Association of genetic variation in *COL11A1* with adolescent idiopathic scoliosis

Hao Yu[1*], Anas M. Khanshour[1*], Aki Ushiki[2,3*], Nao Otomo[4], Yoshinao Koike[4,5], Elisabet Einarsdottir[6], Yanhui Fan[7], Lilian Antunes[8], Yared H. Kidane[1], Reuel Cornelia[1], Rory Sheng[2,3], Yichi Zhang [2,3,9], Jimin Pei[10], Nick V. Grishin[10], Bret M. Evers[11,12], Jason Pui Yin Cheung[13], John A. Herring[14,15], Chikashi Terao[5], You-Qiang Song[7], Christina A. Gurnett[8], Paul Gerdhem[16,17,18], Shiro Ikegawa[4], Jonathan J. Rios[1,15,19,20], Nadav Ahituv[2,3], and Carol A. Wise[1,15,19,20**]

[1]Center for Pediatric Bone Biology and Translational Research, Scottish Rite for Children, Dallas, TX, USA
[2]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA
[3]Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA
[4]Laboratory of Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo, JP
[5]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, JP
[6]Science for Life Laboratory, Department of Gene Technology, KTH-Royal Institute of Technology, Solna, SE
[7]School of Biomedical Sciences, The University of Hong Kong, Hong Kong SAR, CN
[8]Department of Neurology, Washington University in St. Louis, St. Louis, MO, USA
[9]School of Pharmaceutical Sciences, Tsinghua University, Beijing, CN
[10]Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX, USA
[11]Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA
[12]Department of Ophthalmology, University of Texas Southwestern Medical Center, Dallas, TX, USA
[13]Department of Orthopaedics and Traumatology LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, CN
[14]Department of Orthopedic Surgery, Scottish Rite for Children, Dallas, TX, USA
[15]Department of Orthopaedic Surgery, University of Texas Southwestern Medical Center, Dallas, TX, USA
[16]Department of Clinical Science, Intervention & Technology (CLINTEC), Karolinska Institutet, Stockholm, Uppsala University, Uppsala, SE
[17]Department of Surgical Sciences, Uppsala University and
[18]Department of Orthopaedics and Hand Surgery, Uppsala University Hospital, Uppsala, SE
[19]Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX, USA
[20]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, USA

* These individuals contributed equally to the study.
**To whom correspondence should be addressed

Abstract

Adolescent idiopathic scoliosis (AIS) is a common and progressive spinal deformity in children that exhibits striking sexual dimorphism, with girls at more than five-fold greater risk of severe disease compared to boys. Despite its medical impact, the molecular mechanisms that drive AIS are largely unknown. We previously defined a female-specific AIS genetic risk locus in an enhancer near the *PAX1* gene. Here we sought to define the roles of *PAX1* and newly-identified AIS-associated genes in the developmental mechanism of AIS. In a genetic study of 10,519 individuals with AIS and 93,238 unaffected controls, significant association was identified with a variant in *COL11A1* encoding collagen (α1) XI (rs3753841; NM_080629.2_c.4004C>T; p.(Pro1335Leu); $P=7.07e^{-11}$, OR=1.118). Using CRISPR mutagenesis we generated *Pax1* knockout mice (*Pax1*$^{-/-}$). In postnatal spines we found that PAX1 and collagen (α1) XI protein both localize within the intervertebral disc (IVD)-vertebral junction region encompassing the growth plate, with less collagen (α1) XI detected in *Pax1*$^{-/-}$ spines compared to wildtype. By genetic targeting we found that wildtype *Col11a1* expression in costal chondrocytes suppresses expression of *Pax1* and of *Mmp3*, encoding the matrix metalloproteinase 3 enzyme implicated in matrix remodeling. However, this suppression was abrogated in the presence of the AIS-associated *COL11A1*^{P1335L} mutant. Further, we found that either knockdown of the estrogen receptor gene *Esr2*, or tamoxifen treatment, significantly altered *Col11a1* and *Mmp3* expression in chondrocytes. We propose a new molecular model of AIS pathogenesis wherein genetic variation and estrogen signaling increase disease susceptibility by altering a *Pax1-Col11a1-Mmp3* signaling axis in spinal chondrocytes.

Introduction

The human spinal column is a dynamic, segmented, bony and cartilaginous structure that is essential for integrating the brain and nervous system with the axial skeleton while simultaneously providing flexibility in three dimensions[1]. Idiopathic scoliosis is the most common developmental disorder of the spine, typically appearing during the adolescent growth spurt. Adolescent idiopathic scoliosis (AIS) is reported in all major ancestral groups, with a population prevalence of 1.5-3%[2,3]. Children with AIS usually present with a characteristic right-thoracic major curve pattern and a compensatory lumbar curve. Major thoracolumbar and lumbar curves are less frequent[1]. The three-dimensional nature of the deformity results in torsion in the spine that is most significant at the apex of the major curve, and changes in the structures of the vertebrae and ribs may develop as the curve worsens or progresses[1]. Children with thoracic curves, with larger curves at first presentation, and/or with greater remaining growth potential are at increased risk of progression, but this risk decreases sharply after skeletal maturity[1]. Sex is a recognized risk factor for AIS, with girls having at least a five-fold greater risk of progressive deformity requiring treatment compared to boys[4]. This well-documented sexual dimorphism has prompted speculation that levels of circulating endocrine hormones, particularly estrogen, are important exposures in AIS susceptibility[5].

The genetic architecture of human AIS is complex, and underlying disease mechanisms remain uncertain. Heritability studies of Northern European[6,7], North American[8,9], and South Asian[10] ancestral groups suggest that disease risk is multifactorial, caused by genetic and environmental contributions[2,11]. Accordingly, population-based genome-wide association studies (GWAS) in multiple ancestral groups have identified several AIS-associated susceptibility loci, mostly within non-coding genomic regions[11]. In particular, multiple GWAS have implicated noncoding regions

near the *LBX1*[12], *ADGRG6* (also known as *GRP126*)[13], and *BNC2*[14] genes. An association with alleles in an enhancer distal to *PAX1*, encoding the transcription factor paired box 1, was primarily driven by females, suggesting that it contributes to the sexual dimorphism observed in AIS[15]. Subsequent meta-analysis of combined AIS GWAS identified additional susceptibility loci. These included variants in an intron of *SOX6*, a transcription factor, that along with *PAX1*, is important in early spinal column formation[16]. Furthermore, gene enrichment analyses found significant correlation of AIS-associated loci with biological pathways involving cartilage and connective tissue development[17]. A more recent GWAS in a Japanese population identified fourteen additional AIS loci that are candidates for further evaluation[18]. In separate studies, genome sequencing in AIS cases and families identified enrichment of rare variants in the *COL11A2*[19] and *HSPG2*[20] genes, encoding components of the cartilage extracellular matrix (ECM). Hence variation affecting cartilage and connective tissue ECM is an emerging theme in the heterogeneous genetic architecture of AIS.

Pre-clinical animal models are essential tools for accelerating mechanistic understanding of AIS and for therapeutic testing[11]. In zebrafish, several genetic mutants with larval or later-onset spinal deformity have been described, including *ptk7*[21,22], *c21orf59*[23], *ccdc40*[24], *ccdc151*[25], *dyx1c1*, and *kif6*[26]. In rescue experiments, Rebello et al. recently showed that missense mutations in *COL11A2* associated with human congenital scoliosis fail to rescue a vertebral malformation phenotype in a zebrafish *col11a2* knockout line[27]. In mouse, conditional deletion of *Adgrg6* in skeletal cartilage (using *Col2a1*-Cre) produces a progressive scoliosis of the thoracic spine during postnatal development that is marked by herniations within the cartilaginous endplates of involved vertebrae. Progressive scoliosis, albeit to a lesser extent, was also observed when *Adgrg6* was

deleted from committed chondrocytes (using ATC-Cre)[28-30]. These studies demonstrate that cartilage and possibly other osteochondroprogenitor cells contribute to the scoliosis phenotype in these models[29]. Taken together, genetic and functional studies in mouse, although limited, support the hypothesis that deficiencies in biogenesis and/or homeostasis of cartilage, intervertebral disc (IVD), and dense connective tissues undermine the maintenance of proper spinal alignment during the adolescent growth spurt[11].

The combined contribution of reported AIS-associated variants is broadly estimated to account for less than 10% of the overall genetic risk of the disease[18]. To address this knowledge gap, we sought to define novel loci associated with AIS susceptibility in genes encoding proteins of the ECM (i.e. the "matrisome"[31,32]). Here we identify new genetic associations with AIS. Further, our functional assessments support a new disease model wherein AIS-associated genetic variation and estrogen signaling perturb a *Pax1-Col11a1-Mmp3* axis in chondrocytes.

Results

Nonsynonymous variants in matrisome genes are associated with increased risk of AIS.

The "matrisome" has been defined as "all genes encoding structural ECM components and those encoding proteins that may interact with or remodel the ECM"[33]. Proteins comprising the global ECM as currently defined have been identified by both experimental and bio-informatic methods [31]. We assembled 1027 matrisome genes as previously identified[34], including 274 core-matrisome and 753 matrisome-associated genes (N=1,027 total). For the genes encoding these 1,027 proteins, we identified all nonsynonymous common variants (MAF > 0.01) queried by the Illumina HumanCoreExome-24v1.0 beadchip and determined their genotypes in a discovery cohort of 1,358 cases and 12,507 controls, each of European ancestry (**Table 1**). After applying multiple quality control measures (see Methods section), we retained 2,008 variants in 597 matrisome genes for association testing (**Supplemental Table 1).** This sample size was estimated to provide at least 80% power to detect significant associations at the matrisome-wide level ($\alpha \leq 2.5e\text{-}05$), for alleles with population frequency ≥.05 and OR ≥1.5 (**Figure 1 – figure supplement 1a**). Two nonsynonymous variants, in *COL11A1* (rs3753841; NM_080629_c.4004C>T; p.(Pro1335Leu); odds ratio (OR)=1.236 [95% CI=1.134-1.347], P=$1.17E^{-06}$) and *MMP14* (rs1042704; NM_004995_c.817G>A; p.(Asp273Asn); OR=1.239 [95% CI=1.125-1.363], P=$1.89E^{-05}$) were significantly associated with AIS (**Figure 1a**). Given the sexual dimorphism in AIS and our prior observation of a female-predominant disease locus[15], we tested the 2,008 variants separately in females (N=1,157 cases and 7,138 controls). In females, the association with rs3753841 remained statistically significant, whereas rs1042704, near *MMP14*, was not associated with AIS in females (**Figure 1 – figure supplement 1b**). Our study was not sufficiently powered to test males separately.

To validate these results, we sought to replicate the associations of rs3753841 and rs1042704 in four independent AIS case-control cohorts, from North America, Europe, and eastern Asia, representing multiple ethnicities (total N = 9,161 AIS cases, 80,731 healthy controls, **Table 1**). Genotypes for both variants were extracted from these datasets and tested for association by meta-analysis together with the discovery cohort (see Methods). Meta-analysis of all cohorts together increased the evidence for association of both variants with AIS risk (**Figure 1b**). While a similar effect size was noted for rs1042704 in Japanese and Han Chinese cohorts, the results were less significant, likely due to lower minor allele frequencies (East Asian MAF = 0.02 compared to total non-Asian cohort MAF = 0.20) in these populations (**Figure 1 – figure supplement 1c**). Plotting recombination across both regions suggested that these signals were likely confined to blocks of linkage disequilibrium within the *COL11A1* and *MMP14* genes, respectively (**Figure 1 – figure supplement 1d, e**).

Rare dominant mutations in *COL11A1,* often disrupting a Gly-X-Y sequence, can cause Marshall (MRSHS) (OMIM# 154780) or Stickler syndromes (STL2) (OMIM# 604841) marked variously by facial anomalies, sensineural hearing loss, short stature, spondyloepiphyseal dysplasia, eye anomalies, ectodermal features, and scoliosis. Notably, our AIS cohort and particularly individuals carrying the rs3753841 risk allele were negative for co-morbidities or obvious features of Marshall or Stickler syndromes. Thus, variation in *COL11A1* is independently associated with AIS. Analysis of 625 exomes from the discovery cohort (46%) did not reveal a significant burden of rare *COL11A1* variants, nor common variants in linkage disequilibrium (R^2>0.6) with the top SNP rs3753841. These observations suggested that rs3753841 itself could be causal, although our

methods would not detect deep intronic variants that could contribute to the overall association signal.

COL11A1 is expressed in adolescent spinal tissues.

We next characterized *COL11A1* in postnatal spine development. *COL11A1* encodes one of three alpha chains of type XI collagen, a member of the fibrillar collagen subgroup and regulator of nucleation and initial fibril assembly, particularly in cartilage[35]. Spinal deformity is well-described in *Col11a1*-deficient (*cho/cho*) embryos[36]. In mouse tendon, *Col11a1* mRNA is abundant during development but barely detectable at three months of age[37]. We analyzed RNA-seq datasets derived from adolescent human spinal tissues[38], finding that *COL11A1* was upregulated in cartilage relative to bone and muscle. In cartilage, *PAX1* and *COL11A2* showed the strongest expression levels relative to other published human AIS-associated genes[13-15,17,19,20,39] (**Figure 2**). In all, most AIS-associated genes showed the strongest expression levels in cartilage relative to other adolescent spinal tissues.

