

Articles



Family-based integrative genomics identifies neural–metabolic–skeletal pathways underlying scoliosis susceptibility

Xuanchen Li^{1#}, Zhenrui Lin^{2#}, Xiaosheng Chen^{3,4,5}, Zhixiang Zhu^{3,4,5}, Rui Zhang^{3,4,5}, Weijun Wang⁴, Wenyu Zhou^{3,4,5}, Lei Yang^{3,4,5}, Yi He¹, Lihe Jiang^{6*}, Guodong Huang^{1*}, Qian Liang^{3,4,5*}

Abstract

Scoliosis is a complex spinal deformity with strong familial aggregation, yet its genetic basis remains incompletely defined. Here, we performed whole-exome sequencing (WES) in 34 scoliosis families, identifying candidate single nucleotide polymorphisms (SNPs) potentially contributing to disease susceptibility. To connect genomic variation with transcriptional regulation, we integrated these data with cis-expression quantitative trait loci (cis-eQTL) summary statistics from eQTLGen and scoliosis GWAS data from FinnGen R12 using summary-based Mendelian randomization (SMR). After harmonization and filtering ($P_{\text{SMR}} < 0.05$, $P_{\text{HEIDI}} > 0.05$), we identified 31 shared genes overlapping between family-based SNPs and eQTL-associated loci. Among them, *LINC00299* and *NPIPBI* exhibited strong colocalization signals, with *LINC00299* expression inversely correlated with scoliosis risk. Functional annotation further highlighted *NCAM1*, *AGPAT3*, *SMIM12*, and *CTSH* as core pathogenic nodes linking neural development, lipid metabolism, mitochondrial energy regulation, and bone remodeling. Together, these findings reveal a multi-systemic genetic architecture of scoliosis and nominate metabolic-neuro-skeletal pathways as potential targets for precision diagnosis and therapy.

Article history: Submitted 16 March 2026 | Accepted 27 May 2026 | Published 3 July 2026
iCell, Vol.3, No.1 (2026): 23-32. DOI:https://dx.doi.org/10.71373/YIHW8377

Introduction

Scoliosis represents a common structural deformity of the spine, affecting up to 3% of the global population, yet its etiology remains largely elusive^[1-3]. Familial aggregation and twin studies strongly support a genetic component, but known risk loci explain only a fraction of disease heritability. Previous genome-wide association studies (GWAS) have identified several susceptibility regions; however, pinpointing the causal variants and their regulatory mechanisms has been hindered by the complexity of spinal development and by the polygenic nature of the disorder^[4].

To better define the genetic determinants underlying scoliosis, family-based sequencing approaches offer a complementary strategy that reduces population heterogeneity and enhances power to detect rare, high-impact variants. In parallel, advances in expression quantitative trait loci (eQTL) resources and

summary-based Mendelian randomization (SMR) analysis enable systematic integration of transcriptional regulation with GWAS signals to infer putative causal relationships between gene expression and disease phenotypes.

In this study, we combined whole-exome sequencing (WES) from 34 scoliosis families with cis-eQTL data from eQTLGen and scoliosis GWAS data from FinnGen R12^[5-7]. We aimed to identify genes whose regulatory variants exert causal effects on scoliosis susceptibility, characterize their biological functions, and delineate the molecular pathways connecting neural, muscular, metabolic, and skeletal systems in spinal morphogenesis. Our integrative analysis highlights novel pathogenic genes and provides a mechanistic framework for understanding the multi-system basis of scoliosis.

¹Shenzhen Key Laboratory of Neurosurgery, Department of Neurosurgery, The First Affiliated Hospital of Shenzhen University, Shenzhen Second People's Hospital, 3002# Sungang Road, Futian District, Shenzhen 518035, China. ²Guangxi Key Laboratory of AIDS Prevention and Treatment & Guangxi Colleges and Universities Key Laboratory of Prevention and Control of Highly Prevalent Diseases, School of Public Health, Guangxi Medical University, Nanning 530021, Guangxi, China. ³Department of Spine Surgery, The First Affiliated Hospital of Shenzhen University, Shenzhen Second People's Hospital, 3002# Sungang Road, Futian District, Shenzhen 518035, China. ⁴Shenzhen Youth Spine Health Center, The Shenzhen Second People's Hospital, 3002# Sungang Road, Futian District, Shenzhen 518035, China. ⁵Medical Innovation Technology Transformation Center of Shenzhen Second People's Hospital, 3002# Sungang Road, Futian District, Shenzhen 518035, China. ⁶School of Basic Medical Sciences, Youjiang Medical University for Nationalities, Baise 533000, Guangxi, China.

