Evaluation of IRX Genes and Conserved Noncoding Elements in a Region on 5p13.3 Linked to Families with Familial Idiopathic Scoliosis and Kyphosis

Cristina M. Justice,*1 Kevin Bishop,† Blake Carrington,† Jim C. Mullikin,† Kandice Swindle,§ Beth Marosy,** Raman Sood,† Nancy H. Miller,† and Alexander F. Wilson*

*Genomics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, and **McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21224, †Zebrafish Core, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, ‡National Institutes of Health Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville, Maryland 20852, and §Department of Orthopedic Surgery, University of Colorado Anschutz Medical Campus, Aurora, Colorado 80045

ORCID ID: 0000-0002-6190-6821 (C.M.J.)

ABSTRACT Because of genetic heterogeneity present in idiopathic scoliosis, we previously defined clinical subsets (a priori) from a sample of families with idiopathic scoliosis to find genes involved with spinal curvature. Previous genome-wide linkage analysis of seven families with at least two individuals with kyphoscoliosis found linkage (P-value = 0.002) in a 3.5-Mb region on 5p13.3 containing only three known genes, IRX1, IRX2, and IRX4. In this study, the exons of IRX1, IRX2, and IRX4, the conserved noncoding elements in the region, and the exons of a nonprotein coding RNA, LOC285577, were sequenced. No functional sequence variants were identified. An intrafamilial test of association found several associated noncoding single nucleotide variants. The strongest association was with rs12517904 (P = 0.00004), located 6.5 kb downstream from IRX1. In one family, the genotypes of nine variants differed from the reference allele in all individuals with kyphoscoliosis, and two of three individuals with scoliosis, but did not differ from the reference allele in all other genotyped individuals. One of these variants, rs117273909, was located in a conserved noncoding region that functions as an enhancer in mice. To test whether the variant allele at rs117273909 had an effect on enhancer activity, zebrafish transgenesis was performed with overlapping fragments of 198 and 687 bp containing either the wild type or the variant allele. Our data suggests that this region acts as a regulatory element; however, its size and target gene(s) need to be identified to determine its role in idiopathic scoliosis.

KEYWORDS idiopathic scoliosis IRX genes zebrafish transgenesis conserved noncoding regions kyphoscoliosis

Idiopathic scoliosis is defined as a lateral curvature of the spine greater than ten degrees (°) documented by radiographic analysis, and present in the late juvenile or adolescent period in otherwise normal individuals. The prevalence of idiopathic scoliosis in the general population is estimated to be 2–3% (Bunnell 1986; Lonstein and Carlson 1984). Both sporadic [idiopathic scoliosis (IS)] and familial forms [familial idiopathic scoliosis (FIS)] exist. Several genetic analyses have reported linkage of FIS to various candidate regions, including chromosomes 6p (LOD = 1.42), 10q (LOD = 1.60), and 18q (ATA82B02, LOD = 8.26) (Wise et al. 2000); 17p11.2 (D17S799, LOD = 3.2) (Salehi et al. 2002); 19p13.3 (D19S922, LOD = 4.087), and 2q (LOD = 1.72) (Chan et al. 2002); Xq23-26 (GATA172D05, LOD = 2.23) (Justice et al. 2003); 6p25-22 (D6S1031, P-value = 0.0032), 6q14-16 (D6S1031, P-value = 0.0092), 9q32-34 (D9S915, P-value = 0.0005), 16q11-q12 (D16S2623, P-value = 0.0005) and 17p11-q11 (D16S2623, P-value = 0.0005) (Miller et al. 2005); 8q12 (D8S1136, LOD = 2.77) (Gao et al. 2007); 9q31.2-q34.2 (D9S2157, LOD = 3.64) and 17q25.3-qtel

1Corresponding author: Genomics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, 333 Casell Drive, Suite 1200, Baltimore, MD 21224. E-mail: cmj@mail.nih.gov

doi: 10.1534/g3.116.029975
Manuscript received November 4, 2015; accepted for publication April 5, 2016; published Early Online April 12, 2016.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Supplemental material is available online at www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.029975/-/DC1
extension) homeobox genes to identify significantly associated SNVs. In addition, a SNV in a highly conserved element (CNE) surrounding this SNP, rs117273909, acts as an enhancer in vivo, and if the alternate allele affects the enhancer function.