We next sought to characterize *Col11a1* expression in spines of postnatal mice. To detect COL11A1 protein (collagen α1(XI)), we performed immunohistochemistry (IHC) and immunofluorescence (IF) microscopy using a collagen α1(XI) reactive antibody[40] in newborn (P0.5) and adolescent (P28) mice. In spines of P0.5 mice, strong staining was observed in the nucleus pulposus (NP) and in surrounding annulus fibrosus (AF) (**Figure 2b**). In thoracic spines of P28 mice, the compartments of the IVD were more distinct, and strong collagen α1(XI) staining was observed in each (**Figure 2c**). In regions of the cartilage endplate (CEP)-vertebral bone transition, collagen α1(XI) was detected in columnar chondrocytes, particularly in the

hypertrophic zone adjacent to condensing bone (**Figure 2c**). We also examined collagen α1(XI) expression in ribs, as these structures are also involved in the scoliotic deformity[1]. In P28 rib growth plates, as in spine, a biphasic pattern was observed in which collagen α1(XI) reactivity was most pronounced around cells of the presumed resting and pre-hypertrophic/hypertrophic zones (**Figure 2 – figure supplement a,b**). These data show that in mouse, collagen α1(XI) is detectable in all compartments of young postnatal IVD and, at the thoracic level, is particularly abundant in the chondro-osseous junction region of IVD and vertebral growth plate.

Col11a1 is downregulated in the absence of Pax1 in mouse spine and tail.

We previously identified AIS-associated variants within a putative enhancer of *PAX1* encoding the transcription factor Paired Box 1[15,17]. Pax1 is a well-described marker of condensing sclerotomal cells as they form segments that will eventually become the IVD and vertebrae of the spine[41-43]. We generated *Pax1* knock-out mice (*Pax1*$^{-/-}$) using CRISPR-Cas9 mutagenesis and validated them using sequencing and Southern blot (**Figure 3-figure supplement 3a-d**). Homozygous *Pax1*$^{-/-}$ mice were viable and developed kinks in the tail, as observed in other *Pax1*-deficient mice[44]. We next compared the expression of collagen α1(XI) protein in IVD and condensing bone of wildtype and *Pax1*$^{-/-}$ mice by performing IF staining in P28 spines (**Figure 3a**). In wildtype IVD, strong overlapping expression of collagen α1(XI) and PAX1 cells was observed, mostly within the CEP and chondro-osseous interface (**Figure 3a**). PAX1 staining was negative in *Pax1*$^{-/-}$ mice as expected, and collagen α1(XI) staining was dramatically diminished in CEP and the chondro-osseous vertebral borders. Moreover, the IVD in *Pax1*$^{-/-}$ mice was highly disorganized, without discernable NP, AF, and CEP structures as has been reported (**Figure 3-figure supplement e**)[45]. To test the effect of *Pax1* on expression of *Col11a1* and other AIS-

associated genes during embryonic development, RNA was isolated from vertebral tissue dissected from the tails of E12.5 wild-type and *Pax1*$^{-/-}$ mice and subjected to bulk RNA-seq and quantitative real-time PCR (qRT-PCR) (**Figure 3b).** Gene-set enrichment analysis of RNA-seq was most significant for the gene ontology term "extracellular matrix" (**Figure 3c**). By qRT-PCR analysis, expression of *Col11a1, Adgrg6,* and *Sox6* were significantly reduced in female and male *Pax1*$^{-/-}$ mice compared to wild-type mice (**Figure 3d-g**). These data show that loss of *Pax1* leads to reduced expression of *Col11a1* and the AIS-associated genes *Adgrg6* and *Sox6* in affected tissue of the developing tail.

Col11a1 regulates Mmp3 expression in chondrocytes.

COL11A1 has been linked with ECM remodeling and invasiveness in some cancers[46]. In solid tumors, *COL11A1* has been shown to alter ECM remodeling by enhancing *MMP3* expression in response to *TGFβ*[46]. *MMP3* encodes matrix metalloproteinase 3, also known as stromolysin, an enzyme implicated in matrix degradation and remodeling in connective tissues[47]. We confirmed strong *MMP3* mRNA expression, relative to *COL11A1,* in human spinal cartilage and bone, but minimal expression in spinal muscle (**Figure 4 – figure supplement 4a)**. We next cultured costal chondrocytes from P0.5 *Col11a1*$^{fl/fl}$ mice[40] and subsequently removed *Col11a1* by treating with Cre-expressing adenoviruses. After confirming *Col11a1* excision (**Figure 4a**), we compared *Mmp3* expression in these cells to cells treated with GFP-expressing adenoviruses lacking Cre activity. We found that *Mmp3* expression was significantly increased in cells where *Col11a1* mRNA expression was downregulated by about 70% compared to untreated cells (**Figure 4b**). Furthermore, Western blotting in these cells demonstrated a ~2-5-fold increase in pro-, secreted, and active forms of Mmp3 protein when collagen α1(XI) was reduced. The proteolytic processing

per se of precursor MMP3 into active forms[48] did not appear to be affected by *Col11a1* expression (**Figure 4c**). These results suggest that *Mmp3* expression is negatively regulated by *Col11a1* in mouse costal chondrocytes.

To test whether *Col11a1* affects *Mmp3* expression *in vivo,* we bred *Col11a1*$^{fl/fl}$ female mice with *Col11a1*$^{fl/fl}$ *ATC* males carrying the Acan enhancer-driven, doxycycline-inducible Cre (ATC) transgene[49]. ATC has been shown to harbor Cre-mediated recombination activity in most differentiated chondrocytes and in nucleus pulposus within two days of treating pregnant mothers with doxycycline starting at E15.5[49]. ATC activity was confirmed by crossing this line to the R26td$^{[Tomato]}$ reporter that ubiquitously expresses the fluorescent gene Tomato after Cre recombination. Strong Cre activity was seen in P0 pups of mothers treated with doxycycline at E15.5 in the NP, CEP, and annulus fibrosus (AF) of the IVD and in chondrocytes of the growth plates (**Figure 4 – figure supplement 4b**). Pregnant *Col11a1*$^{fl/fl}$ females were treated with doxycycline water from E15.5 to induce Cre expression in differentiated chondrocytes. Excision of *Col11a1* was confirmed in DNA from costal cartilage of *Col11a1*$^{fl/fl}$*ATC* cre-positive offspring (**Figure 4 – figure supplement 4c**). Consistent with results obtained by *in vitro* excision of *Col11a1*, cartilage from mice deficient in *Col11a1* showed ~4-fold upregulation of *Mmp3* mRNA expression relative to *Col11a1*$^{fl/fl}$ mice (**Figure 4d**).

AIS-associated variant in COL11A1 perturbs its regulation of MMP3.

Although low-resolution structures currently available for collagen triple helices are not useful for modeling the effects of individual variants on protein stability, we noted that the AIS-associated variant P1335L occurs at the third position of a Gly-X-Y repeat and consequently could be

structurally important in promoting stability of the triple helix, particularly if it is hydroxylated. We also noted that this variant is predicted to be deleterious by Combined Annotation Dependent Depletion (CADD)[50] and Genomic Evolutionary Rate Profiling (GERP)[51] analysis (CADD CADD= 25.7; GERP=5.75). Further, *COL11A1* missense variants have been shown to evoke transcriptional changes in ECM genes in cancer cells[52]. We therefore tested whether the *COL11A1*^{P1335L} sequence variant alters its regulation of *Mmp3* in chondrocytes. For this, SV40-immortalized cell lines were established from *Col11a1*$^{fl/fl}$ mouse costal chondrocytes and transduced with lentiviral vectors expressing green fluorescent protein (GFP) and *COL11A1*^{wt}, *COL11A1*^{P1335L}, or vector alone. After transduction, GFP-positive cells were grown to 50% confluence and treated with Cre-expressing adenovirus (ad5-Cre) to remove endogenous mouse *Col11a1* (**Figure 5a**). Using a human-specific *COL11A1* qRT-PCR assay, we detected overexpression of *COL11A1*^{wt} and *COL11A1*^{P1335L} compared to untransduced cells regardless of Cre expression (**Figure 5a**). Western blotting with an antibody directed against the HA epitope tag confirmed overexpression of human collagen α1(XI) protein (**Figure 5b**). Endogenous *Mmp3* mRNA and protein upregulation was evident by qRT-PCR and Western blotting, respectively, in untransduced cells treated with Ad5-Cre, as expected. Overexpressing human wildtype *COL11A1* suppressed *Mmp3* expression, consistent with the negative regulation we previously observed (**Figure 5 a,b)**. However, the *COL11A1*^{P1335L} mutant failed to downregulate *Mmp3* expression despite being overexpressed (**Figure 5a,b).** Thus, regulation of *Mmp3* appeared to be perturbed in the presence of the *COL11A1*^{P1335L} variant in these cells.

***Col11a1* and *Mmp3* are responsive to estrogen receptor signaling in chondrocytes.**

The expression of *Col11a1,* and of other ECM genes, is known to be estrogen-responsive in certain tissues, such as ovarian follicular cells[53]. Because of the suspected role of endocrine hormones in AIS, we investigated whether *Col11a1* expression was responsive to estrogen receptor siRNA-mediated knockdown in cultured chondrocytes. We first validated that *Mmp3* mRNA and protein levels were significantly increased after *Col11a1* knockdown in wildtype chondrocytes, as observed by Cre-mediated deletion in *Col11a1*$^{fl/fl}$ chondrocytes (**Figure 6a**). Estrogen receptor 2 (*Esr2),* but not estrogen receptor alpha (*Esr1),* was detected in mouse chondrocytes by qRT-PCR (data not shown). We therefore tested the consequences of *Esr2* siRNA-mediated knockdown on gene expression in chondrocytes. After *Esr2* knockdown, *Col11a1* as well as *Pax1* were significantly upregulated compared to scramble control, while *Mmp3* expression was significantly downregulated (**Figure 6b**). We also performed *Col11a1* knockdowns in these cells and noted upregulation of *Pax1* expression, suggesting a negative feedback loop between *Pax1* and *Col11a1* in these cells (**Figure 6b**). Simultaneous knockdown of *Col11a1* and *Esr2* expression reduced *Mmp3* expression to normal levels, supporting a possible interaction between *Col11a1* and *Esr2* in regulating *Mmp3*. Treating chondrocytes with tamoxifen, an estrogen receptor modulator, also upregulated *Col11a1* expression to similar levels as observed after *Esr2* knockdown, compared to cells treated with DMSO carrier (**Figure 6- figure supplement 6a**). These results suggest that estrogen signaling suppresses *Col11a1* expression. In cultured rat CEP cells, *Esr2* mRNA was downregulated, and *Mmp3* mRNA was upregulated after *Col11a1* knockdown, as observed in mouse chondrocytes (**Figure 6c, Figure 6 – figure supplement 6b**). However, *Esr2* knockdown did not significantly impact *Col11a1* or *Mmp3* expression in these cells (**Figure 6c**). Hence, we conclude that in cultured mouse chondrocytes, ESR2 signaling disrupts the suppression of *Mmp3* by *Col11a1*.

Discussion

Adolescent idiopathic scoliosis has been described in the medical literature for centuries, yet its underlying etiology has remained enigmatic[54]. Given that AIS originates in children who appear to be otherwise healthy, even its tissue of origin has been difficult to discern, and long debated[11]. The advent of powerful genotyping and sequencing methods in the last two decades has led to breakthrough discoveries of genetic loci associated with AIS, most in non-coding regions of the genome that are difficult to interpret biologically[11]. Aggregating these results, however, provided supportive evidence that pathways of cartilage and connective tissue ECM development are relevant in AIS etiology[11,17]. Here, in the largest multi-ethnic human cohort studied to date, we elected to test the hypothesis that alterations in ECM proteins themselves contribute to AIS susceptibility. This approach yielded most significant evidence for a common protein-altering variant in the *COL11A1* gene encoding collagen α1(XI), a minor yet critical component of cartilaginous ECM. Moreover, our studies define a *COL11A1*-mediated disease pathway (**Figure 7**) and point to the chondro-osseous junction of IVD and vertebrae spine as a relevant cellular compartment in AIS etiology.

The results of this study together with the previous observation of *COL11A2* rare variant enrichment in AIS supports a role for the collagen α1(XI) heterotrimer itself in its pathogenesis[19]. Collagen type XI, composed of three chains encoded by the *COL11A1*, *COL11A2*, and *COL2A1* genes (OMIM #s 120280,120290, 120140, respectively), is a minor component of collagen type II fibrils that are abundant in cartilage. Collagen type XI is also broadly expressed in testis, trachea, tendons, trabecular bone, skeletal muscle, placenta, lung, brain neuroepithelium, the vitreous of the eye, and intervertebral discs[55]. In the pericellular space, collagen α1(XI) initiates

fibrillogenesis with collagen type II fibrils, maintaining regular spacing and diameter of the collagen fibrils, while organizing the pericellular surface by interaction with cartilage proteoglycans[56,57]. Purified human collagen type XI, when added back to chondrocytes in *in vitro* culture, stimulates chondrogenesis while inhibiting hypertrophy, as measured by histologic staining, proliferation assays, and relative expression of chondrogenic early marker genes[58]. In newborn and one-month old mice, we found that collagen α1(XI) was abundant in IVD and at the chondro-osseous junction of IVD and vertebrae, particularly concentrated in pre-hypertrophic/hypertrophic chondrocytic cells. In long bone growth plates, Long et al.[59] recently identified eight distinct cell clusters after unsupervised analysis of single cell (scRNAseq) of flow-sorted hypertrophic chondrocytes from Col10a1Cre;Rosa26fs-tdTomato mice. At embryonic stage 16.5 (E16.5), *Col11a1* expression was highest in cells with signatures of pre-hypertrophic to hypertrophic transition, and lowest in cells with osteogenic signatures (M. Hilton, personal communication)[59]. Taken together, these results suggest that collagen α1(XI) normally participates in maintaining growth plate cells in a hypertrophic, pre-osteogenic state, although little is known about its precise molecular function in that compartment, or in the IVD, during spinal development. Spines of *Col11a1*-deficient mice (*cho/cho*) show incompletely formed vertebral bodies, spinal curvatures, and decreased separation between vertebrae, which are themselves less mineralized than in wildtype mice[36]. Notably, common *COL11A1* variants also have been associated with adult lumbar disc herniation (LDH) and degeneration (LDD), as well as DXA-measured bone size, spinal stenosis and spondylolisthesis[60-62].

