swedish–Danish (40,41)] (Supplementary Material, Table S3). Combining the results of the discovery and validation studies provided increased evidence for association for 4 of the 12 novel loci that we successfully tested (Fig. 2 and Supplementary Material, Table S3). Of these four loci, three surpassed a genome-wide significance level of 5.0E-08 and occurred in intronic regions of the *CDH13* [MIM:601364], *SOX6* [MIM:607257] and *ABO* [MIM: 110300] genes. The 16q23.3-associated region also encodes two uncharacterized transcripts, *LOC101928446* and *LOC101928417*, that flank *CDH13* exon 2 (Supplementary Material, Fig. S4A). Other loci that did not surpass genome-wide significance in this meta-analysis were not included in the heritability study, yet may also contribute to the genetic architecture of AIS (Supplementary Material, Table S3).

The risk of progressive deformity is more than 5 times greater in females than in males, and girls are about 10 times more likely to require surgical correction of their deformity (42). Males with AIS may be expected to carry more penetrant and rare AIS risk alleles, but our study was underpowered to study that group independently (6). Instead, we removed male genotypes (9.3% of the total cases in our discovery study) and performed a separate female-specific meta-analysis. SNPs at the *LBX1*, *GPR126* and *PAX1* genes remained the most statistically significant. We noted that the novel association within the *CDH13* gene improved by an order of magnitude ($P = 2.29E-09$; Supplementary Material, Table S4) and surpassed *BNC2* in the female-only analysis despite the reduction in sample size. These results suggest that the risk of AIS conveyed by *CDH13* may be particularly relevant in females. In contrast, the evidence for the novel association with the *ABO* and *SOX6* loci (rs687621, rs1455114) decreased considerably with removal of male subjects. Fifteen additional loci that improved in the female-only analysis and provided suggestive evidence of association are given in Supplementary Material, Table S4.

Estimation risk proportions explained by seven AIS-associated loci

We applied multiple statistical measures to estimate the population attributable risk of disease conferred by the seven AIS-associated loci that surpassed genome-wide significance. The four measures tested gave fairly consistent estimations of the AIS risk proportions explained by these loci (Supplementary Material, Fig. S3), ranging between 2.94 and 6.35% in the combined East Asian cohorts (JP1-2 and HK) and between 2.61 and 5.67% in the combined European ancestry cohorts (Texas-GWAS1-3, MO1 and Swedish–Danish). It is interesting to note that in these analyses the chromosome 10q24.31 locus appears to confer a slightly higher proportion of risk to East Asian versus European ancestry populations (Supplementary Material, Fig. S3).

Gene-functional enrichment analysis

In order to understand the combined effects of SNPs, we sought to discover perturbation of gene ontology (GO) cellular components, biological processes and molecular functions in our study. We obtained 17 GO terms that were statistically significant at the 0.05 false discovery rate (FDR) cutoff. There were 13 GO terms (pathways) significantly impacted in both reference populations (Fig. 4). The top three pathways include the following: cartilage development (GO:0051216, with an average FDR value of 2.47E-03), connective tissue development (GO:0061448, with an average FDR = 2.45E-03) and embryo development (GO:0009790, with an average FDR = 2.45E-03). Genes involved in the cartilage and connective tissue development pathways, and genes with $P \leq 0.05$, are shown in Supplementary Material, Table S8. Applying a less-stringent FDR cutoff as suggested by the gene set enrichment analysis (GSEA) (43) captured additional pathways

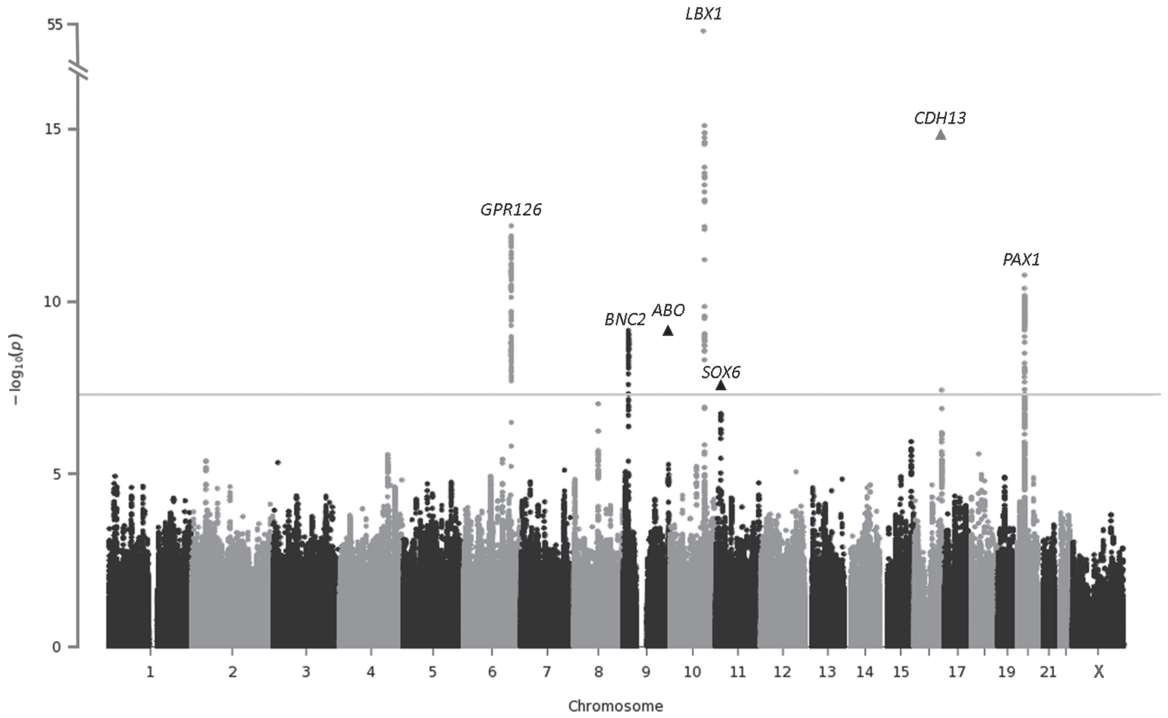

Figure 1. Manhattan plot showing the –log10 *P*-values (y axis) by physical genomic position (x axis) for each SNP genome wide. The black and gray dots represent the *P*-value from six discovery studies (JP1, Texas-GWAS1-3, MO1 and HK). The triangles represent the overall *P*-value after combining the discovery studies with the two independent replication cohorts (JP2 and Swedish–Danish). The horizontal line represents the threshold for genome-wide significance level ($P < 5\times10^{-8}$) after Bonferroni multiple testing correction. For chromosome X, association *P*-values were calculated using only females.

from the AIS meta-analysis that may participate in disease pathogenesis (Supplementary Material, Table S9).

Gene expression and ChIP-seq

The key role of SOX6 in regulating chondrogenesis and cartilage development is well described (44,45). To evaluate other AIS-associated genes in cartilage development, we performed real-time amplification of total RNA from the mouse N1511 chondrocytic cell line (46) and rib cartilage tissue and compared to the expression of *Col2a1*, *Sox5* and *Sox9* (Supplementary Material, Fig. S5). Sox6 was barely detectable in N1511 cells as expected for cells that are naive to chondrogenesis, but in the mouse rib cartilage its expression exceeded *Sox5*. The expression of *Abo* was not detected in either tested cell type. However, we found robust steady-state expression of *Cdh13* in both N1511 cells and rib cartilage, particularly in N1511 cells that are naïve to chondrogenesis. In these cells, *Cdh13* expression was strikingly similar to *Gpr126* and significantly more than each of *Col2a1* and *Sox5*. This was in contrast to its relatively lower expression in rib cartilage, as compared to *Gpr126* and *Sox9*.