These authors contributed equally to this work.

*Correspondence: Qian Liang (liangq2023spine@126.com), Guodong Huang (jxgd211@163.com), Lihe Jiang (jianglihe@ymun.edu.cn)

Mail: Xuanchen Li, xuanchenli313@gmail.com; Zhenrui Lin, 20224302735@sr.gxmu.edu.cn;

Xiaosheng Chen, dr_chenxiaosheng@163.com; Zhixiang Zhu, 844314305@qq.com;

Rui Zhang, zhang.ruino.1@163.com; Wenyu Zhou, drzhouwenyu@163.com;

Lei Yang, yiyuanbgs@126.com; Yi He, heyi810721@163.com; Weijun Wang, 122209671@qq.com

Methods

WES Data from Scoliosis Families

The WES data of scoliosis families included in this study were obtained from 34 scoliosis families previously collected by our research group. The study protocol was approved by the ethics committee (approval number: 20181106003-GZ2021). After obtaining informed consent, peripheral blood samples were collected, and genomic DNA was extracted. WES was performed for sequencing, and following quality control and variant filtering, candidate SNP data were generated for subsequent intersection analysis with the SMR results. Detailed sequencing data, sampling details, and systematic sample information for the 34 families are provided in **Supplementary Table 1**.

cis-eQTL Data

The cis-eQTL summary statistics used in this study were obtained from the eQTLGen Consortium, which provides standardized cis-eQTL summary files suitable for SMR analysis via its official website. This dataset integrates whole-blood samples from 31,684 individuals of European ancestry across 37 independent studies, offering comprehensive and high-quality cis-eQTL data^[8]. The data are available at: <https://www.eqtlgen.org/cis-eqtls.html>.

GWAS Data

The scoliosis GWAS summary statistics were obtained from the FinnGen R12 database, which integrates genomic and health data from 500,000 Finnish biobank participants. FinnGen is a large-scale genomics initiative that has analyzed over 500,000 Finnish biobank samples and linked genetic variants to health data to understand disease mechanisms and susceptibility^[9]. Case-control summary statistics for scoliosis were retrieved using the keyword “scoliosis”. The data are available at: (1) FinnGen R12 summary statistics: <https://www.finnngen.fi>; (2) Scoliosis data: https://storage.googleapis.com/finngen-public-data-r12/summary_stats/release/finngen_R12_M13_SCOLIOSIS.gz.

Data Preprocessing

The raw summary statistics from the FinnGen R12 scoliosis GWAS and the eQTLGen cis-eQTL data were subjected to standardized preprocessing and format harmonization. For the GWAS data, core fields including SNP identifier, effect allele, reference allele, allele frequency, effect size, standard error, and P-value were uniformly extracted, and field naming conventions were standardized. Duplicate and blank SNP records were removed to ensure unique variant identification. Total sample size information was added, and the data were converted into the outcome data format required for SMR analysis. Subsequently, the cis-eQTL and GWAS data underwent allele orientation alignment, genomic coordinate matching, and gene symbol standardization to eliminate format discrepancies and naming inconsistencies across different data sources, ensuring complete matching of variant information between the two datasets. Finally, standardized input files directly applicable to SMR colocalization analysis were generated.