**MATERIALS AND METHODS**

**Subjects**

All probands and their relatives were clinically characterized by a single orthopedic surgeon. Written informed consent was obtained for all study participants, in accordance with the Institutional Review Board of the participating institutions.

**Primer design for genes and conserved noncoding elements**

Primers covering the IRX exons genes, the CNEs, and a long intergenic nonprotein RNA in the linkage region, LOC285577, were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). CNEs were selected for sequencing if they had a LOD score > 100 based on the PhastCons Placental Mammal Conserved Elements, 28-way Multiz Alignment (http://www.genome.ucsc.edu), and 500 kb downstream from IRX1, IRX2, and IRX4 (Supplemental Material, Figure S1 and Table S1). For rs117273909, genotyping of the controls was performed by sequencing. Sequences of primers are available on request.

**PCR and sequencing**

For mutational analysis of the genes, PCR was performed on DNA from 53 individuals from seven FIS families from the KS clinical subset using

| SNP          | Positiona | MAFb | Allelic Association | Genotypic Association |
|--------------|-----------|------|---------------------|-----------------------|
| rs67250895   | 2303160   | 0.22 | 0.03124             | 0.14792               |
| rs139215365  | 2427684   | 0.07 | 0.25035             | 0.03999               |
| rs16870466   | 2799648   | 0.11 | 0.04946             | 0.05745               |
| rs1497457    | 2799759   | 0.31 | 0.03285             | 0.19374               |
| rs2934527    | 2947498   | 0.48 | 0.04338             | 0.09179               |
| rs2934528    | 2947877   | 0.32 | 0.03903             | 0.07172               |
| rs182225473  | 3045672   | 0.04 | 0.03913             | 0.03913               |
| rs14651224   | 3046004   | 0.20 | 0.03438             | 0.03438               |
| rs117273909  | 3182971   | 0.44 | 0.03913             | 0.03913               |
| rs111916055  | 3187995   | 0.03 | 0.03438             | 0.03438               |
| rs16871553   | 3266026   | 0.39 | 0.02924             | 0.02924               |
| rs73733752   | 3278780   | 0.54 | 0.03913             | 0.03913               |
| rs61712864   | 3317668   | 0.18 | 0.04626             | 0.25960               |
| rs73733769   | 3326276   | 0.03 | 0.02653             | 0.02653               |
| rs73032754   | 3428448   | 0.04 | 0.00948             | 0.00965               |
| novel        |           |      |                     |                       |
| rs117494736  | 3459291   | 0.10 | 0.02326             | 0.02326               |
| rs53739844   | 3491460   | 0.03 | 0.03893             | 0.03893               |
| rs35450818   | 3491486   | 0.03 | 0.00737             | 0.00737               |
| rs34560950   | 3512343   | 0.50 | 0.00737             | 0.00737               |
| rs62336074   | 3513171   | 0.47 | 0.28388             | 0.28388               |
| rs35155570   | 3518273   | 0.03 | 0.02409             | 0.02409               |
| rs828332     | 3591325   | 0.03 | 0.00737             | 0.00737               |
| rs12517904   | 3608064   | 0.03 | 0.00004             | 0.16992               |
| rs76205392   | 3617948   | 0.03 | 0.03893             | 0.03893               |
| rs71577554   | 3618245   | 0.03 | 0.02686             | 0.02686               |
| rs10475220   | 3618387   | 0.22 | 0.02686             | 0.02686               |
| rs10475221   | 3618393   | 0.03 | 0.02686             | 0.02686               |
| rs78040936   | 3630480   | 0.20 | 0.02686             | 0.02686               |
| novel        | 4003967   | 0.04 | 0.02645             | 0.02645               |

a Map positions obtained from NCBI (GRCh37/hg19).
b Maximum likelihood estimate of minor allele frequency from founders (FREQ [S.A.G.E., v6.0.1]).

(rs117273909, acts as an enhancer in vivo, and if the alternate allele affects the enhancer function.

Table 1 Association analysis results (P-values < 0.05)
the HotStarTaq amplification protocol (Qiagen). For the sequencing of CNEs and LOC285577, PCR from 46 of these individuals was carried out using the KAPA 2G Fast HS ReadyMix PCR Kit (KAPA Biosystems, Wilmington, MA). The reactions were analyzed on 3730 DNA Sequencers (Applied Biosystems, Grand Island, NY).