To define functional elements that could drive the associations we observed within the *SOX6* and *CDH13* genes, we performed H3K27ac ChIP-seq of mouse cartilage chondrocytes to identify active chromatin regions such as enhancers and promoters (47,48). This identified multiple highly enriched peaks that mapped within *SOX6*- and *CDH13*-associated interval, defined as regions with SNPs in strong linkage disequilibrium ($R^2 >= 0.8$) with the top-associated SNP, Fig 3B. No obvious peaks were found within the *ABO*-associated locus, Fig. 3C. In addition, we also screened potentially relevant H3K27ac ChIP-seq data sets in human stem cells (H1-ESC) and skeletal muscle (HSMM) available in the Encyclopedia of DNA Elements (ENCODE). These comparisons revealed a pattern of active chromatin in human stem cells and skeletal muscle for all three regions (Fig. 3).

Discussion

An obstacle to discovering genetic associations with complex diseases, such as AIS, is insufficient statistical power to detect relatively weak effects contributed by each locus. To overcome this limitation, we organized the largest GWAS of AIS to date through an international consortium-based effort. We sought to achieve further gains in power by applying imputation-driven multi-ethnic meta-analysis methods. The results of this study identified three novel AIS associations, within *CDH13*, *SOX6* and *ABO*, that were not significant in any independent AIS studies. *CDH13* encodes the cadherin 13 or 'T-cadherin' protein. The cadherin family of transmembrane molecules classically mediates Ca^{2+}-dependent homophilic intercellular adhesion. Within this family, T-cadherin is structurally distinct as it lacks transmembrane and cytoplasmic domains and is attached to the plasma membrane via a glycosylphosphatidylinositol anchor (49). During development, T-cadherin is expressed in regions avoided by neural crest cells migrating from the neural tube, suggesting it may act as a negative guidance cue for axonal growth, and it is also highly expressed throughout the early developing sclerotome, with later expression restricted to the caudal sclerotome (50). T-cadherin is also expressed in proprioceptive sensory neurons of the dorsal root ganglion in response to spatially restricted signals from the limb mesenchyme that it innervates (51). A role for *CDH13*-mediated proprioception is plausible in AIS, given that proprioceptive mechanisms are proposed to be essential for maintaining proper spinal alignment and for potentially preventing scoliosis (52). Our data suggest that *CDH13*

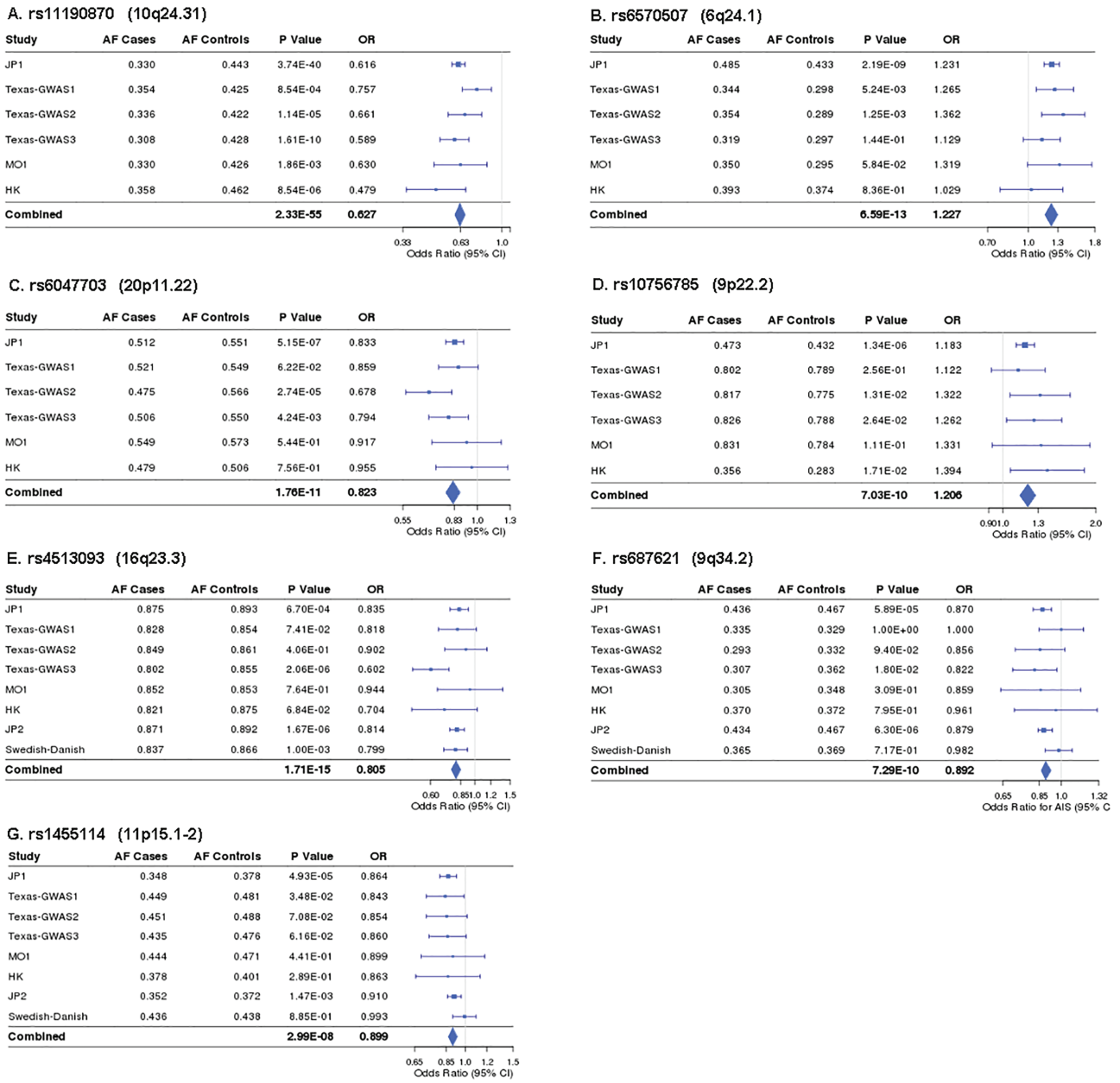

Figure 2. Forest plots of the leading SNPs in seven loci. **(A)–(D):** Loci known to be associated with AIS found in six discovery studies (JP1, Texas-GWAS1-3, MO1 and HK). **(E)–(G):** Three novel loci associated with AIS found in meta-analysis of six discovery studies (JP1, Texas-GWAS1-3, MO1 and HK) and two independent replication cohorts (JP2 and Swedish–Danish). For each study, the square indicates the odds ratio and the line indicates the 95% confidence interval. The square size is proportional to the precision of the estimate. For the combined estimate, the center of the diamond indicates the odds ratio for the meta-analysis, with the left and right corners of the diamond indicating the boundaries of the 95% confidence interval. No significant evidence for heterogeneity across all studies was found in any locus (HetPVal_rs11190870=0.148, HetPVal_rs6570507=0.547, HetPVal_rs6047703=0.285 and HetPVal_rs10756785=0.721, HetPVal_ rs4513093=0.196, HetPVal_ rs687621=0. 381, HetPVal_ rs1455114=0.477).

is expressed in mouse early chondrocytes and rib cartilage, consistent with reports of its expression in adult articular cartilage (50,53). Further tissue-specific functional analyses are required to elucidate the normal physiologic roles of *CDH13* in these tissues and its contributions to spinal development and deformity.