SMR Analysis

Based on the principle of Mendelian Randomization (MR), this study used SMR (Windows 1.3.1) software to perform colocalization and causal inference analysis between gene expression and scoliosis. By integrating cis-eQTL and GWAS summary data, this method overcomes the limitation of traditional GWAS that can only identify genetic associations but cannot pinpoint functional causal genes. Single nucleotide polymorphisms (SNPs) significantly associated with gene expression in the cis-eQTL data were used as instrumental variables, with gene expression level (cis-eQTL) as the exposure and scoliosis (GWAS) as the outcome. The effect estimates of the same variants from the GWAS summary data were integrated to quantitatively infer the potential causal effect of gene expression on the risk of scoliosis.

The significance threshold was set at $P_{\text{SMR}} < 0.05$. Since two different variants located in linkage disequilibrium (LD) may show significant mutual associations, relying solely on the significance of SMR results cannot directly distinguish true causal associations from false positives due to LD, i.e., it is difficult to exclude pleiotropic effects^[10]. To effectively differentiate LD effects from true pleiotropic causal effects, Zhu et al.^[10] proposed the HEIDI test based on instrumental variables. In this study, $P_{\text{HEIDI}} > 0.05$ was used as a filtering criterion to remove false positive associations caused by LD interference. Ultimately, significantly associated genes passing both filters were obtained, along with their association information, generating the *SMR.filter.csv* dataset, which serves as a core candidate set for subsequent functional enrichment and mechanistic analysis.

Family-Based Gene Intersection Analysis

To integrate the results from family-based WES and public omics data, and to improve the reliability and specificity of candidate genes, an intersection analysis was performed between the two independent gene sets. The significant genes identified by SMR analysis were intersected with those carrying key SNPs from the WES analysis of scoliosis families. Venn diagrams were plotted using R software (version 4.4.2) with the ggvenn package^[11] to visualize the overlapping genes. Finally, 31 common genes present in both datasets were identified as core candidates for subsequent functional annotation and mechanistic exploration of scoliosis pathogenesis.

Data Visualization Analysis

All genome-wide data visualization analyses were performed using R software (version 4.4.2). The dplyr and tidyr packages^[12,13] were used for data manipulation, and the CMplot package^[14] was used to generate Manhattan plots. A Manhattan plot was drawn with chromosomes on the x-axis and $-\log_{10}$ P-value on the y-axis for the 31 common genes, providing a visual representation of the genome-wide distribution of scoliosis-associated SNPs across human chromosomes. The Manhattan plot clearly shows the chromosomal localization of significant association loci and preliminarily reveals the distribution pattern of scoliosis susceptibility regions, offering a visual basis for subsequent core gene screening and functional interpretation.

To further validate the colocalization association between gene expression and scoliosis risk, scatter plots of eQTL effect size versus GWAS effect size were generated for each of the 31 common genes using the magick and TeachingDemos packages

^[15,16] in R 4.4.2. In these plots, the x-axis represents the eQTL effect size of the variant on gene expression, and the y-axis represents the GWAS effect size of the variant on scoliosis risk, with the top cis-eQTL and common cis-eQTL sites annotated for each gene. These scatter plots effectively illustrate the correlation trend between eQTL effects and GWAS effects, assist in evaluating the strength of the causal association between gene expression and disease phenotype, and provide visual support for the colocalization validation of core candidate genes.

Functional Enrichment

The Gene Ontology (GO) database provides a standardized annotation system for gene functions and serves as one of the core bioinformatics tools for interpreting gene expression data. GO enrichment analysis systematically annotates gene functions from three dimensions: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC), effectively identifying statistically enriched functional categories and biological characteristics under disease conditions^[17]. In this study, GO enrichment analysis was performed using the clusterProfiler package^[18] in R software (version 4.4.2), and bubble plots were generated. A *P*-value < 0.05 was considered statistically significant for enrichment results^[19].

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a comprehensive molecular function database constructed by the Bioinformatics Center of Kyoto University, covering 17 sub-databases and integrating information on cellular processes, genetic information, human diseases, and more^[20,21]. Enrichment analysis based on KEGG data is a common approach to identify biological pathways and functional modules involved in target genes. In this study, KEGG enrichment analysis was performed using the clusterProfiler package in R 4.4.2, and bubble plots were visualized. A *P*-value < 0.05 was considered statistically significant for KEGG enrichment results^[19].