DNA isolated from blood samples of 100 controls consisting of individuals who married into FIS families, and who did not have FIS, were amplified with rs117273909 primers using GeneAmp High Fidelity PCR System (Applied Biosystems, Grand Island, NY). The products were sequenced on an Applied Biosystems / Hitachi 3730 Genetic Analyzer. Sequencing analysis was performed using Sequencing Analysis version 5.2, and Sequence Scanner version 1.0 (both from Applied Biosystems, Grand Island, NY). Alignments of DNA sequences were done with SeqScape (Applied Biosystems, Grand Island, NY), Sequencher (Gene Codes Corporation, Ann Arbor, MI), and CodonCode Alignment software (v 3.7.1.1).

**Statistical methods**

Data cleaning was carried out on 344 CNE SNVs and 70 insertions/deletions from 46 individuals. Individuals with a genotype missing rate > 10%, and variants (SNVs, insertions and deletions) with either a Polypred (v6.11) value < 99, a missing rate > 10%, and/or two or more Mendelian inconsistencies were removed. The Polypred program identifies heterozygous single nucleotide substitutions, and assigns scores ranging from 99 to 0 to each heterozygous site, where a score of 99 indicates a very good fit and stands for a true positive rate of > 97%. Mendelian inconsistencies were tested using PEDCHECK (O’Connell and Weeks 1998). SNVs that became monomorphic after these steps were removed. Data cleaning reduced the number of individuals for analysis to 38, the number of SNVs to 197, and the number of insertions/deletions to 40. Of the remaining individuals, 22 had scoliosis (of which 14 had KS), 12 were unaffected and four had no curvature information. Tests of Hardy-Weinberg equilibrium were performed on each individual in the lower box. Recombination rates (cm/Mb) are depicted by the purple vertical lines. The plot was created using LocusZoom (https://statgen.sph.umich.edu/locuszoom).

Comparing the likelihood of the data in models with and without a marker, and uses the phenotype and genotype information of the entire family. The degree of lateral curvature was analyzed as a quantitative phenotype, and the phenotype and genotype information of the entire family. The degree of lateral curvature was analyzed as a quantitative phenotype, and genotype (a/a, a/A, or A/A) and allelic (presence of minor allele) tests of association were performed.

**Cloning of putative regulatory elements into zebrafish enhancer detection vector**

Site-directed mutagenesis was used to change the wild type rs117273909 C allele to the variant T allele (Bioinnovatise, Rockville, MD). Four constructs (198bp C allele, 198bp T allele, 687bp C allele, and 687bp T allele) were cloned into the zebrafish enhancer detection (ZED) vector (Bessa et al. 2009), which has insulators that prevent false positive expression due to position effects. Tol2 mRNA was synthesized with the mMessage mMachine SP6 kit (Ambion, Grand Island, NY) using NolI linearized pCS2FA-transposase vector as template DNA (Kwan et al. 2007).

**Zebrafish transgenesis**

Microinjections of all four constructs were performed into one- to two-cell stage zebrafish embryos using 25–45 pg of plasmid DNA mixed with 50 pg of Tol2 mRNA. Embryos were incubated at 28° C with 0.0033% PTU (1-phenyl 2-thiourea) to suppress pigmentation. Embryos with transgene integration were identified by red fluorescent protein (RFP) expression in the skeletal muscle at 48 hr post-fertilization (hpf), and green fluorescent protein (GFP) expression was observed in RFP positive embryos from 48 hpf to 5 dpf. Embryos positive for both RFP and GFP were grown to adulthood. Several germline transmitting founders were identified for each construct, and their progeny were evaluated for patterns of GFP expression from 48 hpf to 5 dpf. Embryos from F2 generation of germline transmitting founders were also evaluated for GFP expression by crossing F1 adult fish with wildtype fish. The only difference in sequence between the C and T allele constructs was at the rs117273909 locus. Screening and imaging of embryos were performed using the Zeiss SteREO Lumar.V12 stereomicroscope with an AxioCam HRC color camera or Zeiss Axio Observer Z1 inverted microscope with an AxioCam MRm black and white camera.
Table 2 SNV genotypes for family 1