SOX6, encoding SRY-box 6 transcription factor, is a member of the Sex-determining Y box transcription factor (SOX) family of transcription factor proteins defined by the conserved high mobility group DNA-binding domain. *SOX6* and the closely related *SOX5* [MIM:604975] as well as *SOX9* [MIM: 608160] share largely overlapping DNA-binding sites and are expressed in the developing sclerotome and/or notochord. The SOX trio cooperatively drives chondrogenesis through transcriptional activation of super enhancers (54). KO mice lacking Sox6 display small, fragmented nucleus pulposi and *Sox6*$^{-/-}$ mice that survive to adulthood develop kinked tails, apparently due to compressions in the IVD that are particularly localized within the nucleus pulposus (45). While *SOX6* is clearly a key player in disc biogenesis and in chondrogenesis, it has also been shown to be essential for correct muscle fiber-type specification in the zebrafish. Surviving zebrafish with homozygous loss of *sox6* display disrupted

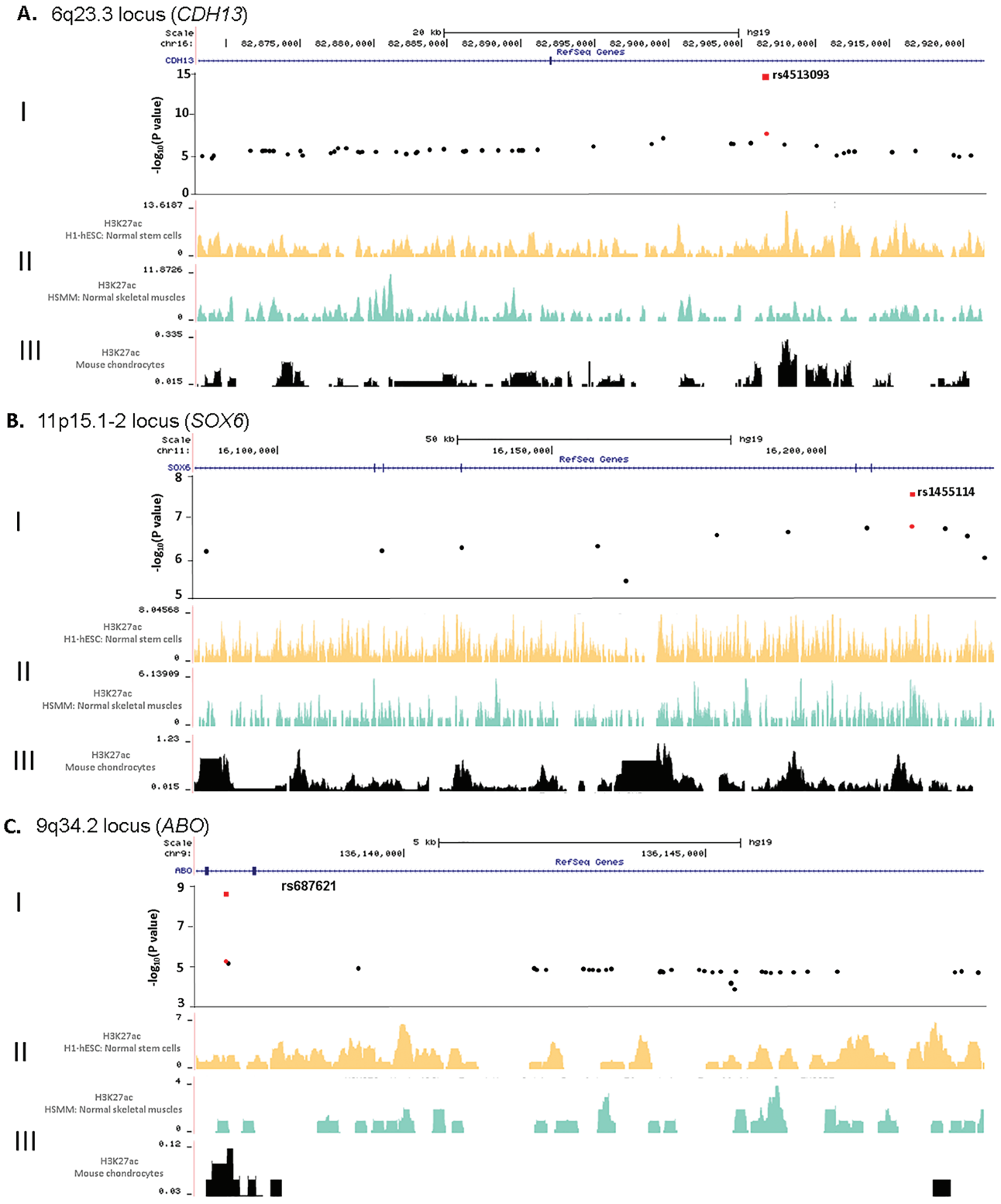

Figure 3. Musculoskeletal tissue H3K27ac ChIPseq annotations for novel AIS-associated genomic regions. (**A**) *CDH13* locus at 16q23.3. (**B**) *SOX6* locus at 11p15.1-2. (**C**) *ABO* locus at 9q34.2. **I:** Association *P*-values for the top SNPs in each region. The black dots represent SNPs from the combined six discovery studies and in high LD (R^2 >=0.8) with the lead SNP (using 1KG_ASN). The red square represents the overall *P*-value after combining the two replications with the discovery studies. **II:** The image from University of California, Santa Cruz Genome Browser at the corresponding genomic region (human genome build 19; hg19) annotated with H3K27ac in the embryonic stem cell (H1-hESC) and skeletal muscle myoblasts (HSMM) cell lines from ENCODE database. **III:** H3K27ac ChIPseq peaks from mouse chondrocytes.

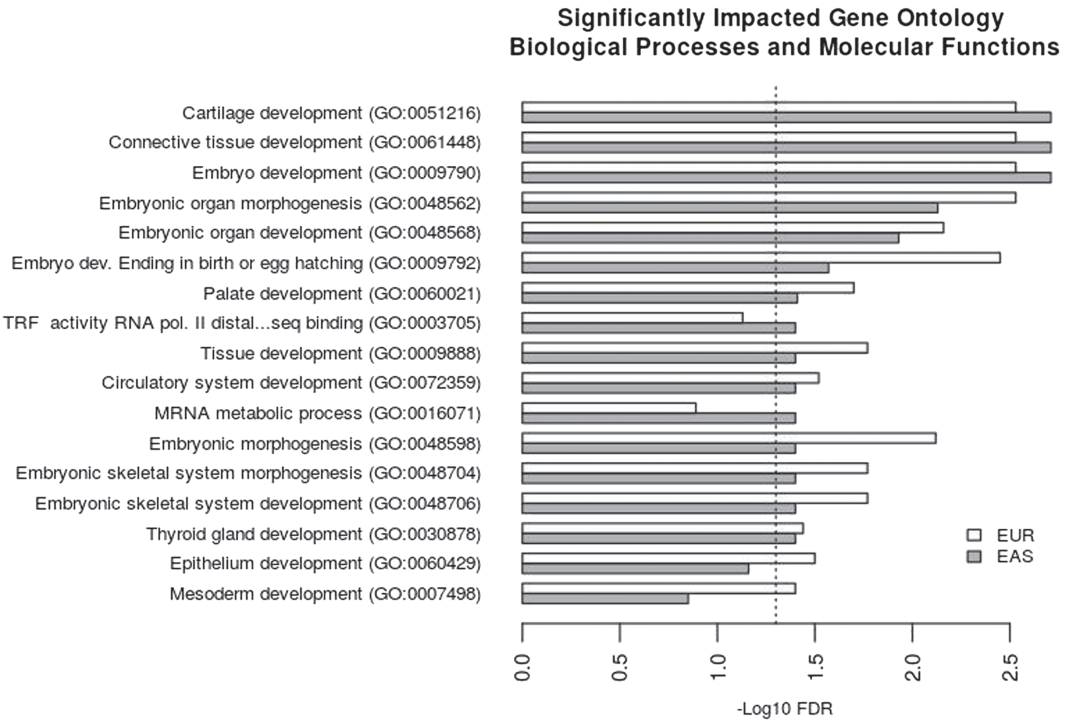

Figure 4. FDR plot for significant gene ontology terms (pathways) using EAS and EUR reference populations. Dotted vertical line indicates 0.05 (=1.3 in Log10 scale) FDR cutoff. All pathways to the right of the dotted line are statistically significant at 0.05 FDR cutoff.

muscle fiber-type identity and a striking spinal curvature that progresses with age (55). As noted, *Gpr126* has been linked to Sox-mediated mechanisms and may be a SOX target, as its RNA levels are reduced when *Sox9* is removed from mouse cartilage or chondrocytes in culture (27,56). We also hypothesize that *SOX6* affects the expression of another AIS candidate gene, *PAX1* (37). *Pax1* expression is a well-characterized marker of the developing sclerotome (25). By E15.5, this expression is normally restricted to the perichondrium and outer annulus of the disc. However, in $Sox5^{-/-}/Sox6^{-/-}$ mutants, *Pax1* is still widely expressed throughout the IVD, suggesting its spatial expression is under the control of one or both Sox factors (45). These observations suggest that the functional mechanism underlying the AIS association may be due to a reduction in *SOX6* activity.