Pax1 is a well-described marker of early spine development, where it activates a gene expression cascade starting at E12.5-13.5 in mouse development[44,63,64]. Our data showed that loss of *Pax1*

leads to decreased expression of *Col11a1*, *Sox6,* and *Adgrg6* in E12.5 tails of both male and female mice. The downregulation of *Col11a1* is consistent with a prior study of gene expression in flow-sorted GFP-labeled *Pax1*$^{-/-}$ embryonic IVD cells[64]. However, from these experiments we cannot discern if *Pax1* directly regulates *Col11a1* in *cis*, or by an indirect effect. It is likely, however, that *Col11a1* expression in developing tail is directly activated by binding SOX transcription factors, as a prior genomic study using chromatin immunoprecipitation and sequencing (ChIP-seq) in rat chondrosarcoma cells identified super enhancers near the *Col11a1* gene that were bound multiple times by SOX9 and SOX6[65]. The SOX5/6/9 trio is known to regulate many of the same genes as PAX1[64], but whether this includes *Col11a1* is unknown.

In mouse postnatal spines, we observed co-localization of collagen α1(XI) and PAX1 proteins specifically within the cartilaginous endplate-vertebral junction region that includes the vertebral growth plate. The endplate, which is important as the site allowing diffusion of nutrients from the circulation into the avascular inner IVD, harbors subpopulations of cells expressing type II collagen presumably organized by collagen type XI[41,43]. While the endplate is continuous with the vertebral growth plate in mice, it is important to note that in humans the endplate and vertebrae become distinctly separate structures with closure of the growth plates at puberty[41]. This is also the site of the ring apophyses that form the insertion of the IVD into vertebrae[66]. Lagging maturity of the ring apophysis, combined with mechanical forces across the IVD in the horizontal plane, has been proposed as an initiating factor leading to rotatory decompensation in the adolescent spine in AIS[66,67]. Recently, Sun et al. reported the discovery of a vertebral skeletal stem cell (vSSC) residing in the endplate and marked by expression of the genes *Zic1* and *Pax1,* along with other cell surface markers[68]. These vSSCs also express high levels of *Col11a1* (M. Greenblatt, personal

communication). It is interesting to consider that AIS-associated variation in collagen α1(XI), perhaps together with mechanical forces, could alter the differentiation trajectory of this cell niche. Altogether, extant data and our results strongly suggest that cell populations at the IVD-vertebral junction region are relevant in AIS pathogenesis. Further investigation is warranted to understand the developmental programmes of cells in this region of the spine.

Matrix metalloproteinase 3, also known as stromolysin, is a secreted enzyme expressed in connective tissues and in regions of endochondral ossification[69]. MMP3 has degradative activity toward a variety of ECM components, including proteoglycans, fibronectin, laminin, but notably not type I collagen[70]. Additionally, in chondrocytes MMP3 also has been shown to translocate to the nucleus, where it activates transcription of connective tissue growth factor (*CTGF/CCN2*) by binding to an element known as transcription enhancer dominant in chondrocytes (TRENDIC)[71,72]. Our observations of a *Col11a1-Mmp3* signaling axis in chondrocytes and CEP cells raise the possibility that *Col11a1* variation may have consequences for both MMP3 enzymatic activity levels and for *MMP3*-mediated transcriptional programming in these cells. *COL11A1* missense variants, usually altering glycine or proline in Gly-X-Y repeats in the collagen α1(XI) helical domain as with $COL11A1^{P1335L}$, are reported to be frequent in cutaneous squamous cell carcinomas and have been linked to transcriptional changes and tumor invasiveness[52]. The mechanisms by which chondrocytes or other cells sense such single amino acid changes in collagen α1(XI) and induce subsequent transcriptional responses are unknown but may involve direct interaction with integrins in the pericellular space[52].

We found that *Col11a1* expression is sensitive to estrogen receptor blockade or knockdown in chondrocytes. Type XI collagen is also a key player in organizing the pericellular space, which is critical for transmitting mechanical forces from the ECM to the cell[73]. Thus, it is interesting to consider that type XI collagen may effectively act as a receptor for environmental cues, i.e., mechanical forces and estrogen signaling, in the adolescent spine. Our study provides new insights into the regulation and signaling role of *Col11a1* in chondrocytes, and suggests potential mechanisms by which its genetic variation contributes to AIS susceptibility.

Methods

Discovery study

The cases in the discovery stage (USA TX: n=1,358) were recruited at Scottish Rite for Children as approved by the Institutional Review Board of the University Texas Southwestern Medical Center as previously described[74]. Subjects were genotyped on the Illumina HumanCoreExome BeadChip (Illumina, San Diego, CA, USA). For controls, we utilized 12,507 non-AMD GRU (non-age related macular degeneration general research use) subjects of the European ancestry downloaded from dbGaP web site (https://www.ncbi.nlm.nih.gov/gap/) from the International Age-Related Macular Degeneration Genomics Consortium study (IAMDGC: phs001039.v1.p1.c1; https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001039.v1.p1). The subjects from the IAMDGC study were also genotyped on the Illumina HumanCoreExome Beadchip-24v1.0 platform[75]. We merged cases and controls and applied quality controls to the genotypes for 468,801 overlapping SNPs using PLINK.1.9. [76] as described in [17]. In summary, samples with sex inconsistencies or from duplicated or related individuals or ancestral outliers as identified by principal component analysis (PCA) were removed, leaving 13,865 samples in the analysis. Genotypes were corrected for strand direction, and SNPs with call-rate per marker <95%, deviating from Hardy-Weinberg equilibrium (HWE) (cutoff P-value = 10^{-4}), or with significant missingness rate between cases and controls (cutoff P-value = 10^{-4}) were removed, leaving 341,759 SNPs in the analysis. Genotypes for SNPs across autosomal chromosomes were imputed using Minimac3 with the 1000G-Phase3.V.5 reference panel as described in the instructions available from the software website[77]. Protein-coding changes were annotated with ANNOVAR using RefSeq-based transcripts[78]. External

databases included allele frequencies from gnomAD[79]; variant pathogenicity in Clinvar[80]; Combined Annotation Dependent Depletion (CADD) scores[81]; Genomic Evolutionary Rate Profiling (GERP) scores [82] and protein domains in IntroPro [83]. Only bi-allelic common (MAF>0.01) protein-altering SNPs with imputation quality Rsq >=0.3 within matrisome genes[34] were included for further analysis. Matrisome genes used can be found in the Molecular Signature Database (MsigDB)[84,85] (https://www.gsea-msigdb.org/gsea/msigdb/cards/NABA_MATRISOME). Genetic association for the imputed allele dosages in the discovery cohort (USA TX) was performed in Mach2dat [86] using logistic regression with gender and ten principal components as covariates. The genomic regions of the associated loci were visualized with LocusZoom software[87] utilizing linkage disequilibrium (LD) information from 1000 Genomes EUR populations.

Meta-analysis study

For the meta-analysis stage we utilized four cohorts [USA MO: n=2,951(1,213 cases and 1,738 controls], Swedish-Danish populations [SW-D: n= 4,627 (1,631 cases and 2,996 controls), as described in [17,88]], Japan [JP: n=79,211(5,327 cases and 73,884 controls)] and Hong Kong [HK: n=3,103 (990 cases and 2,113 controls)] to check significant candidates from the discovery study. Summary statistics across the discovery study and the 4 replication cohorts [total N=103,757 (10,519 cases and 93,238 controls)], were combined as previously described[17] using METAL[89].

USA MO cohort: Whole exome sequencing (WES) data from 1,213 unrelated idiopathic scoliosis cases of European ancestry with spinal curvature greater than 10˚ Cobb angle were derived from the Adolescent Idiopathic Scoliosis 1000 Exomes Study (dbGAP accession number: phs001677), and recruited from St. Louis Children's Hospital, and St. Louis Shriners Hospital for Children.

Patients and/or parents provided study consent and IRB approval was obtained from each contributing institution. For controls, exome data from 1,738 unrelated samples of European ancestry were provided by investigators at Washington University School of Medicine in St. Louis, MO (dbGAP accession numbers: phs000572.v8.p4 and phs000101.v5.p1), and Oregon Health & Science University in Portland, OR (https://gemini.conradlab.org/). Exome data were aligned to the human genome reference (GRCh37) using BWA-MEM (v0.7.15). Variant calling of single nucleotide variants (SNVs) and insertion and deletion variants (INDELs) were generated first for each single sample in cases and controls and then combining all samples with joint genotyping method, described in GATK Best-Practices (Genome Analysis Toolkit (GATK v3.5) https://gatk.broadinstitute.org/hc/en-us/sections/360007226651-Best-Practices-Workflows). All cases and controls samples were classified as unrelated and of European ancestry using relationship inference [90] and principal component analysis [76]. Association analysis of variants rs3753841 and rs1042704 were performed using logistic regression adjusting for sex and PCs in PLINK[76].

JP cohort: Informed consents were obtained from all the subjects or their parents and the ethics committees of RIKEN and participating institutions approved this study. 5,327 case subjects were recruited from collaborating hospitals (Japanese Scoliosis Clinical Research Group) as previously described[18]. For controls, 73,884 subjects were randomly selected from the BioBank Japan Project, and subjects were genotyped on Illumina Human BeadChips as previously described [14,18]. Imputation and association analyses in JP were performed as previously described[91].

HK cohort: 3,103 subjects were recruited at The Duchess of Kent Children's Hospital as approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (IRB approval number: UW 08-158). All 990 cases were characterized by Cobb

angles greater than 40 degrees with onset age between 10 and 18 years old. Congenital, neuromuscular and syndromic scoliosis samples were excluded. We used 2,113 controls from the Chinese population with no spinal deformities on MRI scans [92]. Cases and controls were genotyped using the Illumina Infinium OmniZhongHua-8 BeadChip and analyzed with GenomeStudio 2.0 software. The quality control approach adopted the GWA tutorial developed by Andries et al [93]. The filtered genotyping data of cases and controls was phased and imputed using SHAPEIT [94] and IMPUTE2 [95], respectively. Logistic model association analysis was performed using PLINK 1.9[76].

Stratification-by-sex test

To investigate sex-specificity in the *COL11A1* and *MMP14* loci, we performed stratification-by-sex analysis in the discovery study (USA_TX). Association for the imputed allele dosages in rs3753841 and rs37538 were computed separately for females (1,157 cases and 7138 controls) using logistic regression with ten principal components as covariates in Mach2dat[86].

RNAseq of human tissues

RNAseq was performed as previously described[96]. Read counting and transcript quantification were performed using HTSeq[97]. Finally, reads were normalized using DESeq2 tools[98] and TPM values were generated using the Kalisto pipeline[99].

Animal studies

All mouse and rat work was conducted per IACUC approved protocols and in accordance with AALAC and NIH guidelines.

Generation of *Pax1* knockout mice: Mouse work was approved by the UCSF IACUC, protocol number AN181381. Two gRNAs were designed to target the 5′ and 3′ ends of *Pax1* gene (gRNA sequence shown in **Figure 5-figure supplement 5a)** using the gRNA design tool on the Integrated DNA Technologies (IDT, Newark, NJ, USA) website and selected based on low off-target and high on-target scores. The knockout allele was generated using *i*-GONAD[100] as previously described [101].

To validate proper generation of knockout, mice were analyzed by genotyping with primers shown in Appendix A Key Resources Table, Sanger sequencing of PCR-amplified DNA, and Southern blot (**Figure 5-figure supplement 5a)**. For Southern blot analyses, genomic DNA were treated with NcoI (Cat #R0193, New England Biolabs, MA, USA) and fractionated by agarose gel electrophoreses. Following capillary transfer onto nylon membranes, blots were hybridized with Digoxigenin (DIG)-labeled DNA probes (corresponding to chr2:147,202,083-147,202,444; mm9) amplified by the PCR DIG Probe Synthesis Kit (Cat #11636090910, Sigma-Aldrich, MO, USA). The hybridized probe was immunodetected with antidigoxigenin Fab fragments conjugated to alkaline phosphatase (Cat # 11093274910, Sigma-Aldrich, MO, USA) and visualized with a CDP star (Cat #11685627001, Sigma-Aldrich, MO, USA) according to the manufacturer's protocol. Chemiluminescence was detected using the FluorChem E (Cat #92-14860-00, ProteinSimple, CA, USA).

Col11a1$^{fl/fl}$ mice: The *Col11a1*$^{fl/fl}$ mouse line was kindly provided by Dr. Lou Soslowsky with permission from Dr. David Birk. All *Col11a1*$^{fl/fl}$ and wildtype mice were handled per UTSW IACUC, protocol number 2016-101455.

Other mice: Cartilage was harvested from C57B46 wild type mice for siRNA-mediated knockdowns experiments.

Histologic methods

For thin cryostat sections, P0.5 mouse whole body was fixed in 4% paraformaldehyde for 6 hours followed by 10% sucrose for 12 hours, then transferred to 18% sucrose for 24 hours. Tissues were then embedded in optimal cutting temperature compound (OCT) and sectioned using low-profile blades on a Thermo Shandon Cryostar NX70 cryostat and all sections were lifted on APES clean microscope slides. For whole mount images, samples were treated similarly with the addition of 2% polyvinylpyrrolidone (PVP) during the cryoprotection step and frozen in 8% gelatin (porcine) in the presence of 20% sucrose and 2% PVP. Samples were sectioned at a thickness of 10 μm. Slides were stored at -80˚ C until ready for use. For P28 and older mice, spines were removed then fixed, decalcified, and embedded in OCT. Spines were processed by making 7 um thick lateral cuts the length of the spine.