| SNV          | Conserved Regiona | P-Valueb | Allelesc | MAFd | Unaffected | Scoliosis | Kyphoscoliosis |
|--------------|-------------------|----------|----------|------|------------|-----------|----------------|
| rs114080324  | 2947499           | Yes      | 0.116    | G/A  | 0.006      | GG        | GG AG AG AG   |
| rs183225473  | 3045672           | No       | 0.039    | G/C  | 0.006      | GG        | GG GG CG CG   |
| rs100688728  | 3045983           | Yes      | 0.034    | T/G  | 0.27       | TT        | TT TT TT TT   |
| rs146531224  | 3046004           | No       | 0.039    | G/A  | 0.006      | GG        | GG GG AG AG   |
| rs117273909  | 3182971           | Yes      | 0.039    | C/T  | 0.005      | CC        | CC CC CC CT   |
| rs111916055  | 3187995           | No       | 0.034    | A/G  | 0.028      | AA        | AA AA AA AG   |
| Novel        | 3188045           | No       | 0.083    | T/G  | NA         | TT        | TT TT TT AT   |
| rs16871553   | 3269602           | No       | 0.029    | C/T  | 0.199      | TT        | TT TT TT CT   |
| rs73733752   | 3278780           | No       | 0.039    | A/G  | 0.03       | AA        | AA AA AA AA   |

Data availability
The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

RESULTS

Variant analysis of IRX genes exons and CNEs
Sequencing of the IRX1, IRX2, and IRX4 exons in 53 individuals did not identify any functional sequence variants. Sequencing of the CNEs surrounding the IRX genes in 46 individuals identified 197 SNVs suitable for analysis, of which 16 were \( \pm 500 \) kb of IRX4, 70 were \( \pm 500 \) kb of IRX2, and 160 were \( \pm 500 \) kb of IRX1, with eight SNVs overlapping the conserved regions of IRX4 and IRX2 and 41 SNVs overlapping the conserved regions of IRX1 and IRX2. Of the 197 SNVs, 23 were novel.

Intrafamilial tests of association between the genotypes of the 197 SNVs and the quantitative trait (scoliotic curvature) resulted in 30 SNVs (Table 1) with nominal significance (\( P \)-values < 0.05). Only rs12517904 (\( P \)-value = 0.00004, Figure 1) was significant after adjusting for multiple testing (Bonferroni correction \( P \)-value = 0.00013). No association was identified with the 40 insertions/deletions.

We searched for variants cosegregating with KS. In one family, the genotypes of nine SNVs differed from the reference allele in all individuals affected with KS, and two out of three individuals with FIS (Table 2); unaffected family members, and all other individuals genotyped for this study, did not differ from the reference allele for any of these nine SNVs. The one family member with FIS that did not differ from the reference allele had a scoliotic curve of 36\( ^\circ \), and no hyperkyphosis. Two out of three individuals with FIS (Table 2) had two out of three SNVs differing from the reference allele in all individuals genotyped for this study, did not differ from the reference allele for any of these nine SNVs. The one family member with FIS that did not differ from the reference allele had a scoliotic curve of 36\( ^\circ \), and no hyperkyphosis.

Three of these SNVs were located in highly conserved regions, based on whole-genome alignment of vertebrates. Of these three, only rs117273909 (\( P \)-value = 0.039 for association analysis using all families genotyped in this study) was found to be conserved in all 96 vertebrate species in which this CNE and the corresponding CNE were present (www.genome.ucsc.edu). Rs117273909 is located in the ~841 kb gene desert between IRX1 (413 kb downstream) and IRX2 (431 kb downstream), and thus could function as an enhancer for either IRX1 or IRX2, or for both.

We determined the frequency of the variant T allele of rs117273909 in a Caucasian population by sequencing 100 controls consisting of individuals who married into our FIS families, and who did not have a history of FIS. Of these 100 controls, 90 matched the reference sequence (C), one was heterozygous (C/T), and nine failed to amplify, resulting in a T allele frequency of 0.00515. In the 1000 Genomes Project (www.1000genomes.org), rs117273909 was found to be C/T in 27 of 5008 genotypes (T allele frequency of 0.00539), and heterozygous A/T in three out of 5008 genotypes.