The novel association with the 9q34.2 locus (represented by SNP rs687621) occurs within the *ABO* gene (Supplementary Material, Fig. S4C). *ABO* glycosyltransferases catalyze the transfer of carbohydrates to the H antigen, forming the antigenic structures of the *ABO* blood group [MIM:616093]. Nonsynonymous and deletion variants in *ABO* determine the A, B and O blood types. Other diseases and traits associated within *ABO* include protection against malaria, regulation of interleukin-10 levels and other inflammatory biomarkers as well as susceptibility to Graves' disease, venous thromboembolism and childhood obesity (57–61). *ABO* is expressed at varying levels in a variety of tissues (GTEx Portal, see web resources section), and its downregulation is also associated with certain cancers (62). However, we did not see *ABO* gene expression in the mouse N1511 chondrocytic cell line and rib cartilage (Supplementary Material, Fig. S5). We sought to determine whether blood type-determining SNPs could be contributing to the association with AIS. We found that these SNPs were correlated to varying degrees with rs687621 in both European and Asian populations, and accordingly, these SNPs produced some evidence for association with AIS (Tables S5, S6). Although historically, AIS has not been associated with any particular blood type, we cannot exclude the possibility that blood-determining variants functionally contribute to AIS. We hypothesize, however, that the association we observed is driven by other 9q34.2 variants in linkage disequilibrium with rs687621, possibly affecting *ABO* glycosyltransferases activity in AIS-relevant tissues.

Other loci that did not surpass genome-wide significance in this meta-analysis (Supplementary Material, Table S3) may also contribute to the genetic architecture of AIS. For example, SNP rs7822863 approached genome-wide significance ($P = 9.3E-08$) and is encoded in an 8q13.3 intergenic region between the *EYA1* [MIM: 601653] and *MSC* [MIM: 603628] genes. *EYA1*, which encodes Eyes Absent 1, is a transcriptional co-activator and phosphatase. A heterozygous point mutation and deletions involving *EYA1* cause otofaciocervical syndrome type 1 (OTFCS2 [MIM:166780]), a condition marked by facial and skeletal anomalies including vertebral defects, low-set clavicles, winged scapulae, sloping shoulders as well as cup-shaped low-set ears, preauricular fistulas, hearing loss, branchial defects and mild intellectual disability (63). OTFC is rare but genetically heterogeneous and interestingly is also caused by recessively inherited mutations in *PAX1* (64,65), another AIS-associated gene [MIM#615560] (66). Mutations in *EYA1* also cause dominant branchiootic syndrome 1 [MIM:602588] (67) that is phenotypically similar to otofaciocervical syndrome (OTFC). The flanking gene *MSC* encodes musculin, a potent transcriptional repressor of myogenesis (68). Another top SNP, rs12050567, is encoded within the chromosome 15q26.1 gene *CHD2* [MIM:602119] (Supplementary Material, Table S3). While complete removal of *CHD2* is associated with embryonic and perinatal lethality, heterozygous $Chd2^{+/m}$ mice display

pronounced scoliosis with other developmental anomalies. Some of these features are also recapitulated in a human female patient carrying a balanced *de novo* translocation, t(X;15)(p22.2;q26.1)dn that disrupts *CHD2* encoded in 15q26.1, but no other obvious gene at the Xp22.2 locus (69). Other top meta-analysis loci are candidates for additional evaluation in AIS cohorts to determine their genetic contributions to disease.

Gene-functional enrichment analysis of our meta-analysis identified several gene classes that are over-represented in AIS and consequently may contribute to its pathogenesis. This analysis provided evidence for significant enrichment of genes involved in cartilage and connective tissue development pathways. In particular, along with *SOX6* and *SOX9*, we noted other genes within the cartilage development (GO:0051216) pathway that have been linked to scoliosis. *RUNX3* (OMIM 600210) is involved in maintaining spinal alignment where deletion of Runx3 in the peripheral nervous system in mouse induced peripubertal scoliosis (52) and the over-expression of Runx2 in mice caused scoliosis due to IVD degeneration (70). In a prior study, rare variant study of 391 severe AIS cases and 843 controls, pathway burden analysis identified significant enrichment of nonsynonymous and splice-site variants in extracellular matrix genes. While this association was mostly driven by variants in *COL11A2* (OMIM 120290), nominal association was also noted with variants in *COL11A1* (OMIM 120280) (71). Both COL11A1 and COL11A2 encode fibrillary collagens most abundantly found in cartilage (72). Continued integration of these results with other data sets, such as functional and physical protein–protein interaction networks, canonical pathways and rare-variant burden tests should prove useful for predicting pathways that are perturbed in AIS. Overall, these results expand the catalog of AIS candidate genes available for tissue-specific genetic targeting and other detailed characterizations that will elucidate the mechanistic underpinnings of AIS. Our study also demonstrates the power to identify novel candidate genes by stringent meta-analysis of common variant GWASs derived from genetically diverse populations.

Materials and Methods

Sampling information and genotyping platforms

Japan Cohorts (JP1-2). A total of 4841 case subjects (2104 in JP1 and 2737 in JP2) were recruited from collaborating hospitals (Japanese Scoliosis Clinical Research Group) as previously described (17). All patients underwent clinical examinations and radiologic evaluations by expert scoliosis surgeons. Informed consents were obtained from all the subjects or their parents and the ethics committees of RIKEN and participating institutions approved this study. For controls, 10 736 subjects in JP1 and 56 815 in JP2 were randomly selected from the BioBank Japan Project (see Web Resources) as previously described (17). All subjects in JP1 were genotyped on the Illumina Human (Hap550, Hap600, OmniExpress, Illumina, San Diego, CA) arrays as previously described (14). Cases in JP2 were genotyped using a multiplex PCR-based Invader assay (73), and the genotypes were called by visual inspection for all SNPs. Controls in JP2 were genotyped using the Illumina HumanOmniExpressExome BeadChip or a combination of HumanOmniExpress and HumanExome Beadchip and then imputed using minimac3 (74) with 1000G-Phase3.V.5.

Texas cohorts (Texas-GWAS1, Texas-GWAS2, Texas-GWAS3). The cases in Texas GWAS1-3 as well as controls in Texas-GWAS2 were recruited at Texas Scottish Rite Hospital for Children as approved by the Institute Review Board of the University Texas Southwestern Medical Center as previously described (36). Only subjects of self-reported as European (non-Hispanic white) ancestry were included in the present study. Phenotypes of all controls were reviewed to exclude individuals with musculoskeletal or neurological disorders. GWAS1 cases were genotyped as previously described (36). For Texas GWAS1 controls, we used genotypes by permission from the Database of Genotypes and Phenotypes (dbGaP) (see Web Resources) from the following three data sets: (i) National Institute of Neurological Disorders and Stroke CIDR: Genome-Wide Association Study in Familial Parkinson Disease (phs000126.v1.p1), where we used 861 European ancestry individuals that were negative for neurological disease by self-report. (ii) National Institute of Diabetes and Digestive and Kidney Diseases IBD Genetics Consortium: Crohn's Disease Genome-Wide Association Study (phs000130.v1.p1), where we used 505 subjects with non-Jewish European ancestry who were without history of inflammatory bowel disease. (iii) National Institute on Aging: Late Onset Alzheimer's Disease Family Study: Genome-Wide Association Study for Susceptibility Loci (phs000168.v1.p1), where we used 890 unrelated European ancestry healthy control subjects. GWAS2 cases and controls were genotyped on the Illumina HumanOmniExpress Array as previously described (37). GWAS3 cases were genotyped on the Illumina HumanCoreExome beadchip. For Texas GWAS3 controls, we utilized a single data set of individuals downloaded from dbGaP web site (see Web Resources) from Geisinger Health System-MyCode, eMERGE III Exome Chip Study under phs000957.v1.p1.