Collagen α1(XI) was detected by immunohistochemistry staining using affinity-purified antisera against peptide (C) YGTMEPYQTETPRR-amide (Genescript, NJ, USA) as described[40] and secondary horseradish peroxidase (HRP) conjugated affinity purified secondary antibody (Cat # AP187P, Millipore-Sigma Aldrich, MO, USA). Briefly, frozen sections were equilibrated to room temperature for 1 hour, then fixed with 4% paraformaldehyde in PBS at 4 degrees Celsius for 20 minutes. Slides were washed, treated with 3% H_2O_2 in methanol for 10 minutes to block endogenous peroxidase, washed, and transferred to PBS with 0.05% TWEEN 20 (Cat #P3563-10PAK, Sigma-Aldrich, MO, USA) pH 7.4. Slides were blocked with 0.5 % goat serum in PBS mix with 0.2% Triton 100 (Cat #T8787, Sigma-Aldrich, MO, USA) at room temperature for 1.5 hours. The primary collagen α1(XI) affinity purified antibody was applied at 0.40 mg/ml and slides were incubated overnight at 4°C. Afterward slides were washed in PBS Tween 20 for three times

and treated with goat anti- rabbit-HRP for 1.5 hours, then washed three times in PBS Tween 20. After applying 3,3'-diaminobenzidine (DAB) solution, slides were washed and counterstained with Mayer's hematoxylin (Cat #MHS80, Sigma-Aldrich, MO, USA), washed, dehydrated, and mounted.

For collagen α1(XI) and PAX1 IF studies, P0.5 mice, P28 spine and ribs sections were fixed in 4% paraformaldehyde (PFA) for 20 minutes then washed with PBS + 0.1% Triton three times, before incubation with 10% normal goat serum in PBS + 0.1% Triton for 30 min to block the background. Slides were incubated with goat anti-mouse collagen α1(XI) antibody at 1:500 dilution and mouse anti-rat PAX1(Cat #MABE1115M, Sigma-Aldrich, MO, USA), in PBS + 0.1% Triton + 1% normal goat serum at 4 °C overnight. Secondary antibodies used were 1:5000 anti-rat Alexa488 and anti-mouse Alexa594 conjugated antibodies (Cat #A32740 Invitrogen, CA, USA). The sections were mounted using ProLong Gold with DAPI (Cat #S36964 Invitrogen, CA, USA) for imaging as described[102]. All images were taken with Carl Zeiss Axio Imager.M2 fluorescence microscope (Zeiss,Oberkochen, DE) .

Rib cartilage and IVD cell culture

Mouse costal chondrocytes were isolated from the rib cage and sternum of P0.5 mice. Rat IVD was removed intact from 1-month female rats and immediately separated into NP, AF, and CEP isolates. Subsequently, tissues were incubated and shaken with 2 mg/ml Pronase solution (Cat #10165921001 Sigma-Aldrich, Inc., St. Louis, MO, USA) for 1 hour, followed by 1.5 hours digestion with 3 mg/ml Collagenase D solution. (Cat #11088882001 Sigma-Aldrich, Inc., St. Louis, MO, USA), then 5 hours digestion with 0.5 mg/ml Collagenase D solution before 3 times

PBS wash. Filtered, dissociated cells were seeded in Dulbecco's modified Eagle's medium (DMEM; Cat #MT15017CV Thermo Fisher Scientific, MA, USA) containing 10% fetal bovine serum (FBS), 100 µg/ml streptomycin and 100 IU/ml penicillin. Remaining cartilage tissues underwent further digestion in 86 U/ml type 1 collagenase (Cat # SCR103 Sigma-Aldrich, Inc., St. Louis, MO, USA) overnight. Cells were collected and cultured in DMEM with 10% FBS plus 100 µg/ml streptomycin and 100 IU/ml penicillin.

SV40 immortalization and transfection of primary chondrocytes.

Col11a1$^{fl/fl}$ mouse costal chondrocytes were isolated from the rib cage and sternum of P0.5 mice. The cells were transduced with pRRLsin-sv40 T antigen-IRES-mCherry lenti-virus[103] for 48 hours, then sorted for mCherry-positive cells by flow cytometry. mCherry-positive cells were then infected with plv-eGFP, plv-eGFP-COL11A1-HA, plveGFP-COL11A1^{P1335L}-HA constructs. After expansion, GFP-positive cells were sorted by flow cytometry and seeded in 24-well plates.

Adenovirus treatment.

SV40 induced *Col11a1*$^{fl/fl}$ mouse costal chondrocytes were grown to 50% confluency. Afterward, cells were treated with 2ul Ad5-CMV-cre adenovirus (titer 1.8x10^{11}pfu/ml) and Ad5-CMV-eGFP (titer 1.65x10^{10}pfu/ml) as control. Both virus strains were from the Gene Vector Core facility, Baylor College of Medicine. After 48-hours the cells were harvested for mRNA and protein lysate.

RNA-seq and qRT-PCR

For *Pax1* knockout studies, total RNA was collected from E12.5 tails using TRIzol (Cat #15596026, Thermo Fisher Scientific, MA, USA) and converted to cDNA using ReverTra Ace qPCR-RT master mix with genomic DNA (gDNA) remover (Cat #FSQ-301, Toyobo, Osaka, Japan). Sequencing was done using an Illumina Novaseq platform and the data were analyzed using Partek Flow (Version 10.0) and Gene ontology[104]. qPCR was performed using SsoFast EvaGreen supermix (Cat #1725205, Bio-Rad, CA, USA). Primer sequences used for qPCR are shown in Appendix A, Key Resources table.

To quantify the expression level of *Col11a1*, *Mmp3*, and marker genes in IVD compartments and rib cartilage, cultured cells were collected in RNeasy (Qiagen, Inc.) for RNA purification. Taqman Reverse Transcription Kit (Cat #4387406 Thermo Fisher Scientific, MA, USA) was used to reverse transcribe mRNA into cDNA. Following this, RT-qPCR was performed using a Power SYBR Green PCR Master Mix Kit (Cat #1725271, Bio-Rad, CA, USA). The primer sequences for the genes used in this study are listed in the Appendix A Key Resources Table. Gene expression was calculated using the ΔΔCT method after normalizing to GAPDH.

siRNA knockdown

Mouse rib cartilage cells seeded in 6 well-plates were 60-80% confluent at transfection. Lipofectamine RNAiMAX reagent (Cat #13778030 Thermo Fisher, Waltham, MA, USA) was diluted (9ul in 500ul) in Opti-MEM Medium (Cat # 31985070 Thermo Fisher MA, USA). 30 pmol siRNA was diluted in 500ul Opti-MEM medium, then added to diluted Lipofectamine RNAiMAX Reagent. siRNA-lipid complex was added to cells after 5-minute incubation at room temperature. Cells were harvested after 72 hours.

Western blotting

For MMP3 western blotting, a total of 30 μg protein mixed with SDS-PAGE buffer was loaded on 12% SDS-polyacrylamide gel for electrophoresis. For collagen α1(XI) western blotting, 50 μg protein mixed with SDS-PAGE buffer was loaded on 4-20% SDS-polyacrylamide gel. The separated proteins were then transferred to nitrocellulose membranes (Cat #77010 Thermo Fisher Waltham, MA, USA) at 100 volts for 2-3 hours. The membrane was first incubated with blocking buffer containing 5% defatted milk powder, and then exposed to 0.1 mg/mL mouse anti-rabbit Mmp3 (Cat #ab214794 Abcam, Cambridge, MA, USA) or anti-rabbit Col11a1 (Cat #PA5-38888 Thermo Fisher Waltham, MA, USA) overnight. The samples were then washed thoroughly with TBS buffer, followed by incubation with HRP-labeled antirabbit IgG secondary antibodies 1:5000, (Cat #32460 Thermo Fisher Waltham, MA, USA) overnight. The membranes were then washed with TBS buffer. GAPDH was detected by a rabbit anti-mouse antibody (Cat #14C10 Cell Signaling, MA, USA) and used as the internal loading control.

Data availability

Summary data for GWAS3 is deposited in the NHGRI-EBI GWAS catalog (accession #GCST90179478). Otherwise, data generated or analyzed in this study are included in the manuscript.

Acknowledgements

We thank the patients and their families who participated in these studies. We are grateful to the clinical investigators who referred patients into the study from Japan Scoliosis Clinical Research Group (JSCRG), Scottish Rite for Children Clinical Group (SRCCG). 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, ScoliGeneS are included Appendix B. We also thank Drs. Carlos Cruchaga, Sanjay Jain, Matthew Harms, and Don Conrad for allowing us to use exome data from their cohort studies. We are grateful to Dr. James Lupski and the Baylor-Hopkins Center for Mendelian Genomics for providing exome sequencing of AIS cases. The University of Pennsylvania is acknowledged for providing the *Col11a1*$^{fl/fl}$ mouse line. We thank Drs. D. Burns for help interpreting histological studies. We thank Dr. M. Hilton for sharing scRNAseq data as described in reference (51). This study was funded by JSPS KAKENHI Grants (22H03207 to SI), the Swedish Research Council (number K-2013-52X-22198-01-3 and 2017-01639 to PG and EE), the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet (to PG), Center for Innovative medicine (CIMED), Karolinska Institutet (to PG), the Department of Research and Development of Vasternorrland County Council (to PG), and Karolinska Institutet research funds (to PG) and Research Grants Council of Hong Kong (No:17114519 to YQS), The National Institutes of Health GM127390 (to NG) and the Eunice Kennedy Shriver National Institute of Child Health & Development of the National Institutes of Health (P01HD084387 to CAW and NA). Content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The data used for the analyses described in the USA MO cohort were obtained from the database of Genotypes and Phenotypes (dbGaP) Adolescent Idiopathic Scoliosis 1000 Exomes

Study (phs001677.v1.p1) which included investigators at Washington University in St Louis (M. Dobbs), Shriners Hospital for Children, St Louis (M. Dobbs), University of Colorado (N. Miller), University of Iowa (S. Weinstein and J. Morcuende), University of Wisconsin (P. Giampietro), and Hospital for Special Surgery (C. Raggio). Sequencing of these samples was funded by NIH NIAMS 1R01AR067715 (M. Dobbs and CG). Use of the dbGAP dataset phs001039.v1.p1 is gratefully acknowledged. All contributing sites and additional funding information for dbGAP dataset phs001039.v1.p1_are acknowledged in this publication: Fritsche et al. (2016) Nature Genetics 48 134–143, (doi:10.1038/ng.3448); The International AMD Genomics consortium's web page is: http://eaglep.case.edu/iamdgc_web/, and additional information is available on: http://csg.sph.umich.edu/abecasis/public/amd015/. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing computing resources that have contributed to the research results reported within this paper. URL: http://www.tacc.utexas.edu

Author contributions

H.Y., A.M.K., A.U., X-Y. L, N.O, Y.K., E.E., Y.F., L.A., Y.H.K., R.C., R.S, Y.Z., J.P. performed genetic, bioinformatic, or proteomic analyses or molecular biology experiments. N.V.G., B.M.E., J.P.Y.C., and J.A.H. provided technical or clinical advice. C.T., Y.-Q.S., C.A.G., P.G., and S.I. supervised genetic experiments and provided summary data. J.J.R., N.A., and C.A.W. supervised the work. C.A.W conceived the study and wrote the manuscript. All co-authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Richards, B.S., Sucato, D.J., Johnston C.E. Scoliosis. in *Tachdjian's Pediatric Orthopaedics*, Vol. 1 (ed. Herring, J.A.) (Elsevier, 2020).
2. Wise, C.A. The Genetic Architecture of Idiopathic Scoliosis. in *Molecular Genetics of Pediatric Orthopaedic Disorders* (ed. Carol A. Wise, J.J.R.) 71-90 (Springer, New York, 2014).
3. Hresko, M.T. Clinical practice. Idiopathic scoliosis in adolescents. *N Engl J Med* **368**, 834-41 (2013).
4. Karol, L.A.M., Charles E. Johnston, II, MD, Richard H. Browne, PhD, Michael Madison, PhD. Progression of the Curve in Boys Who Have Idiopathic Scoliosis. *The Journal of Bone and Joint Surgery* **75-A**, 1804-1810 (1993).
5. Liang, Z.T., Guo, C.F., Li, J. & Zhang, H.Q. The role of endocrine hormones in the pathogenesis of adolescent idiopathic scoliosis. *FASEB J* **35**, e21839 (2021).
6. Wynne-Davies, R. Familial (idiopathic) scoliosis. A family survey. *J Bone Joint Surg Br* **50**, 24-30 (1968).
7. Grauers, A., Rahman, I. & Gerdhem, P. Heritability of scoliosis. *Eur Spine J* **21**, 1069-74 (2012).
8. Riseborough, E.J. & Wynne-Davies, R. A genetic survey of idiopathic scoliosis in Boston, Massachusetts. *J Bone Joint Surg Am* **55**, 974-82 (1973).
9. Kruse, L.M., Buchan, J.G., Gurnett, C.A. & Dobbs, M.B. Polygenic threshold model with sex dimorphism in adolescent idiopathic scoliosis: the Carter effect. *J Bone Joint Surg Am* **94**, 1485-91 (2012).
10. Tang, N.L. *et al.* Genetic epidemiology and heritability of AIS: A study of 415 Chinese female patients. *J Orthop Res* **30**, 1464-9 (2012).
11. Wise CA, S.D., Ushiki A, Khanshour A, Kidane YH, Makki N, Gurnett CA, Gray RS, Rios JJ, Ahituv N, Solnica-Krezel L. The cartilage matrisome in adolescent idiopathic scoliosis. *Bone Research* **8**(2020).
12. Takahashi, Y. *et al.* A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. *Nat Genet* **43**, 1237-40 (2011).
13. Kou, I. *et al.* Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nat Genet* **45**, 676-9 (2013).
14. Ogura, Y. *et al.* A Functional SNP in BNC2 Is Associated with Adolescent Idiopathic Scoliosis. *Am J Hum Genet* **97**, 337-42 (2015).
15. Sharma, S. *et al.* A PAX1 enhancer locus is associated with susceptibility to idiopathic scoliosis in females. *Nat Commun* **6**, 6452 (2015).
16. Smits, P. & Lefebvre, V. 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**, 1135-48 (2003).
17. Khanshour, A.M. *et al.* Genome-wide meta-analysis and replication studies in multiple ethnicities identify novel adolescent idiopathic scoliosis susceptibility loci. *Hum Mol Genet* **27**, 3986-3998 (2018).
18. Kou, I. *et al.* Genome-wide association study identifies 14 previously unreported susceptibility loci for adolescent idiopathic scoliosis in Japanese. *Nat Commun* **10**, 3685 (2019).