Based on the association with KS (\( P \)-value = 0.039), the presence of a nonreference allele cosegregation with KS in a single family, the fact that rs117273909 is located in a noncoding fragment (element_603, hg19 chr5:3182218-3183271, VISTA Enhancer Browser, http://enhancer.lbl.gov), which drives expression in the hindbrain in five out of 10 transgenic mice embryos (Visel et al. 2007), and the highly conserved nature of this SNV, we gave this SNV priority in our effort to determine if this SNV was functional. We then used zebrafish transgenesis to evaluate the effect of a C to a T allele substitution (rs117273909).

Regulatory activity of CNE containing rs117273909
We performed zebrafish transgenesis to determine if changing the allele at rs117273909 from a C (wild type allele) to a T would result in a change in its regulatory activity. Tena et al. (2011) tested a 1172 bp fragment encompassing rs117273909, which did not show any regulatory activity in Xenopus, possibly due to the presence of repressors; therefore, we chose to test smaller fragments for regulatory activity. For the zebrafish transgenesis assay, the wild type (C allele) and variant (T allele) versions of a 198 bp fragment (hg19 chr5:3182938-3183135), and an overlapping larger 687 bp fragment (hg19 chr5:3182466-3183152), both selected based on strong conservation across vertebrates, were tested for regulatory activity (Figure S2). Despite strong RFP expression, no consistent GFP expression patterns were observed in the embryos injected with any of the four constructs (Figure S3).

Due to the mosaic nature of the transgene expression during transient transgenesis, we generated multiple stable transgenic lines for all four constructs by screening founders for germline transmission. Variable GFP expression patterns, ranging from weak and diffused expression to strong and specific expression, were observed in the F1 and F2 progeny of multiple germline transmitting founders for each of the four constructs (Figure 2, A–P). These data may indicate positional effects of the transgene integration in independent founders for each of the four constructs. The specific GFP expression patterns were observed in pineal gland, pharyngeal arches, and brain. However, there was no association of these expression patterns with any particular construct.

DISCUSSION
Our results suggest that noncoding fragments as small as 198 bp can act as regulatory elements in zebrafish, which is in contrast to a report that a 1172 bp fragment (TA3235, 131:818307–820479 bp, xenTro2)
is also possible that unidentifiable lateral and thoracic curvature in individuals in these families. Future mental patterning and be responsible for the variation of the degree of regions further upstream from the individual SNV can be determined only by SNVs may play a role in the phenotype, but the relevance of each that may lie in regulatory regions that in promoter to the right tissue at the correct time.

RNA polymerase, transactivators, and transcription factors to the pro-

1.6 Mb. This looping mechanism may help facilitate the delivery of
ters interact with more than one differ- ences in where expression occurred, but affected the level and/or

regions, or the type of activity present (repressors vs. enhancers). We
due to positional effects, something the ZED vector should mini-
mize, or could be the result of the involvement of this regulatory re-
gion with a homeobox gene. The expression of our conserved fragment in the pineal gland is of interest, since pinealectomised chickens de-

expression trend with one of the two alleles at rs117273909, we did not

be due to positional effects, something the ZED vector should mini-
mize, or could be the result of the involvement of this regulatory re-
gion with a homeobox gene. The expression of our conserved fragment in the pineal gland is of interest, since pinealectomised chickens develop scoliosis (Machida et al. 1993).

There appeared to be differences in the level of expression and the timing of GFP expression between the wild type and alternate allele at rs117273909. de la Calle-Mustienes et al. (2005) noticed that DNA sequence differences in zebrafish and mouse CNEs did not result in differences in where expression occurred, but affected the level and/or timing of the transcription. The change in timing and regulation may be due to the sequence degeneration, which interferes with loop for-

mation. Tena et al. (2011) identified a three-dimensional architecture that forms through CCCTC-binding, present in the Irx clusters of mice, zebrafish, and Xenopus, which brings the Irx1/3 and Irx2/5 promoters together, and demonstrated that cis-regulatory elements in the Irx clusters interact with more than one Irx promoter, up to distances of 1.6 Mb. This looping mechanism may help facilitate the delivery of RNA polymerase, transactivators, and transcription factors to the pro-
moter to the right tissue at the correct time.