Missouri ohort (MO1). For MO1, AIS patients of European ancestry were included in the present study. All protocols were approved by the institutional review board. MO1 cases were genotyped on the Affymetrix 6.0 Array as described previously (75). For MO1 controls, we used 7681 European ancestry subjects downloaded from dbGaP web site (see Web Resources) from the Atherosclerosis Risk in Communities (ARIC) study (phs000280.v3.p1). The ARIC study is a prospective epidemiologic study conducted in US communities using 16 569 DNA samples, of which 13 443 were successfully genotyped using the Affymetrix 6.0 SNP array and passed the Broad Institute's quality control (QC) process. More information about the ARIC study can be found in 'The ARIC investigators' (76).

Hong Kong cohort (HK). A total of 539 subjects were recruited at The Duchess of Kent Children's Hospital as approved by the Institute Review Board of the University of Hong Kong (IRB approval number: UW 08-158). Cases and controls were genotyped with Affymetrix Genome-Wide Human SNP Array 6.0 and Affymetrix Human Mapping 500K Array Set at the Centre of Genomics Sciences in the University of Hong Kong, respectively. All controls in the HK cohort were reportedly healthy and excluded for any bone, muscle or neurological disorders.

Swedish–Danish cohort. A total of 1421 case females of European ancestry with a Cobb angle $>$ 15 degree were included in this cohort. Patients were recruited as approved by the institute review board in Stockholm (290/2006, 2009/1124-31/2, 2012/1595-31/2), Lund (LU 200–95, LU 280–99, LU 363–02, 567/2008, 2014/804) and Southern Denmark (S-2011002) as previously described in the Scoliosis and Genetics in Scandinavia study (ScoliGeneS) (Grauers, Wang *et al.* 2015). Cases genotyping was performed

using MassARRAY® System, combined with iPLEX® chemistry as described previously (77). For controls, we used a total of 1996 females of European ancestry from the following 2 cohorts: the Osteoporosis Prospective Risk Assessment study (OPRA) and the PEAK study as previously described (41). The OPRA cohort was genotyped using the Illumina Human Omni Express Exome chip and then imputed using IMPUTE 2.1.2. The PEAK-25 cohort was genotyped using the Infinium® Global Screening Array and then imputed using Eagle phasing v3 and Minimac. To determine concordance between actual genotypes and imputed genotypes, the concordance between genotyping performed using MassARRAY® System and iPLEX® (Agena Bioscience, San Diego, CA) chemistry was compared with the results from the genotyping and subsequent imputation in the Human Omni Express Exome chip in 100 individuals in the PEAK cohort. Concordance ranged between 95% and 100% for the different genotypes in this study.

Imputation and association

Genotypes for JP1 were imputed as previously described (17). For the other five data sets (HK, Texas-GWAS1, Texas-GWAS2, Texas-GWAS3 and MO1), the subjects in each study were checked for sex inconsistencies using PLINK.1.9 (78). Duplicated or related individuals were removed as described by Anderson *et al.* (79). We used the principal component analysis (80) on the data projected onto HapMap3 samples as recommended by Mitchell *et al.* (81) to correct for possible stratification due to underlying population differences. We applied initial per-SNP quality control measures using PLINK including genotyping call-rate per marker (>95%), deviation from Hardy–Weinberg equilibrium (cut-off $P = 10^{-4}$) and the missingness rate between cases and controls (cut-off $P = 10^{-4}$). Genotypes for autosomal chromosomes were imputed for each of HK, Texas-GWAS1-2 and MO1 data sets applying minimac2 (82) with the 1000G-Phase3.V.5 reference panel as described in the instructions provided by the software developer available from the software website (see Web Resources). Texas-GWAS3 genotypes were imputed using Minimac3 (74) as described in the instructions provided by the software developer available from the software website (see Web Resources) with the same reference panel used for other studies. Genotypes imputation for chromosome X in Texas-GWAS1–3, MO1 and the HK data were also done in Minimac3 (74) using only female subjects. Only unique bi-allelic common SNPs (minor allele frequency (MAF) > 0.05) with imputation quality Rsq >0.3 for the data imputed in minimac2,3 and info >0.4 for IMPUT2 were included for further analysis. To ensure that the shared SNPs among tested discovery studies did not have strand issues or any other obvious allele miscoding we plotted the alternative allele frequency in each study compared to their expected frequencies in the 1000 Genomes Project (83) (Supplementary Material, Fig. S1). The number of subjects and SNPs that remained after quality control in each cohort are shown in Supplementary Material, Tables S1 and S2, respectively.

Genome-wide association analysis in JP1 was performed for the imputed allele dosages in R package snpStats (84) by using a score test in a generalized linear model as described before (17). For JP2, the allele dosage was calculated then the association test was performed using logistic regression with gender correction using Mach2dat (38). Genome-wide association for the imputed allele dosages in each of the HK, Texas-GWAS1-3 and MO1 data sets was performed by Mach2dat (38) using logistic regression with gender and 10 principal components as covariates. For the Swedish–Danish cohort, the logistic regression was performed in IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY.

Meta-analysis

For the meta-analysis, we applied three pre-meta-analysis quality assurances. First, to maximize the overlap in the number of SNPs between discovery studies, we harmonized SNP identifiers by creating a unique SNP-ID called ChrPosID which uses the unique format 'chr: position' as recommended by (83). We successfully overlapped 3 821 032 by ChrPosID. Second, to ensure proper meta-analysis for SNPs shared among ethnicities, we required each SNP to have the identical effect and non-effect alleles throughout the discovery studies. As a post-meta-analysis QC, we checked whether observed effect sizes were homogeneous across studies, by calculating I^2 statistics and the Cochran's Q-test for heterogeneity (HetPVal) per SNP (85) as implemented in METAL (39). Then to perform the meta-analysis, we first controlled for inflation by applying genomic control correction of the individual GWASs results. Summary statistics across all six discovery studies were then combined using the inverse-variance-based method as implemented in METAL (39).

Estimation of AIS risk proportions explained by seven associated loci

To estimate the variation contributing to AIS risk for the loci identified in this study we tested the most commonly used measures as recommended by Witte *et al.* (86) and implemented in the INDI-V software (87). This included Approx. Herit. (approximation of heritability expressed as a proportion of total heritability), Sibling RR (sibling recurrence risk), Family RR (genetic variance on log relative risk scale explained) and pAUC (the proportion based on the area under the receiver–operating curve). For common low-risk variants, as we have for AIS in this study, the measures are fairly uniform. However, they can include different types and amounts of information (86).

Gene-functional enrichment analysis

We performed pathway enrichment analysis using Meta-Analysis Gene-Set Enrichment of variaNT Associations (MAGENTA) (88). Briefly, MAGENTA uses SNP *P*-values that are derived from genetic association studies to generate gene-based *P*-values. Genes are assigned *P*-values based on minimum SNP *P*-value (aka 'best SNP') among all SNPs that are located in a pre-defined genetic boundary. Following this, gene *P*-values are adjusted to account for confounding factors such as physical gene size, genetic distance and number of independent SNPs per kilobase (kb) for each gene. Gene *P*-values are then subjected to a GSEA (43) technique which was previous developed for transcriptional studies to compute significantly enriched pathways. In the present study, association *P*-values for all 3 493 832 SNPs from the discovery stage of the meta-analysis were used without prior filtering to account for the effects of all SNPs. We collected gene functional annotation data for 5917 various GO cellular components, biological processes and molecular functions from the Molecular Signatures Database (MsigDB, v6.1) (43). In addition, independent SNPs for Asian (EAS) and European (EUR) populations were generated using genotypes from the 1000-genome project and default pruning settings in PLINK (78). We tested our analysis on these two populations since our meta-study contains data from mixed populations. Gene boundaries were set at 110 kb upstream and 40 kb downstream as recommended by MAGENTA. Finally, nominal GSEA *P*-values were exported from MAGENTA output files into the R statistical package and *P*-values were then adjusted for multiple hypothesis testing using the FDR method.