19. Haller, G. *et al.* A polygenic burden of rare variants across extracellular matrix genes among individuals with adolescent idiopathic scoliosis. *Hum Mol Genet* **25**, 202-9 (2016).
20. Baschal, E.E. *et al.* Exome sequencing identifies a rare HSPG2 variant associated with familial idiopathic scoliosis. *G3 (Bethesda)* **5**, 167-74 (2014).
21. Hayes, M. *et al.* ptk7 mutant zebrafish models of congenital and idiopathic scoliosis implicate dysregulated Wnt signalling in disease. *Nat Commun* **5**, 4777 (2014).
22. Van Gennip, J.L.M., Boswell, C.W. & Ciruna, B. Neuroinflammatory signals drive spinal curve formation in zebrafish models of idiopathic scoliosis. *Sci Adv* **4**, eaav1781 (2018).
23. Jaffe, K.M. *et al.* c21orf59/kurly Controls Both Cilia Motility and Polarization. *Cell Rep* **14**, 1841-9 (2016).
24. Becker-Heck, A. *et al.* The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. *Nat Genet* **43**, 79-84 (2011).
25. Bachmann-Gagescu, R. *et al.* The ciliopathy gene cc2d2a controls zebrafish photoreceptor outer segment development through a role in Rab8-dependent vesicle trafficking. *Hum Mol Genet* **20**, 4041-55 (2011).
26. Konjikusic, M.J. *et al.* Mutations in Kinesin family member 6 reveal specific role in ependymal cell ciliogenesis and human neurological development. *PLoS Genet* **14**, e1007817 (2018).
27. Rebello, D. *et al.* COL11A2 as a candidate gene for vertebral malformations and congenital scoliosis. *Hum Mol Genet* **32**, 2913-2928 (2023).
28. Long, F., Zhang, X.M., Karp, S., Yang, Y. & McMahon, A.P. Genetic manipulation of hedgehog signaling in the endochondral skeleton reveals a direct role in the regulation of chondrocyte proliferation. *Development* **128**, 5099-108 (2001).
29. Liu, Z. *et al.* Dysregulation of STAT3 signaling is associated with endplate-oriented herniations of the intervertebral disc in Adgrg6 mutant mice. *PLoS Genet* **15**, e1008096 (2019).
30. Liu, Z. *et al.* An adhesion G protein-coupled receptor is required in cartilaginous and dense connective tissues to maintain spine alignment. *Elife* **10**(2021).
31. Naba, A. *et al.* The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics* **11**, M111 014647 (2012).
32. Naba, A., Hoersch, S. & Hynes, R.O. Towards definition of an ECM parts list: an advance on GO categories. *Matrix Biol* **31**, 371-2 (2012).
33. Hynes, R.O. & Naba, A. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol* **4**, a004903 (2012).
34. Naba, A. *et al.* The extracellular matrix: Tools and insights for the "omics" era. *Matrix Biol* **49**, 10-24 (2016).
35. Fernandes, R.J., Weis, M., Scott, M.A., Seegmiller, R.E. & Eyre, D.R. Collagen XI chain misassembly in cartilage of the chondrodysplasia (cho) mouse. *Matrix Biol* **26**, 597-603 (2007).
36. Hafez, A. *et al.* Col11a1 Regulates Bone Microarchitecture during Embryonic Development. *J Dev Biol* **3**, 158-176 (2015).
37. Wenstrup, R.J. *et al.* Regulation of collagen fibril nucleation and initial fibril assembly involves coordinate interactions with collagens V and XI in developing tendon. *J Biol Chem* **286**, 20455-65 (2011).

38. Makki, N. *et al.* Genomic characterization of the adolescent idiopathic scoliosis associated transcriptome and regulome. *bioRxiv*, 2020.03.02.973735 (2020).
39. Gao, X. *et al.* CHD7 gene polymorphisms are associated with susceptibility to idiopathic scoliosis. *Am J Hum Genet* **80**, 957-65 (2007).
40. Sun, M. *et al.* Collagen XI regulates the acquisition of collagen fibril structure, organization and functional properties in tendon. *Matrix Biol* **94**, 77-94 (2020).
41. Chan, W.C., Au, T.Y., Tam, V., Cheah, K.S. & Chan, D. Coming together is a beginning: the making of an intervertebral disc. *Birth Defects Res C Embryo Today* **102**, 83-100 (2014).
42. Aszodi, A., Chan, D., Hunziker, E., Bateman, J.F. & Fassler, R. Collagen II is essential for the removal of the notochord and the formation of intervertebral discs. *J Cell Biol* **143**, 1399-412 (1998).
43. Smith, L.J., Nerurkar, N.L., Choi, K.S., Harfe, B.D. & Elliott, D.M. Degeneration and regeneration of the intervertebral disc: lessons from development. *Dis Model Mech* **4**, 31-41 (2011).
44. Wilm, B., Dahl, E., Peters, H., Balling, R. & Imai, K. Targeted disruption of Pax1 defines its null phenotype and proves haploinsufficiency. *Proc Natl Acad Sci U S A* **95**, 8692-7 (1998).
45. Wallin, J. *et al.* The role of Pax-1 in axial skeleton development. *Development* **120**, 1109-21 (1994).
46. Wu, Y.H., Chang, T.H., Huang, Y.F., Huang, H.D. & Chou, C.Y. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene* **33**, 3432-40 (2014).
47. Mudgett, J.S. *et al.* Susceptibility of stromelysin 1-deficient mice to collagen-induced arthritis and cartilage destruction. *Arthritis Rheum* **41**, 110-21 (1998).
48. Sun, S. *et al.* The active form of MMP-3 is a marker of synovial inflammation and cartilage turnover in inflammatory joint diseases. *BMC Musculoskelet Disord* **15**, 93 (2014).
49. Dy, P. *et al.* Sox9 directs hypertrophic maturation and blocks osteoblast differentiation of growth plate chondrocytes. *Dev Cell* **22**, 597-609 (2012).
50. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med* **13**, 31 (2021).
51. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15**, 901-13 (2005).
52. Lee, C.S. *et al.* Mutant collagen COL11A1 enhances cancerous invasion. *Oncogene* **40**, 6299-6307 (2021).
53. Zalewski, A., Cecchini, E.L. & Deroo, B.J. Expression of extracellular matrix components is disrupted in the immature and adult estrogen receptor beta-null mouse ovary. *PLoS One* **7**, e29937 (2012).
54. Wise, C.A., Sharma, S. Current Understanding of Genetic Factors in Idiopathic Scoliosis. in *The Genetics and Development of Scoliosis* (ed. Kusumi, K., Dunwoodie, S.) 167-190 (Springer, New York, 2010).
55. Yoshioka, H. *et al.* Developmental pattern of expression of the mouse alpha 1 (XI) collagen gene (Col11a1). *Dev Dyn* **204**, 41-7 (1995).
56. Smith, G.N., Jr., Hasty, K.A. & Brandt, K.D. Type XI collagen is associated with the chondrocyte surface in suspension culture. *Matrix* **9**, 186-92 (1989).

57. Luo YY, S.P., Kehlet SN,Karsdal MA. Type XI Collagen. in *Biochemistry of Collagens, Laminins and Elastin* (ed. MA, K.) 99-106 (Academic Press, London, UK, 2019).
58. Li, A., Wei, Y., Hung, C. & Vunjak-Novakovic, G. Chondrogenic properties of collagen type XI, a component of cartilage extracellular matrix. *Biomaterials* **173**, 47-57 (2018).
59. Long, J.T. *et al.* Hypertrophic chondrocytes serve as a reservoir for marrow-associated skeletal stem and progenitor cells, osteoblasts, and adipocytes during skeletal development. *Elife* **11**(2022).
60. Jiang, H., Yang, Q., Jiang, J., Zhan, X. & Xiao, Z. Association between COL11A1 (rs1337185) and ADAMTS5 (rs162509) gene polymorphisms and lumbar spine pathologies in Chinese Han population: an observational study. *BMJ Open* **7**, e015644 (2017).
61. Mio, F. *et al.* A functional polymorphism in COL11A1, which encodes the alpha 1 chain of type XI collagen, is associated with susceptibility to lumbar disc herniation. *Am J Hum Genet* **81**, 1271-7 (2007).
62. Styrkarsdottir, U. *et al.* GWAS of bone size yields twelve loci that also affect height, BMD, osteoarthritis or fractures. *Nat Commun* **10**, 2054 (2019).
63. Rodrigo, I., Hill, R.E., Balling, R., Munsterberg, A. & Imai, K. Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development* **130**, 473-82 (2003).
64. Sivakamasundari, V. *et al.* A developmental transcriptomic analysis of Pax1 and Pax9 in embryonic intervertebral disc development. *Biol Open* **6**, 187-199 (2017).
65. Liu, C.F. & Lefebvre, V. The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Res* **43**, 8183-203 (2015).
66. Costa, L. *et al.* Ossification and Fusion of the Vertebral Ring Apophysis as an Important Part of Spinal Maturation. *J Clin Med* **10**(2021).
67. Castelein, R.M., Pasha, S., Cheng, J.C. & Dubousset, J. Idiopathic Scoliosis as a Rotatory Decompensation of the Spine. *J Bone Miner Res* **35**, 1850-1857 (2020).
68. Sun, J. *et al.* A vertebral skeletal stem cell lineage driving metastasis. *Nature* **621**, 602-609 (2023).
69. Ortega, N., Behonick, D.J. & Werb, Z. Matrix remodeling during endochondral ossification. *Trends Cell Biol* **14**, 86-93 (2004).
70. Sellers, A. & Murphy, G. Collagenolytic enzymes and their naturally occurring inhibitors. *Int Rev Connect Tissue Res* **9**, 151-90 (1981).
71. Eguchi, T. *et al.* Novel transcription-factor-like function of human matrix metalloproteinase 3 regulating the CTGF/CCN2 gene. *Mol Cell Biol* **28**, 2391-413 (2008).
72. Cui, N., Hu, M. & Khalil, R.A. Biochemical and Biological Attributes of Matrix Metalloproteinases. *Prog Mol Biol Transl Sci* **147**, 1-73 (2017).
73. Xu, X., Li, Z., Leng, Y., Neu, C.P. & Calve, S. Knockdown of the pericellular matrix molecule perlecan lowers in situ cell and matrix stiffness in developing cartilage. *Dev Biol* **418**, 242-7 (2016).
74. Sharma, S. *et al.* Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. *Hum Mol Genet* **20**, 1456-66 (2011).
75. Fritsche, L.G. *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet* **48**, 134-43 (2016).

76. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
77. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
78. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
79. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
80. Landrum, M.J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062-D1067 (2018).
81. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894 (2019).
82. Davydov, E.V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* **6**, e1001025 (2010).
83. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* **49**, D344-D354 (2021).
84. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
85. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
86. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).
87. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).
88. Ameur, A. *et al.* SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. *Eur J Hum Genet* **25**, 1253-1260 (2017).
89. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
90. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
91. de Klerk, J.B. *et al.* A patient with mevalonic aciduria presenting with hepatosplenomegaly, congenital anaemia, thrombocytopenia and leukocytosis. *J Inherit Metab Dis* **11 Suppl 2**, 233-6 (1988).
92. Song, Y.Q. *et al.* Lumbar disc degeneration is linked to a carbohydrate sulfotransferase 3 variant. *J Clin Invest* **123**, 4909-17 (2013).
93. Marees, A.T. *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res* **27**, e1608 (2018).
94. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2011).
95. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
96. Makki, N. *et al.* Genomic characterization of the adolescent idiopathic scoliosis-associated transcriptome and regulome. *Hum Mol Genet* **29**, 3606-3615 (2021).

97. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-9 (2015).
98. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
99. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-7 (2016).
100. Gurumurthy, C.B. *et al.* Creation of CRISPR-based germline-genome-engineered mice without ex vivo handling of zygotes by i-GONAD. *Nat Protoc* **14**, 2452-2482 (2019).
101. Ushiki, A. *et al.* Deletion of CTCF sites in the SHH locus alters enhancer-promoter interactions and leads to acheiropodia. *Nat Commun* **12**, 2282 (2021).
102. Yu, H., Harrison, F.E. & Xia, F. Altered DNA repair; an early pathogenic pathway in Alzheimer's disease and obesity. *Sci Rep* **8**, 5600 (2018).
103. Jha, K.K., Banga, S., Palejwala, V. & Ozer, H.L. SV40-Mediated immortalization. *Exp Cell Res* **245**, 1-7 (1998).
104. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).