In summary, tests of association identified several significant SNVs that may lie in regulatory regions that influence gene expression. These SNVs may play a role in the phenotype, but the relevance of each individual SNP can be determined only by in vitro and in vivo assays. It is also possible that unidentified SNVs in highly conserved regulatory regions further upstream from the IRX family may disrupt develop-

damental patterning and be responsible for the variation of the degree of lateral and thoracic curvature in individuals in these families. Future work will focus on completely sequencing the region with the most significant association, as well as performing additional in vivo assays of

regulatory regions surrounding the other associated SNVs identified in this study, including rs1251709, and the SNVs in LOC285577, of which little is known to date. Once we have a better understanding of what tissues are involved in the scoliosis phenotype, we will be able to define the regulatory regions using both in vitro and in vivo assays.

ACKNOWLEDGMENTS

The ZED vector was kindly provided by Jose Bessa (Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide, Carretera de Utrera Km1, 41013 Sevilla, Spain). We greatly appreciate the helpful suggestions provided by Valer Gotea, Laura Elnitski, and Tyra Wolfsberg. We also would like to thank the families who participated in this study. This research was supported, in part, by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health (NIH). Some of the results were obtained with the program S.A.G.E., which was supported by grant 1 P41 RR03655 from the National Center for Research Resources. Grant funding was provided by the Scoliosis Research Society, the National Scoliosis Foundation, Scoliosis Association, Inc., the Institute de France Foundation Yves Cotrel, the Orthopedic Pediatric Society of North America, and the Orthopedic Research and Education Foundation, and NIH grant 1-R01-AR048862-01A1. The control samples were sequenced by the University of Colorado Center for DNA Sequencing and Analysis Core (http://DNASequencingCore.ucdenver.edu), which is supported by a NIH/National Cancer Institute Cancer Center Core Support Grant (P30 CA046934).

LITERATURE CITED

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W. J. Kent et al., 2004 Ultraconserved elements in the human genome. Science 304(5675): 1321–1325.

Bergofsky, E. H., 1979 Respiratory failure in disorders of the thoracic cage. Am. Rev. Respir. Dis. 119(4): 643–669.

Bessa, J., J. J. Tena, E. de la Calle-Mustienes, A. Fernandez-Minan, S. Naranjo et al., 2009 Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. Dev. Dyn. 238(9): 2409–2417.

Bosse, A., A. Stoykova, K. Nieselt-Struwe, K. Chowdhury, N. G. Copeland et al., 2000 Identification of a novel mouse Iroquois homeobox gene, Irox5, and chromosomal localisation of all members of the mouse Iroquois gene family. Dev. Dyn. 218(1): 160–174.
Bunnell, W. P., 1986 The natural history of idiopathic scoliosis before skeletal maturity. Spine 11(8): 773–776.

Bürglin, T. R., 1997 Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res. 25(21): 4173–4180.

Cavodeassi, F., J. Modolell, and J. L. Gomez-Skarmeta, 2001 The Iroquois family of genes: from body building to neural patterning. Development 128(15): 2847–2855.

Chan, V., C. C. Fong, K. D. Luk, B. Yip, M. K. Lee et al., 2002 A genetic locus for adolescent idiopathic scoliosis linked to chromosome 19p13.3. Am. J. Hum. Genet. 71(2): 401–406.

de la Calle-Mustienes, E., C. G. Feijoó, M. Manzanares, J. J. Tena, E. Rodriguez-Segovia et al., 2005 A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. Genome Res. 15(8): 1061–1072.

Edery, P., P. Margaritte-Jeannin, B. Biot, A. Labalme, J. C. Bernard et al., 2011 New disease gene location and high genetic heterogeneity in idiopathic scoliosis. Eur. J. Hum. Genet. 19(8): 865–869.

Gao, X., D. Gordon, D. Zhang, R. Browne, C. Helms et al., 2007 CHD7 gene polymorphisms are associated with susceptibility to idiopathic scoliosis. Am. J. Hum. Genet. 80(5): 957–965.

Gómez-Skarmeta, J. L., and J. Modolell, 2002 Iroquois genes: genomic organization and function in vertebrate neural development. Curr. Opin. Genet. Dev. 12(4): 403–408.