Gene expression in a clonal chondrocytic cell line

Total RNA was extracted from the clonal chondrocytic cell line N1511 (46) using RNeasy Mini Kit (Qiagen, MD, USA) and from postnatal day 29 mouse rib cartilage using RNeasy Lipid Tissue Mini Kit (Qiagen, MD, USA) according to the manufacturer's instructions. cDNA was synthesized using High-Capacity RNA-to-cDNA kit (Applied Biosystems, CA, USA) according to the manufacturer's instructions. Primer sequences for real-time amplification are given in Supplementary Material, Table S7. The qPCR was performed on a 7500 real-time PCR system (Applied Biosystems, CA, USA). Amplification reactions consisted of iTaq Universal SYBR Green Supermix (Bio-Rad, CA, USA), 10 µm forward and reverse primers (Sigma-Aldrich, MO, USA) and 1 µL of 1:5 diluted cDNA template. Primers were designed using Primer3 and tested for specificity in silico using NCBI Primer-Blast (see Web Resources). The amplification was carried out with an initial holding step at 50°C for 2 min followed by a denaturation step at 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min in a 10 µL reaction volume. Each biologic replicate ($N = 3$) included three technical replicates. Melting curves were generated after the 40 cycles to confirm specificity.

Chromatin ChIP-seq

Costal chondrocytes from the anterior rib cage and sternum of postnatal day (P) 2–4 mice were isolated in 2 biological replicas as previously described (26). Briefly, rib cages were dissected, soft tissues were removed and rib cages digested at 37°C for 1 h in 2 mg/ml pronase and then washed in phosphate-buffered saline (PBS). Rib cages were then further digested for 1 h using 3 mg/ml collagenase D in Dulbecco's Modified Eagle Medium at 37°C and 5% CO2 followed by 3 PBS washes to remove digested tissue. Remaining cartilage was then further digested for 4–6 h in 3 mg/ml collagenase D and then filtered through a 45 µm cell strainer to obtain chondrocytes. Chondrocytes were cross-linked using 1% formaldehyde by standard techniques (89). ChIP was performed using an antibody against H3K27ac (Millipore, Burlington, MA 05-1334) using the LowCell# ChIP kit (Diagenode, Seraing, Belguim). Illumina sequencing libraries were generated using the Accel-NGS 2S Plus DNA Library Kit (Swift Biosciences, Ann Arbor, MI). Sequencing was done on a HiSeq 4000 and computational analyses was performed using Bowtie (90) and MACS (91).

Supplementary Materials

Supplementary Materials are available at *HMG* online.

Acknowledgements

We thank the patients and their families who participated in these studies. We thank Dr. Julia Kozlitina from the McDermott Center for Human Growth and Development at UT Southwestern Medical Center, Dallas, Texas, for helpful statistical advice and discussions on the AIS risk statistics. We are grateful to the clinical investigators who referred patients into the study from Japan Scoliosis Clinical Research Group (JSCRG), Texas Scottish Rite Hospital for Children Clinical Group (TSRHCCG) and St Louis, Missouri Group (SLMG). We are grateful for the ScoliGeneS study group for the help with patient recruitment and analyses of the Swedish–Danish cohort. The names and affiliations for JSCRG, TSRHCCG, SLMG and ScoliGeneS are included in the Supplemental data.

Conflict of Interest statement. None declared.

Funding

Japan Agency For Medical Research and Development (AMED) (contract no. 17ek0109212h0001 to S.I.); Scoliosis Research Society and NIH (NICHD P01 HD084387 to C.A.W.); the Swedish Research Council (number K-2013-52X-22198-01-3 and 2017-01639 to P.G. and E.E.); the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet (to P.G.); the Swedish Society of Spinal Surgeons (to P.G.); the Scoliosis Research Society (to P.G.).