Appendix A. Key Resources Table

siRNA and Primer Sequences

Mouse Esr2 siRNA	CAAGUGUUACGAAGUAGGAdT
Mouse Col11a1 siRNA	GAAAGAAGGUGCAAAGGGUdT
Mouse Mmp3 F	CTCTGGAACCTGAGACATCACC
Mouse Mmp3 R	AGGAGTCCTGAGAGATTTGCGC
Mouse Col11a1 F	AGGAGAGTTGAGAATTGGGAATC
Mouse Col11a1 R	TGGTGATCAGAATCAGAAGTT
Mouse Col11a2 F	CTCATCTTCCTGCATCAGAC
Mouse Col11a2 R	ACTTGGAAAGCGAGGTCCT
Mouse Adgrg6 F	AGAGGATGGACTGAGGCTGTGT
Mouse Adgrg6 R	CCAGGCTTGTTTGGACATGGTTG
Mouse Sox6 F	GCATAAGTGACCGTTTTGGCAGG
Mouse Sox6 R	GGCATCTTTGCTCCAGGTGACA
Mouse Mmp14 F	GCCTTCTGTTCCTGATAA
Mouse Mmp14 R	CCATCCTTCCTCTCGTAG
Mouse Pax1 F	AACCAGCACGGAGTATACAGC
Mouse Pax1 R	TGTAAGCTACCGAGTGCATCC
Mouse Esr2 F	GGTCCTGTGAAGGATGTAAGGC
Mouse Esr2 R	TAACACTTGCGAAGTCGGCAGG
Mouse Gapdh F	CATCACTGCCACCCAGAAGACTG
Mouse Gapdh R	ATGCCAGTGAGCTTCCCGTTCAG
Rat Sfrp2 F	CGTGAAACGGTGGCAGAAG
Rat Sfrp2 R	CGGATGCTGCGGGAGAT
Rat Krt19 F	AAGACACACTGGCAGAAACG
Rat Krt19 R	GATTCTGCCGCTCACTATCA
Rat Mmp12 F	TTGGCCATTCCTTGGGGCTGC
Rat Mmp12 R	TGTTGGTGGCTGGACTCCCAGG
Mouse Pax1 F (Fig. 5)	CCGCACATTCAGTCAGCAAC
Mouse Pax1 R (Fig. 5)	CATCTTGGGGGAGTAGGCAG
Mouse Col11a1 F (Fig. 5)	CACAAAACCCCTCGATAGAAGTG
Mouse Col11a1 R (Fig. 5)	CCTGTGATCAGGAACTGCTGAA
Mouse Adgrg6 F (Fig. 5)	TCCTGTCCATCTCTGGCTCA
Mouse Adgrg6 R (Fig. 5)	CACAAGACAGAGCTGCTCCA
Mouse Sox6 F (Fig. 5)	TGCGACAGTTCTTCACTGTGG
Mouse Sox6 R (Fig. 5)	CGTCCATCTTCATACCATACG
Mouse β -Actin F (Fig. 5)	GGCACCACACCTTCTACAATG
Mouse β -Actin R (Fig. 5)	GGGGTGTTGAAGGTCTCAAAC
Pax1-genotyping F	CAGAACCTGGAATGCTGTGCTC

Pax1-genotyping R	AAAGGGTTGCAGTGCCTTCAC

Appendix B. Clinical groups

Texas Scottish Rite Hospital for Children Clinical Group (TSRHCCG)

Lori A. Karol[1], Karl E. Rathjen[1], Daniel J. Sucato[1], John G. Birch[1], Charles E. Johnston III[1], Benjamin S. Richards[1], Brandon Ramo[1], Amy L. McIntosh[1], John A. Herring[1], Todd A. Milbrandt[2], Vishwas R. Talwakar[3], Henry J. Iwinski[42], Ryan D. Muchow[3], J. Channing Tassone[4], X. -C. Liu[4], Richard Shindell[5], William Schrader[6], Craig Eberson[7], Anthony Lapinsky[8], Randall Loder[9] and Joseph Davey[10]

1. Department of Orthopaedic Surgery, Texas Scottish Rite Hospital for Children, Dallas, Texas, USA.

2. Department of Orthopaedic Surgery, Mayo Clinic, Rochester, Minnesota, USA

3. Department of Orthopaedic Surgery, Shriners Hospitals for Children, Lexington, Kentucky, USA.

4. Department of Orthopaedic Surgery, Children's Hospital of Wisconsin, Milwaukee, Wisconsin, USA.

5. OrthoArizona, Phoenix, Arizona, USA.

6. Departments of Orthopedics, Sports Medicine, and Surgical Services, Akron Children's Hospital, Akron, Ohio, USA.

7. Pediatric Orthopaedics and Scoliosis, Hasbro Children's Hospital, Providence, Rhode Island, USA.

8. University of Massachusetts Memorial Medical Center, Worcester, Massachusetts, USA.

9. Indiana University-Purdue University Indianapolis, Indianapolis, Indiana, USA.

10. University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA.

Japan Scoliosis Clinical Research Group (JSCRG)

Kota Watanabe[1], Nao Otomo[1,2], Kazuki Takeda[1,2], Yoshiro Yonezawa[1,2], Yoji Ogura[1,2], Yohei Takahashi[1,2], Noriaki Kawakami[3], Taichi Tsuji[4], Koki Uno[5], Teppei Suzuki[5], Manabu Ito[6], Shohei Minami[7], Toshiaki Kotani[7], Tsuyoshi Sakuma[7], Haruhisa Yanagida[8], Hiroshi Taneichi[9], Satoshi Inami[9], Ikuho Yonezawa[10], Hideki Sudo[11], Kazuhiro Chiba[12], Naobumi Hosogane[12], Kotaro Nishida[13], Kenichiro Kakutani[13], Tsutomu Akazawa[14], Takashi Kaito[15], Kei Watanabe[16], Katsumi Harimaya[17], Yuki Taniguchi[18], Hideki Shigemats[19], Satoru Demura[20], Takahiro Iida[21], Ryo Sugawara[22], Katsuki Kono[23], Masahiko Takahata[24], Norimasa Iwasaki[24], Eijiro Okada[1], Nobuyuki Fujita[1], Mitsuru Yagi[1], Masaya Nakamura[1], Morio Matsumoto[1]

1. Department of Orthopaedic Surgery, Keio University School of Medicine, Tokyo, Japan

2. Laboratory of Bone and Joint Diseases, Center for Integrative Medical Sciences, RIKEN, Tokyo, Japan

3. Department of Orthopaedic Surgery, Meijo Hospital, Nagoya, Japan

4. Department of Orthopaedic Surgery, Toyota Kosei Hospital, Nagoya, Japan

5. Department of Orthopaedic Surgery, National Hospital Organization, Kobe Medical Center, Kobe, Japan

6. Department of Orthopaedic Surgery, National Hospital Organization, Hokkaido Medical Center, Sapporo, Japan

7. Department of Orthopaedic Surgery, Seirei Sakura Citizen Hospital, Sakura, Japan

8. Department of Orthopaedic Surgery, Fukuoka Children's Hospital, Fukuoka, Japan

9. Department of Orthopaedic Surgery, Dokkyo Medical University School of Medicine, Mibu, Japan

10. Department of Orthopaedic Surgery, Juntendo University School of Medicine, Tokyo, Japan

11. Department of Advanced Medicine for Spine and Spinal Cord Disorders, Hokkaido University Graduate School of Medicine, Sapporo, Japan

12. Department of Orthopaedic Surgery, National Defense Medical College, Tokorozawa, Japan

13. Department of Orthopaedic Surgery, Kobe University Graduate School of Medicine, Kobe, Japan

14. Department of Orthopaedic Surgery, St. Marianna University School of Medicine, Kawasaki, Japan

15. Department of Orthopaedic Surgery, Osaka University Graduate School of Medicine, Suita, Japan

16. Department of Orthopaedic Surgery, Niigata University Hospital, Niigata, Japan

17. Department of Orthopaedic Surgery, Graduate School of Medical Sciences, Kyushu University Beppu Hospital, Fukuoka, Japan

18. Department of Orthopaedic Surgery, Faculty of Medicine, The University of Tokyo, Tokyo, Japan

19. Department of Orthopaedic Surgery, Nara Medical University, Nara, Japan

20. Department of Orthopaedic Surgery, Kanazawa University School of Medicine, Kanazawa, Japan

21. Department of Orthopaedic Surgery, Dokkyo Medical University Koshigaya Hospital, Koshigaya, Japan

22. Department of Orthopaedic Surgery, Jichi Medical University, Simotsuke, Japan

23. Department of Orthopaedic Surgery, Kono Othopaedic Clinic, Tokyo, Japan
24. Department of Orthopaedic Surgery, Hokkaido University, Sapporo, Japan

Scoliosis and Genetics in Scandinavia (ScoliGeneS) study group

Tian Cheng[1], Juha Kere[2], Aina Danielsson[3], Kristina Åkesson[4], Ane Simony[5], Mikkel Andersen[5], Steen Bach Christensen[5], Maria Wikzén[6], Luigi Belcastro[6]

1. Department of Clinical Sciences and Technology, Karolinska Institutet, Huddinge, Sweden

2. Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

3. Department of Orthopedics, Sahlgrenska University Hospital, Gothenburg, Sweden

4. Department of Orthopedics and Clinical Sciences, Lund University, Skane University Hospital, Malmö, Sweden

5. Sector for Spine Surgery and Research, Middelfart Hospital, Middelfart, Denmark

6. Department of Reconstructive Orthopaedics, Karolinska University Hospital, Stockholm, Sweden.

Cohort	Ethnicity	Stage	Subjects	Cases		Controls	
				Male	Female	Male	Female
USA (TX)	NHW	Discovery	13,865	201	1,157	5,369	7,138
USA (MO)	NHW	Replication	2,951	201	1,012	1,049	689
SW-D	NHW	Replication	4,627	222	1,409	505	2,491
JP	EAS (Japanese)	Replication	79,211	323	5,004	40,205	33,679
HK	EAS (Han Chinese)	Replication	3,103	178	812	858	1,255
Total			103,757	10,519		93,238	

Table 1. Study cohorts. USA (TX): Texas cohort, USA (MO): Missouri cohort, SW-D: Swedish-Danish cohort, JP: Japanese cohort, HK: Hong Kong cohort, NHW: Non-Hispanic White, EAS: East Asian.

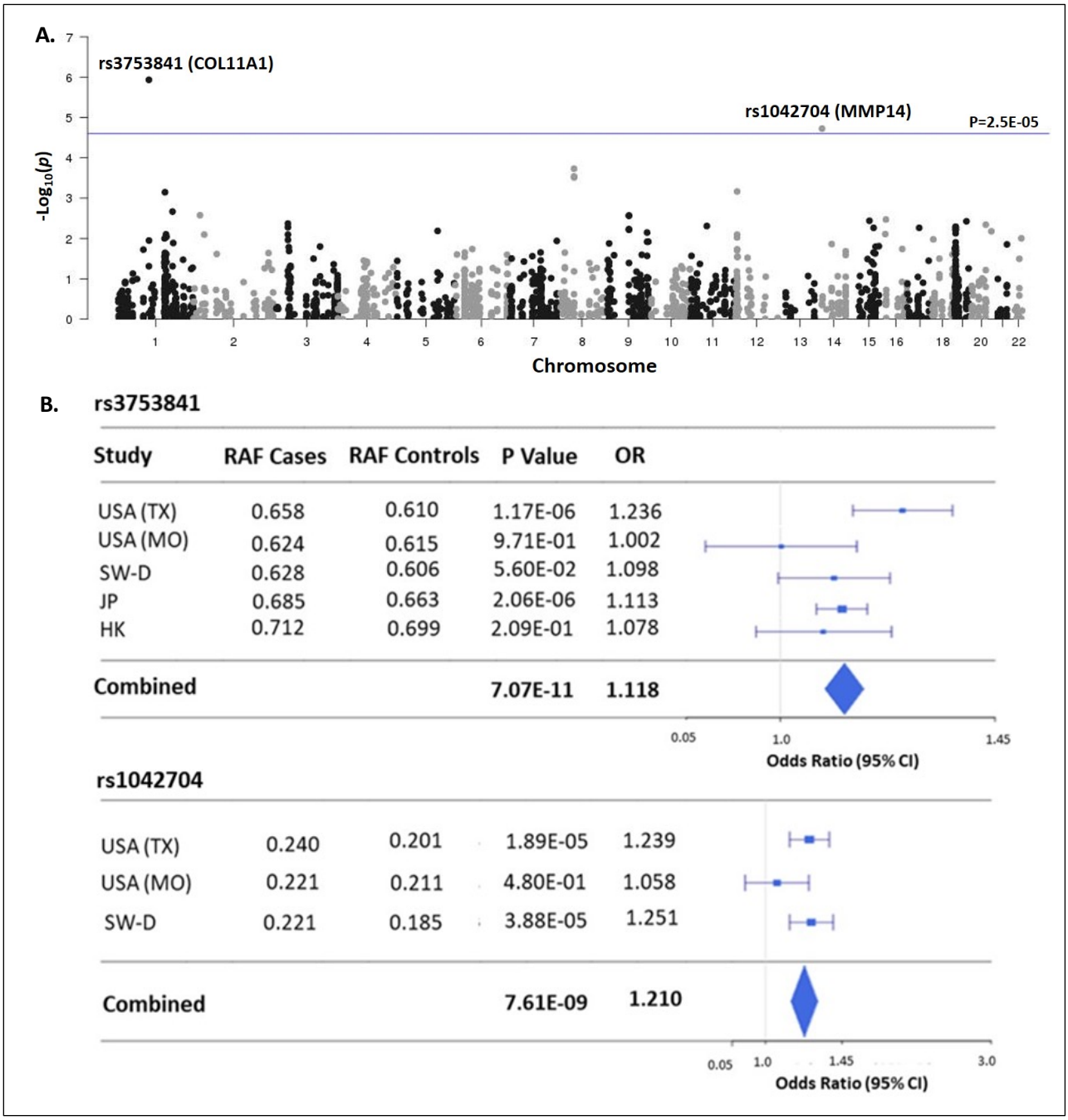
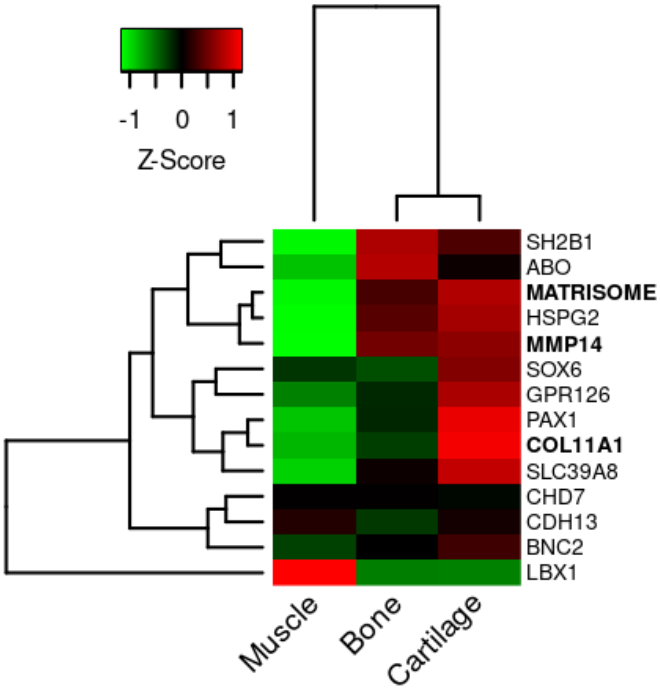
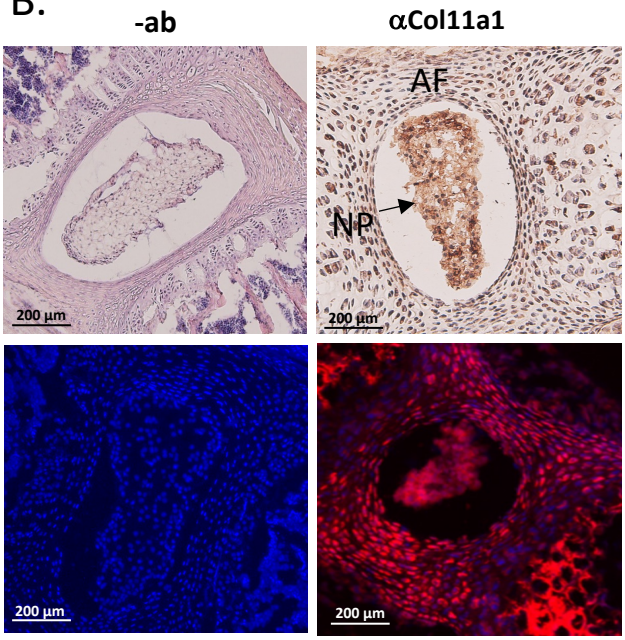