Goode, D. K., P. Snell, S. F. Smith, J. E. Cooke, and G. Elgar, 2005 Highly conserved regulatory elements around the SHH gene may contribute to the maintenance of conserved synteny across human chromosome 7q36.3. Genomics 86(2): 172–181.

Justice, C. M., N. H. Miller, B. Marosy, J. Zhang, and A. F. Wilson, 2003 Familial idiopathic scoliosis: evidence of an X-linked susceptibility locus. Spine 28(6): 589–594.

Kou, I., Y. Takahashi, T. A. Johnson, Y. Ogura et al., 2013 Identification of a susceptibility locus for severe adolescent idiopathic scoliosis on chromosome 17q24.3. PLoS One 8(9): e72802.

Miyake, A., I. Kou, Y. Takahashi, T. A. Johnson, Y. Ogura et al., 2013 Identification of a susceptibility locus for adolescent idiopathic scoliosis on chromosome 17q24.3. PLoS One 8(9): e72802.

Nobrega, M. A., I. Ovcharenko, V. Afzal, and E. M. Rubin, 2003 Scanning human gene deserts for long-range enhancers. Science 302(5644): 413.

Ocaka, L., C. Zhao, J. A. Reed, N. D. Ebenezer, G. Brice et al., 2008 Assignment of two loci for autosomal dominant adolescent idiopathic scoliosis to chromosomes 9q31.2–q34.2 and 17q25.3–qtel. J. Med. Genet. 45(2): 87–92.

O’Connell, J. R., and D. E. Weeks, 1998 PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am. J. Hum. Genet. 63(1): 259–266.

Ogura, K., K. Matsumoto, A. Kuroiwa, T. Isobe, T. Otoguro et al., 2001 Cloning and chromosome mapping of human and chicken Iroquois (IRX) genes. Cytogeten. Cell Genet. 92(3–4): 320–325.

Peters, T., R. Dildrop, K. Ausmeier, and U. Rüther, 2000 Organization of mouse Iroquois homeobox genes in two clusters suggests a conserved regulation and function in vertebrate development. Genome Res. 10(10): 1453–1462.

Raggo, C. L., P. F. Giampietro, S. Dobrin, C. Zhao, D. Dorschort et al., 2009 A novel locus for adolescent idiopathic scoliosis on chromosome 12p. J. Orthop. Res. 27(10): 1366–1372.

Salehi, L. B., M. Mangino, S. De Serio, D. De Cicco, F. Capon et al., 2002 Assignment of a locus for autosomal dominant idiopathic scoliosis (IS) to human chromosome 17p11. Hum. Genet. 111(4–5): 401–404.

Sandelin, A., P. Bailey, S. Bruce, P. G. Engström, J. M. Klos et al., 2004 Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genomics 5(1): 99.

Takahashi, Y., I. Kou, A. Takahashi, T. A. Johnson, K. Kono et al., 2011 A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. Nat. Genet. 43(12): 1237–1240.

Tana, J. J., M. E. Alonso, E. de la Calle-Mustienes, E. Splinter, W. de Laat et al., 2011 An evolutionarily conserved three-dimensional structure in the vertebrate Irx clusters facilitates enhancer sharing and coregulation. Nat. Commun. 2: 310.

Visel, A., S. Minovitsky, I. Dubchak, and L. A. Pennacchio, 2007 VISTA Enhancer Browser—a database of tissue-specific human enhancers. Nucleic Acids Res. 35(Database issue): D88–D92.

Wigginton, J. E., D. J. Cutler, and G. R. Abecasis, 2005 A note on exact tests of Hardy-Weinberg equilibrium. Am. J. Hum. Genet. 76(5): 887–893.

Wise, C. A., R. Barnes, J. Gillum, J. A. Herring, A. M. Bowcock et al., 2000 Localization of susceptibility to familial idiopathic scoliosis. Spine 25(18): 2372–2380.

Woodle, A., M. Goodson, D. K. Goode, P. Snell, G. K. McEwen et al., 2005 Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 3(1): e7.

Wu, Q., N. Tommerup, S. Ming Wang, and L. Hansen, 2006 A novel primate specific gene, CEI, is located in the homeobox gene IRX4 promoter in Homo sapiens. Gene 371(2): 167–173.