References

1. Hresko, M.T. (2013) Clinical practice. Idiopathic scoliosis in adolescents. *N. Engl. J. Med.*, **368**, 834–841.
2. Herring, J.A. (2008) *Tachdjian's Pediatric Orthopaedics*. WB Saunders. Philadelphia.
3. Little, D.G., Song, K.M., Katz, D. and Herring, J.A. (2000) Relationship of peak height velocity to other maturity indicators in idiopathic scoliosis in girls. *J. Bone Joint Surg. Am.*, **82**, 685–693.
4. Martin, C.T., Pugely, A.J., Gao, Y., Mendoza-Lattes, S.A., Ilgenfritz, R.M., Callaghan, J.J. and Weinstein, S.L. (2014) Increasing hospital charges for adolescent idiopathic scoliosis in the United States. *Spine (Phila Pa 1976)*, **39**, 1676–1682.
5. Raggio, C.L. (2006) Sexual dimorphism in adolescent idiopathic scoliosis. *Orthop. Clin. North Am.*, **37**, 555–558.
6. Kruse, L.M., Buchan, J.G., Gurnett, C.A. and Dobbs, M.B. (2012) Polygenic threshold model with sex dimorphism in adolescent idiopathic scoliosis: the Carter effect. *J. Bone Joint Surg. Am.*, **94**, 1485–1491.
7. Sharma, S. and Wise, C. (2010) Dunwoodie, K.a. (ed.), In *The Genetic Architecture of Adolescent Idiopathic Scoliosis*. New York: Springer-Verlag, pp. 167–190.
8. Andersen, M.O., Thomsen, K. and Kyvik, K.O. (2007) Adolescent idiopathic scoliosis in twins: a population-based survey. *Spine (Phila Pa 1976)*, **32**, 927–930.
9. Kesling, K.L. and Reinker, K.A. (1997) Scoliosis in twins. A meta-analysis of the literature and report of six cases. *Spine (Phila Pa 1976)*, **22**, 2009–2014.
10. Riseborough, E.J. and Wynne-Davies, R. (1973) A genetic survey of idiopathic scoliosis in Boston, Massachusetts. *J. Bone Joint Surg. Am.*, **55**, 974–982.
11. Grauers, A., Danielsson, A., Karlsson, M., Ohlin, A. and Gerdhem, P. (2013) Family history and its association to curve size and treatment in 1,463 patients with idiopathic scoliosis. *Eur. Spine J.*, **22**, 2421–2426.
12. Tang, N.L., Yeung, H.Y., Hung, V.W., Di Liao, C., Lam, T.P., Yeung, H.M., Lee, K.M., Ng, B.K. and Cheng, J.C. (2012) Genetic epidemiology and heritability of AIS: a study of 415 Chinese female patients. *J. Orthop. Res.*, **30**, 1464–1469.
13. Ward, K., Ogilvie, J., Argyle, V., Nelson, L., Meade, M., Braun, J. and Chettier, R. (2010) Polygenic inheritance of adolescent idiopathic scoliosis: a study of extended families in Utah. *Am. J. Med. Genet. A*, **152A**, 1178–1188.
14. 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.
15. 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.
16. 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.
17. 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.
18. Jagla, K., Dolle, P., Mattei, M.G., Jagla, T., Schuhbaur, B., Dretzen, G., Bellard, F. and Bellard, M. (1995) Mouse Lbx1 and human LBX1 define a novel mammalian homeobox gene family related to the Drosophila lady bird genes. *Mech. Dev.*, **53**, 345–356.
19. Gross, M.K., Moran-Rivard, L., Velasquez, T., Nakatsu, M.N., Jagla, K. and Goulding, M. (2000) Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. *Development*, **127**, 413–424.
20. Brohmann, H., Jagla, K. and Birchmeier, C. (2000) The role of Lbx1 in migration of muscle precursor cells. *Development*, **127**, 437–445.
21. Schafer, K., Neuhaus, P., Kruse, J. and Braun, T. (2003) The homeobox gene Lbx1 specifies a subpopulation of cardiac neural crest necessary for normal heart development. *Circ. Res.*, **92**, 73–80.
22. Gross, M.K., Dottori, M. and Goulding, M. (2002) Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron*, **34**, 535–549.
23. Muller, T., Brohmann, H., Pierani, A., Heppenstall, P.A., Lewin, G.R., Jessell, T.M. and Birchmeier, C. (2002) The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron*, **34**, 551–562.
24. Guo, L., Yamashita, H., Kou, I., Takimoto, A., Meguro-Horike, M., Horike, S., Sakuma, T., Miura, S., Adachi, T., Yamamoto, T. *et al.* (2016) Functional investigation of a non-coding variant associated with adolescent idiopathic scoliosis in zebrafish: elevated expression of the ladybird homeobox gene causes body axis deformation. *PLoS Genet.*, **12**, e1005802.
25. 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.
26. Karner, C.M., Long, F., Solnica-Krezel, L., Monk, K.R. and Gray, R.S. (2015) Gpr126/Adgrg6 deletion in cartilage models idiopathic scoliosis and pectus excavatum in mice. *Hum. Mol. Genet.*, **24**, 4365–4373.
27. Henry, S.P., Liang, S., Akdemir, K.C. and de Crombrugghe, B. (2012) The postnatal role of Sox9 in cartilage. *J. Bone Miner. Res.*, **27**, 2511–2525.
28. Dy, P., Wang, W., Bhattaram, P., Wang, Q., Wang, L., Ballock, R.T. and Lefebvre, V. (2012) Sox9 directs hypertrophic maturation and blocks osteoblast differentiation of growth plate chondrocytes. *Dev. Cell*, **22**, 597–609.
29. Wei, Z., Wang, W., Bradfield, J., Li, J., Cardinale, C., Frackelton, E., Kim, C., Mentch, F., Van Steen, K., Visscher, P.M. *et al.* (2013) Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. *Am. J. Hum. Genet.*, **92**, 1008–1012.
30. Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M. *et al.* (2014) Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.*, **46**, 989–993.
31. Benyamin, B., He, J., Zhao, Q., Gratten, J., Garton, F., Leo, P.J., Liu, Z., Mangelsdorf, M., Al-Chalabi, A., Anderson, L. *et al.* (2017) Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nat. Commun.*, **8**, 611.
32. Khanshour, A.M. and Wise, C.A. (2018) Machida, M., W., S., & Dubousset, J. (eds), In *Current Understanding of Genetic Factors in Idiopathic Scoliosis*. Springer, Tokyo.
33. Londono, D., Kou, I., Johnson, T.A., Sharma, S., Ogura, Y., Tsunoda, T., Takahashi, A., Matsumoto, M., Herring, J.A., Lam, T.P. *et al.* (2014) A meta-analysis identifies adolescent idiopathic scoliosis association with LBX1 locus in multiple ethnic groups. *J. Med. Genet.*, **51**, 401–406.
34. Ogura, Y., Takeda, K., Kou, I., Khanshour, A., Grauers, A., Zhou, H., Liu, G., Fan, Y.H., Zhou, T., Wu, Z. *et al.* (2018) An international meta-analysis confirms the association of BNC2 with adolescent idiopathic scoliosis. *Sci. Rep.*, **8**, 4730.
35. Giampietro, P.F., Pourquie, O., Raggio, C., Ikegawa, S., Turnpenny, P.D., Gray, R., Dunwoodie, S.L., Gurnett, C.A., Alman, B., Cheung, K. *et al.* (2018) Summary of the first inaugural joint meeting of the International Consortium for Scoliosis Genetics and the International Consortium for Vertebral Anomalies and Scoliosis, March 16-18, 2017, Dallas, Texas. *Am. J. Med. Genet. A*, **176**, 253–256.
36. 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.
37. 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.
38. Li, Y., Willer, C., Sanna, S. and Abecasis, G. (2009) Genotype imputation. *Annu. Rev. Genomics Hum. Genet.*, **10**, 387–406.
39. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
40. Grauers, A., Wang, J., Einarsdottir, E., Simony, A., Danielsson, A., Akesson, K., Ohlin, A., Halldin, K., Grabowski, P., Tenne, M. *et al.* (2015) Candidate gene analysis and exome sequencing confirm LBX1 as a susceptibility gene for idiopathic scoliosis. *Spine J.*, **15**, 2239–2246.
41. Garg, G., Kumar, J., McGuigan, F.E., Ridderstrale, M., Gerdhem, P., Luthman, H. and Akesson, K. (2014) Variation in the MC4R gene is associated with bone phenotypes in elderly Swedish women. *PLoS One*, **9**, e88565.
42. Karol, L.A.M., Johnston, C.E., II, MD, Browne, R.H., PhD, Madison, M., PhD. (1993) Progression of the curve in boys who have idiopathic scoliosis. *J. Bone Joint Surg. Am.*, **75-A**, 1804–1810.
43. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-

wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **102**, 15545–15550.
44. Smits, P., Li, P., Mandel, J., Zhang, Z., Deng, J.M., Behringer, R.R., de Crombrugghe, B. and Lefebvre, V. (2001) The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev. Cell*, **1**, 277–290.
45. Smits, P. and Lefebvre, V. (2003) Sox5 and Sox6 are required for notochord extracellular matrix sheath formation, notochord cell survival and development of the nucleus pulposus of intervertebral discs. *Development*, **130**, 1135–1148.
46. Kamiya, N., Jikko, A., Kimata, K., Damsky, C., Shimizu, K. and Watanabe, H. (2002) Establishment of a novel chondrocytic cell line N1511 derived from p53-null mice. *J. Bone Miner. Res.*, **17**, 1832–1842.
47. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A. and Wysocka, J. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, **470**, 279–283.
48. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U S A*, **107**, 21931–21936.
49. Philippova, M., Joshi, M.B., Kyriakakis, E., Pfaff, D., Erne, P. and Resink, T.J. (2009) A guide and guard: the many faces of T-cadherin. *Cell Signal*, **21**, 1035–1044.
50. Fredette, B.J., Miller, J. and Ranscht, B. (1996) Inhibition of motor axon growth by T-cadherin substrata. *Development*, **122**, 3163–3171.
51. Poliak, S., Norovich, A.L., Yamagata, M., Sanes, J.R. and Jessell, T.M. (2016) Muscle-type identity of proprioceptors specified by spatially restricted signals from limb mesenchyme. *Cell*, **164**, 512–525.
52. Blecher, R., Krief, S., Galili, T., Biton, I.E., Stern, T., Assaraf, E., Levanon, D., Appel, E., Anekstein, Y., Agar, G. *et al.* (2017) The proprioceptive system masterminds spinal alignment: insight into the mechanism of scoliosis. *Dev Cell*, **42**, 388–399 e383.
53. Wang, Q., Cai, J., Wang, J., Xiong, C., Yan, L., Zhang, Z., Fang, Y. and Zhao, J. (2014) Down-regulation of adiponectin receptors in osteoarthritic chondrocytes. *Cell Biochem. Biophys.*, **70**, 491–497.
54. Liu, C.F. and Lefebvre, V. (2015) The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Res.*, **43**, 8183–8203.
55. Jackson, H.E., Ono, Y., Wang, X., Elworthy, S., Cunliffe, V.T. and Ingham, P.W. (2015) The role of Sox6 in zebrafish muscle fiber type specification. *Skelet. Muscle.*, **5**, 2.
56. Ikegawa, S. (2016) Genomic study of adolescent idiopathic scoliosis in Japan. *Scoliosis Spinal Disord.*, **11**, 5.
57. Comuzzie, A.G., Cole, S.A., Laston, S.L., Voruganti, V.S., Haack, K., Gibbs, R.A. and Butte, N.F. (2012) Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One*, **7**, e51954.
58. Zhao, S.X., Xue, L.Q., Liu, W., Gu, Z.H., Pan, C.M., Yang, S.Y., Zhan, M., Wang, H.N., Liang, J., Gao, G.Q. *et al.* (2013) Robust evidence for five new Graves disease risk loci from a staged genome-wide association analysis. *Hum. Mol. Genet.*, **22**, 3347–3362.
59. Johansson, A., Alfredsson, J., Eriksson, N., Wallentin, L. and Siegbahn, A. (2015) Genome-wide association study identifies that the ABO blood group system influences interleukin-10 levels and the risk of clinical events in patients with acute coronary syndrome. *PLoS One*, **10**, e0142518.
60. Fry, A.E., Griffiths, M.J., Auburn, S., Diakite, M., Forton, J.T., Green, A., Richardson, A., Wilson, J., Jallow, M., Sisay-Joof, F. *et al.* (2008) Common variation in the ABO glycosyltransferase is associated with susceptibility to severe Plasmodium falciparum malaria. *Hum. Mol. Genet.*, **17**, 567–576.
61. Tang, W., Teichert, M., Chasman, D.I., Heit, J.A., Morange, P.E., Li, G., Pankratz, N., Leebeek, F.W., Pare, G., de Andrade, M. *et al.* (2013) A genome-wide association study for venous thromboembolism: the extended cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Genet. Epidemiol.*, **37**, 512–521.
62. Dabelsteen, E. and Gao, S. (2005) ABO blood-group antigens in oral cancer. *J. Dent. Res.*, **84**, 21–28.
63. Rickard, S., Parker, M., van't Hoff, W., Barnicoat, A., Russell-Eggitt, I., Winter, R.M. and Bitner-Glindzicz, M. (2001) Oto-facio-cervical (OFC) syndrome is a contiguous gene deletion syndrome involving EYA1: molecular analysis confirms allelism with BOR syndrome and further narrows the Duane syndrome critical region to 1 cM. *Hum. Genet.*, **108**, 398–403.
64. Patil, S.J., Das Bhowmik, A., Bhat, V., Satidevi Vineeth, V., Vasudevamurthy, R. and Dalal, A. (2018) Autosomal recessive otofaciocervical syndrome type 2 with novel homozygous small insertion in PAX1 gene. *Am. J. Med. Genet. A*, **176**, 1200–1206.
65. Paganini, I., Sestini, R., Capone, G.L., Putignano, A.L., Contini, E., Giotti, I., Gensini, F., Marozza, A., Barilaro, A., Porfirio, B. *et al.* (2017) A novel PAX1 null homozygous mutation in autosomal recessive otofaciocervical syndrome associated with severe combined immunodeficiency. *Clin. Genet.*, **92**, 664–668.
66. Pohl, E., Aykut, A., Beleggia, F., Karaca, E., Durmaz, B., Keupp, K., Arslan, E., Palamar, M., Yigit, G., Ozkinay, F. *et al.* (2013) A hypofunctional PAX1 mutation causes autosomal recessively inherited otofaciocervical syndrome. *Hum. Genet.*, **132**, 1311–1320.
67. Vincent, C., Kalatzis, V., Abdelhak, S., Chaib, H., Compain, S., Helias, J., Vaneecloo, F.M. and Petit, C. (1997) BOR and BO syndromes are allelic defects of EYA1. *Eur. J. Hum. Genet.*, **5**, 242–246.
68. Lu, J., Webb, R., Richardson, J.A. and Olson, E.N. (1999) MyoR: a muscle-restricted basic helix-loop-helix transcription factor that antagonizes the actions of MyoD. *Proc. Natl. Acad. Sci. USA*, **96**, 552–557.
69. Kulkarni, S., Nagarajan, P., Wall, J., Donovan, D.J., Donell, R.L., Ligon, A.H., Venkatachalam, S. and Quade, B.J. (2008) Disruption of chromodomain helicase DNA binding protein 2 (CHD2) causes scoliosis. *Am. J. Med. Genet. A.*, **146A**, 1117–1127.
70. Sato, S., Kimura, A., Ozdemir, J., Asou, Y., Miyazaki, M., Jinno, T., Ae, K., Liu, X., Osaki, M., Takeuchi, Y. *et al.* (2008) The distinct role of the Runx proteins in chondrocyte differentiation and intervertebral disc degeneration: findings in murine models and in human disease. *Arthritis Rheum.*, **58**, 2764–2775.
71. Haller, G., Alvarado, D., McCall, K., Yang, P., Cruchaga, C., Harms, M., Goate, A., Willing, M., Morcuende, J.A., Baschal, E. *et al.* (2016) A polygenic burden of rare variants across extracellular matrix genes among individuals with adolescent idiopathic scoliosis. *Hum. Mol. Genet.*, **25**, 202–209.

72. Kuivaniemi, H., Tromp, G. and Prockop, D.J. (1997) Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. *Hum. Mutat.*, **9**, 300–315.
73. Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. and Nakamura, Y. (2001) A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.*, **46**, 471–477.
74. Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M. *et al.* (2016) Next-generation genotype imputation service and methods. *Nat. Genet.*, **48**, 1284–1287.
75. Buchan, J.G., Alvarado, D.M., Haller, G., Aferol, H., Miller, N.H., Dobbs, M.B. and Gurnett, C.A. (2014) Are copy number variants associated with adolescent idiopathic scoliosis? *Clin. Orthop. Relat. Res.*, **472**, 3216–3225.
76. Folsom, A.R., Chambless, L.E., Ballantyne, C.M., Coresh, J., Heiss, G., Wu, K.K., Boerwinkle, E., Mosley, T.H. Jr., Sorlie, P., Diao, G. *et al.* (2006) An assessment of incremental coronary risk prediction using C-reactive protein and other novel risk markers: the atherosclerosis risk in communities study. *Arch. Intern. Med.*, **166**, 1368–1373.
77. Einarsdottir, E., Grauers, A., Wang, J., Jiao, H., Escher, S.A., Danielsson, A., Simony, A., Andersen, M., Christensen, S.B., Akesson, K. *et al.* (2017) CELSR2 is a candidate susceptibility gene in idiopathic scoliosis. *PLoS One*, **12**, e0189591.
78. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.
79. Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. and Zondervan, K.T. (2010) Data quality control in genetic case-control association studies. *Nat. Protoc.*, **5**, 1564–1573.
80. 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.
81. Mitchell, B.D., Fornage, M., McArdle, P.F., Cheng, Y.C., Pulit, S.L., Wong, Q., Dave, T., Williams, S.R., Corriveau, R., Gwinn, K. *et al.* (2014) Using previously genotyped controls in genome-wide association studies (GWAS): application to the Stroke Genetics Network (SiGN). *Front. Genet.*, **5**, 95.
82. Fuchsberger, C., Abecasis, G.R. and Hinds, D.A. (2015) minimac2: faster genotype imputation. *Bioinformatics*, **31**, 782–784.
83. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Magi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E. *et al.* (2014) Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.*, **9**, 1192–1212.
84. Clayton, D. and Leung, H.T. (2007) An R package for analysis of whole-genome association studies. *Hum. Hered.*, **64**, 45–51.
85. Cochran, W.G. (1954) The combination of estimates from different experiments. *Biometrics*, **10**, 28.
86. Witte, J.S., Visscher, P.M. and Wray, N.R. (2014) The contribution of genetic variants to disease depends on the ruler. *Nat. Rev. Genet.*, **15**, 765–776.
87. Nolan, C. and Benyamin, B. (2014) Centre for Neurogenetics & Statistical Genomics, Queensland Brain Institute, The University of Queensland. *Vol.*, 2018.
88. Segre, A.V., DIAGRAM Consortium, MAGIC investigators, Groop, L., Mootha, V.K., Daly, M.J. and Altshuler, D. (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.*, **6**, 1–19.
89. VanderMeer, J.E., Smith, R.P., Jones, S.L. and Ahituv, N. (2014) Genome-wide identification of signaling center enhancers in the developing limb. *Development*, **141**, 4194–4198.
90. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
91. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.