Figure 1. Matrisome-wide association study. a. Manhattan plot showing –log10 P-values (y axis) versus chromosomal position (x axis) for the 2,008 common coding variants tested in the discovery study USA (TX). The horizontal line represents the threshold for significance level (P-value<2.5×10-5) after Bonferroni multiple testing correction. **b.** Tests of association for SNPs rs3753841 and rs1042704 in discovery and independent replication cohorts. RAF - reference allele frequency; OR - odds ratio; CI-confidence interval.

A.


-1
0
1
Z-Score
SH2B1
ABO
MATRISOME
HSPG2
MMP14
SOX6
GPR126
PAX1
COL11A1
SLC39A8
CHD7
CDH13
BNC2
LBX1
Muscle
Bone
Cartilage

B.


-ab
αCol11a1
AF
NP
200 µm
200 µm
200 µm
200 µm

C.

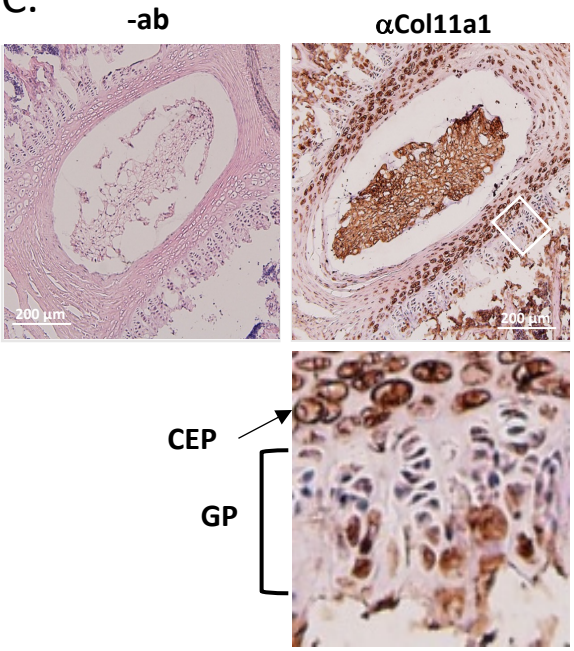

-ab
αCol11a1
200 µm
200 µm
CEP
GP

Figure 2. *Col11a1* and *Mmp14* expression in spine.

a) A heatmap of transcript per million (TPM) values of *COL11A1*, *MMP14*, and other published genes associated with AIS. The average TPM value of matrisome genes is represented as MATRISOME.

b) Detection of collagen type XI alpha chain 1 in P0.5 mouse spine. Immunohistochemistry (IHC) shown at top, with immunofluorescence (IF) staining below. Negative controls lacking primary antibody are shown at left. Results are representative of N≥3 technical replicates in whole spines.

c) Detection of collagen α1(XI) in P28 mouse spine. Negative antibody IHC control shown at left; antibody positive IHC shown at right. Enlarged, rotated view of white boxed area shows a biphasic staining pattern. CEP – cartilage endplate; GP – growth plate. Results are representative of N≥3 technical replicates in whole spines.

A.

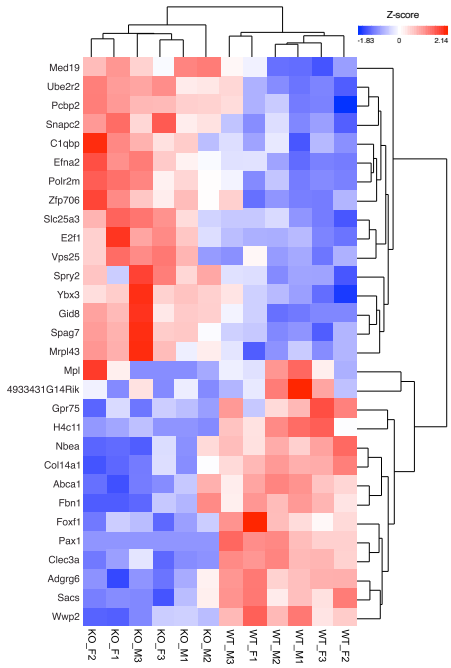
WT - ab

WT -+ab

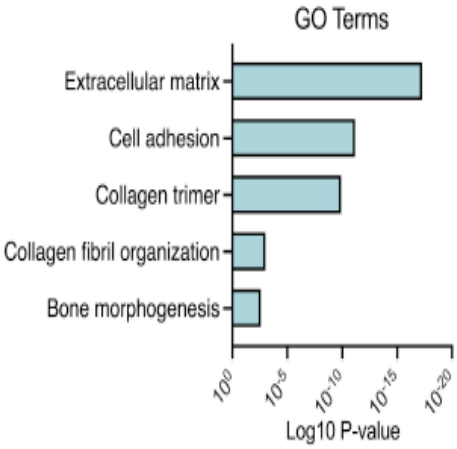
PAX1 KO

PAX1 Col11a1 DAPI

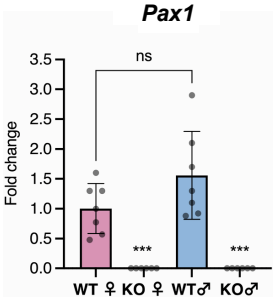
B.



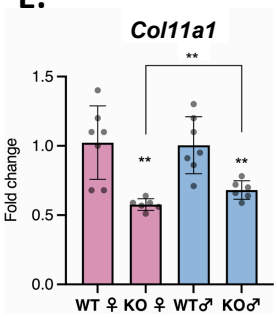
C.



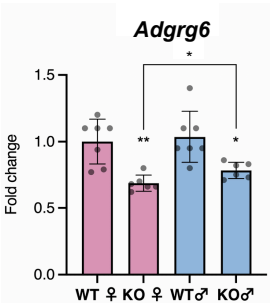
D.



E.



F.



G.

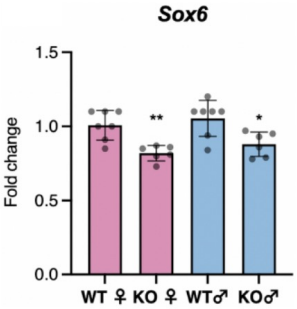


Figure 3. Assessing *Pax1* regulation of *Col11a1*.

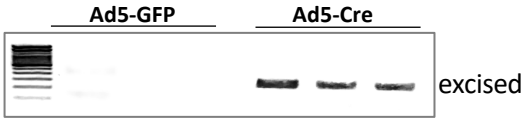
a) IF staining of P28 IVD from thoracic regions of *Pax1*$^{-/-}$ (bottom and wildtype littermate (middle, top) mice using PAX1- (green) and collagen α1(XI)- specific (red) antibodies and DAPI nuclear counterstain. Antibody-negative controls are shown at top. Results are representative of N≥3 technical replicates in whole spines.

b) Heatmap of differentially expressed genes (*P*-value<0.0001) in E12.5 tails of WT and *Pax1*$^{-/-}$ *mice*

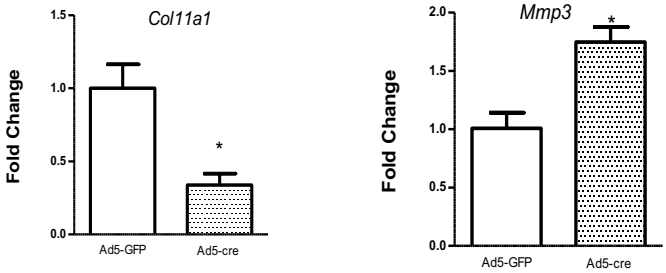
c) Gene ontology (GO) analysis of differentially expressed genes in E12.5 tail WT and *Pax1*$^{-/-}$ null mice

d-g) Gene expression levels dissected from E12.5 mouse tail from wild type (WT) and *Pax1*$^{-/-}$ (KO) mice as determined by qRT-PCR. Each value represents the ratio of each gene expression to that of β-actin, and values are mean ± standard deviation. The expression value of WT female group was arbitrarily set at 1.0. Each dot represents one embryo and statistical differences were determined using a two-sided unpaired t-test (*P**<0.05, *P***<0.01, *P****<0.001).

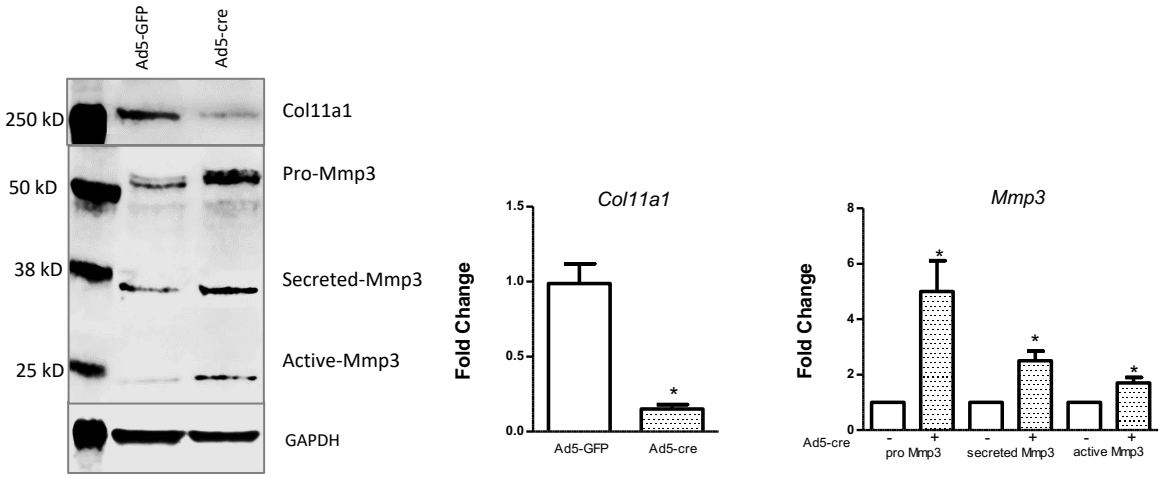
A.


Ad5-GFP
Ad5-Cre
excised

B.


Col11a1
Fold Change
1.5
1.0
0.5
0.0
*
Ad5-GFP
Ad5-cre
Mmp3
Fold Change
2.0
1.5
1.0
0.5
0.0
*
Ad5-GFP
Ad5-cre

C.


Ad5-GFP
Ad5-cre
250 kD
Col11a1
50 kD
Pro-Mmp3
38 kD
Secreted-Mmp3
25 kD
Active-Mmp3
GAPDH
Col11a1
Fold Change
1.5
1.0
0.5
0.0
*
Ad5-GFP
Ad5-cre
Mmp3
Fold Change
8
6
4
2
0
*
*
*
Ad5-cre
-
+
-
+
-
+
pro Mmp3
secreted Mmp3
active Mmp3

D.

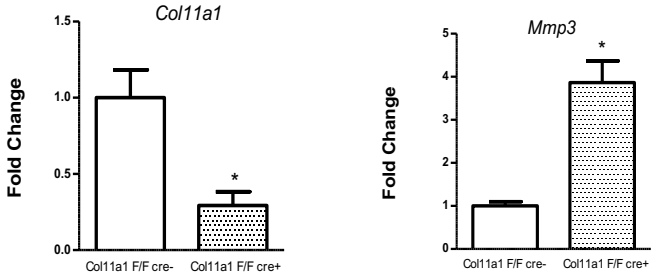

Col11a1
Fold Change
1.5
1.0
0.5
0.0
*
Col11a1 F/F cre-
Col11a1 F/F cre+
Mmp3
Fold Change
5
4
3
2
1
0
*
Col11a1 F/F cre-
Col11a1 F/F cre+

Figure 4. *Col11a1* regulation of *Mmp3* expression in cartilage.

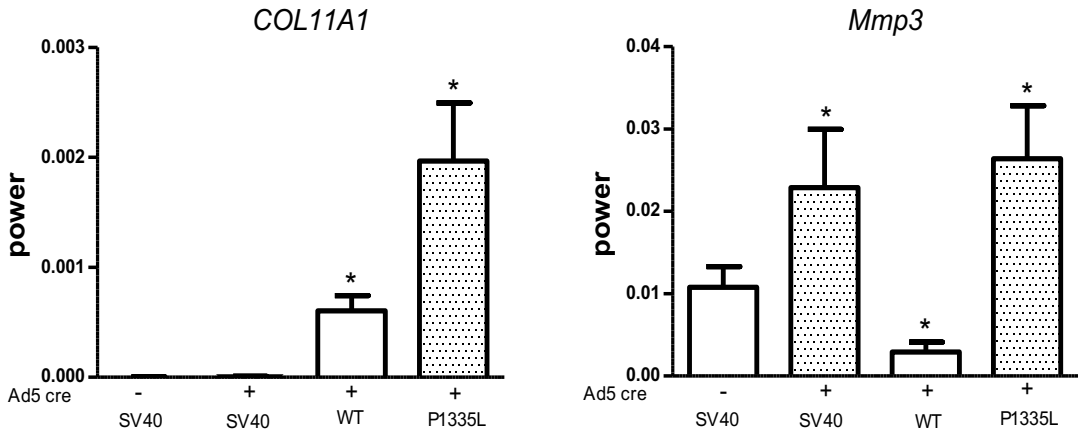
a) PCR assay of *Col11a1* excision in *Col11a1*$^{fl/fl}$ cultured costal chondrocytes.

b) Gene expression levels from *Col11a1*$^{fl/fl}$ cultured costal chondrocytes transduced with GFP (Ad5-GFP, left) or Cre-expressing adenovirus (Ad5-cre, right) as determined by qRT-PCR. Each value represents the ratio of each gene expression to that of *GAPDH*, and values are mean ± standard deviation. The expression value of control Ad5-GFP results was arbitrarily set at 1.0. Statistical differences were determined using a two-sided unpaired t test ($P^*<0.05$). Results shown for N≥3 biologic replicates, each including 3 technical replicates.

c) Western blot detection of collagen α1(XI), MMP3, and GAPDH loading control in cultured costal chondrocytes after Ad5-GFP or Ad5-cre transduction. Results are representative of N= 4 biologic replicates. Protein size ladder is shown in lane 1. Quantification of bands detected by Western blotting, where scramble results were set to 1.0, is shown at right.

d) Gene expression levels from dissected *Col11a1*$^{fl/fl}$ *ATC* costal cartilage, analyzed as described in a). Results shown for N=3 biologic replicates, each including 3 technical replicates.

A.


COL11A1
Mmp3
power
0.003
0.002
0.001
0.000
0.04
0.03
0.02
0.01
0.00
*
Ad5 cre
-
+
SV40
SV40
WT
P1335L

B.

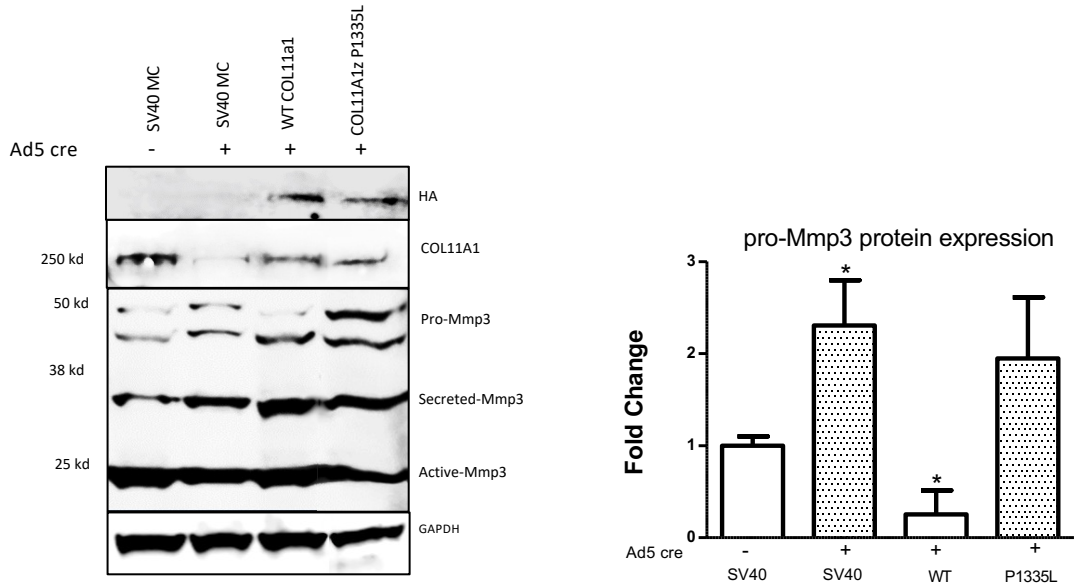
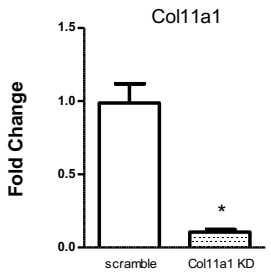

SV40 MC
SV40 MC
WT COL11a1
COL11A1z P1335L
Ad5 cre
-
+
HA
COL11A1
250 kd
50 kd
Pro-Mmp3
38 kd
Secreted-Mmp3
25 kd
Active-Mmp3
GAPDH
pro-Mmp3 protein expression
Fold Change
3
2
1
0
*
SV40
WT
P1335L

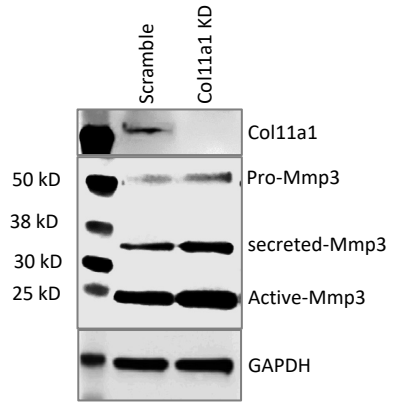
Figure 5. *Col11a1*^{P1335L} regulation of *Mmp3* expression in lentiviral transduced mouse GPCs.

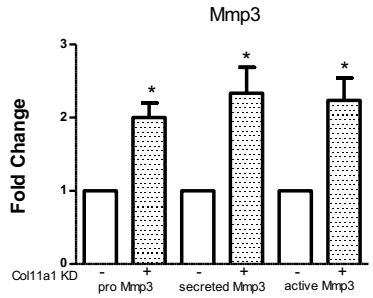
a) qRT-PCR of human *COL11A1* and endogenous mouse *Mmp3*. "SV40" refers to immortalized mouse costal chondrocytes transduced with the lentiviral vector only. Significant fold changes ($P \leq 0.05$) relative to vector-only transfected cells are shown by *. Results shown for N=4 biologic replicates, each including 3 technical replicates.

b) Western blot corresponding to experiments shown in (a) using HA antibody to detect epitope-tagged human collagen α1(XI), COL11A1 antibody to detect mouse and human collagen α1(XI), MMP3 antibody to detect endogenous mouse MMP3, and GAPDH. Quantification at right shows MMP3 expression in each experiment (of N=4) relative to GAPDH.

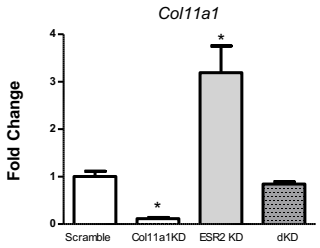
A.

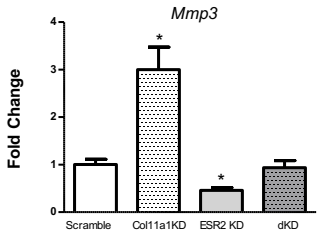

Col11a1
Fold Change
1.5
1.0
0.5
0.0
*
scramble
Col11a1 KD

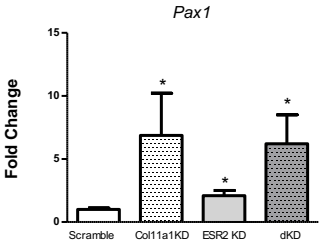

Scramble
Col11a1 KD
Col11a1
50 kD
Pro-Mmp3
38 kD
secreted-Mmp3
30 kD
25 kD
Active-Mmp3
GAPDH

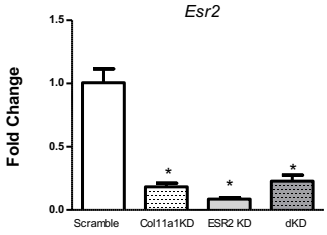

Mmp3
Fold Change
3
2
1
0
*
*
*
Col11a1 KD
-
+
-
+
-
+
pro Mmp3
secreted Mmp3
active Mmp3

B.

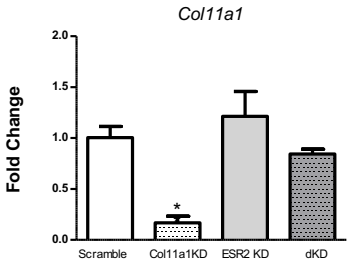

Col11a1
Fold Change
4
3
2
1
0
*
*
Scramble
Col11a1KD
ESR2 KD
dKD

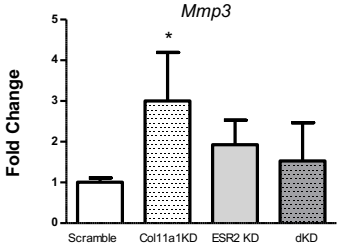

Mmp3
Fold Change
4
3
2
1
0
*
*
Scramble
Col11a1KD
ESR2 KD
dKD


Pax1
Fold Change
15
10
5
0
*
*
*
Scramble
Col11a1KD
ESR2 KD
dKD


Esr2
Fold Change
1.5
1.0
0.5
0.0
*
*
*
Scramble
Col11a1KD
ESR2 KD
dKD

C.


Col11a1
Fold Change
2.0
1.5
1.0
0.5
0.0
*
Scramble
Col11a1KD
ESR2 KD
dKD


Mmp3
Fold Change
5
4
3
2
1
0
*
Scramble
Col11a1KD
ESR2 KD
dKD

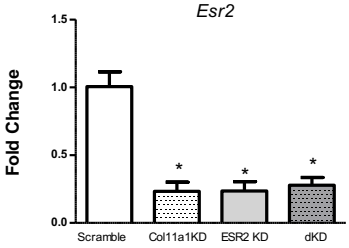

Esr2
Fold Change
1.5
1.0
0.5
0.0
*
*
*
Scramble
Col11a1KD
ESR2 KD
dKD

Figure 6. Effects of estrogen receptor beta on *Col11a1-Mmp3* signaling axis.

a) Representative Western blot (of N=4 biologic replicates) of cultured costal chondrocytes after scramble or *Col11a1*-specific siRNA knockdown. Protein size ladder is shown in lane 1. Quantification of bands detected by Western blotting are shown at right, where scramble results were set to 1.0.

b) Gene expression levels of *Col11a1* (left), *Mmp3* (middle), and *Esr2* (right) mRNA in cultured costal chondrocytes showing fold change relative to the scramble control. dKD = double *Col11a1-Esr2*-specific siRNA knockdowns. Results are representative of N≥3 biologic replicates, each including 3 technical replicates.

c) Gene expression levels from rat CEP cells, as described in (b).

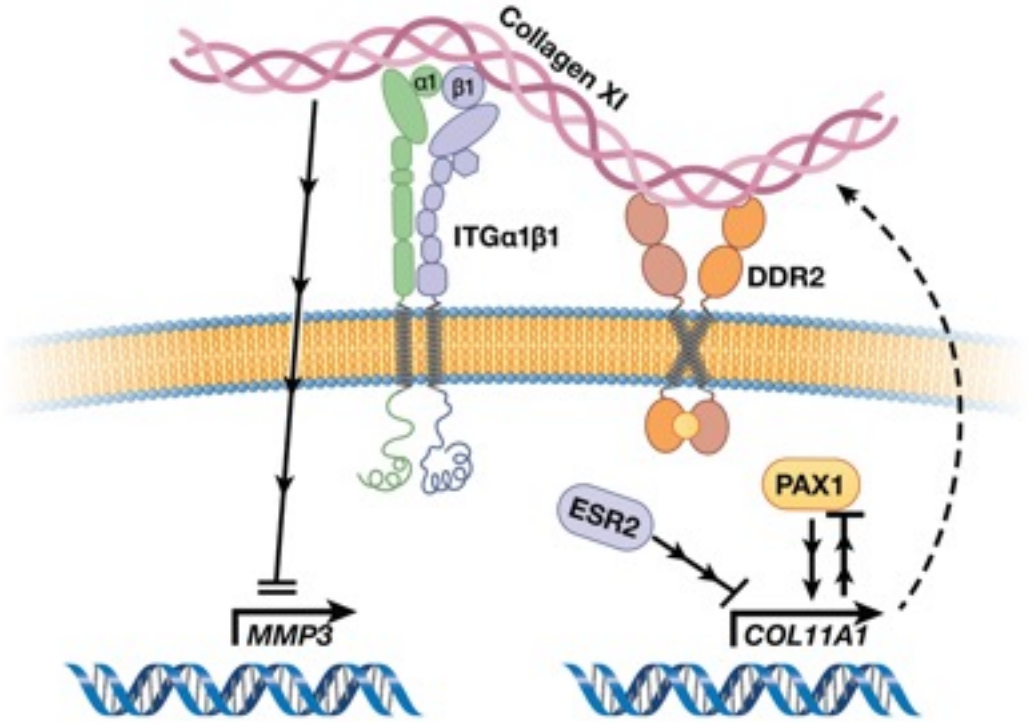

Collagen XI
α1
β1
ITGα1β1
DDR2
PAX1
ESR2
MMP3
COL11A1

Figure 7. Cartoon depiction of a collagen XI-mediated signaling axis in chondrocytes. Collagen XI is held in the pericellular space by integrins and DDR2. COL11A1, under the regulation of ESR2 and PAX1, signals through unknown mechanisms and inhibits MMP3 transcription.