Results

Family-based analysis of scoliosis cohorts

Based on the integrated analysis of WES data from 34 scoliosis families and SMR results, a total of 31 common genes present in both datasets were identified. These genes passed stringent filtering criteria (*P*_{SMR} < 0.05 and *P*_{HEIDI} > 0.05), indicating that they carry key genetic variants at the family level while also showing causal associations with scoliosis risk at the population level. The *P*_{SMR} and *P*_{HEIDI} values for these 31 common genes are detailed in **Table 1**.

Table 1. SMR and HEIDI test *P*-values for the 31 common genes associated with scoliosis

Correlation with loci	Gene	<i>P</i> _{SMR}	<i>P</i> _{HEIDI}
Positive	<i>NCAM1</i>	0.000922659	0.07680346
Positive	<i>AGPAT3</i>	0.004800337	0.1413867
Positive	<i>SMIM12</i>	0.00501161	0.05906706
Negative	<i>CTSH</i>	0.00691361	0.8436791
Negative	<i>LINC00299</i>	0.01113481	0.06945593
Negative	<i>NPIPB11</i>	0.01167407	0.2094704
Positive	<i>NAPRT</i>	0.01293904	0.2441823
Negative	<i>HSPBP1</i>	0.0159643	0.9743946
Negative	<i>TMEM132D</i>	0.01659424	0.8338818
Negative	<i>DEAF1</i>	0.01783776	0.5167472
Positive	<i>TSHR</i>	0.01815261	0.9751504
Positive	<i>ALOX12-AS1</i>	0.0189947	0.4511633
Negative	<i>L3MBTL4</i>	0.02060692	0.5297665
Positive	<i>PRCP</i>	0.02117731	0.6506786
Positive	<i>EIF3E</i>	0.02500594	0.5799573
Positive	<i>KCNN3</i>	0.02503178	0.8401287
Negative	<i>CRYGN</i>	0.02639662	0.9189739
Positive	<i>CAMK2D</i>	0.02760503	0.2768263
Negative	<i>NMNAT3</i>	0.02938423	0.3255898
Negative	<i>NUP214</i>	0.03011871	0.6775703
Not determined	<i>DLG2</i>	0.03200031	0.3766503

Positive	<i>NMT1</i>	0.03203178	0.1570335
Negative	<i>HEG1</i>	0.03310556	0.1746253
Negative	<i>GOLGA8O</i>	0.03618912	0.2041633
Positive	<i>SYNJ2BP</i>	0.03664463	0.420418
Positive	<i>VDR</i>	0.03686418	0.5702664
Negative	<i>GRB10</i>	0.03821055	0.2389569
Negative	<i>ACOT7</i>	0.03850855	0.6677867
Negative	<i>DOCK7</i>	0.04185344	0.8608807
Negative	<i>LDHAL6A</i>	0.04815689	0.6968079

Colocalization validation at the *LINC00299* locus

The Manhattan plot illustrates the distribution of the 31 common genes across all chromosomes (Figure 1A), revealing that multiple genomic loci are significantly associated with scoliosis susceptibility. Colocalization analysis indicated that the long non-coding RNA *LINC00299* exhibits a significant cis-eQTL effect in the context of scoliosis genetic susceptibility.

The single-gene scatter plot for *LINC00299* (Figure 1B) showed that most data points fall along a negative correlation trend line, indicating that alleles increasing its expression tend to decrease the risk of scoliosis. That is, higher expression of *LINC00299* is

associated with lower disease susceptibility, suggesting that down-regulation of this gene may promote abnormal spinal development. The top cis-eQTL for this locus (marked by a red triangle) also perfectly aligns with the overall fitted trend. Statistical analysis showed a significant SMR causal association test ($P_{SMR} = 0.01113481$, $P < 0.05$) and a HEIDI heterogeneity test ($P_{HEIDI} = 0.06945593$, $P > 0.05$) indicating no horizontal pleiotropy, thus excluding false colocalization bias.

Further analysis using the regional colocalization plot (Figure 1C) revealed that within the 8.5-9.0 Mb region on chromosome 2, the significant association peak from the scoliosis GWAS and the strong cis-eQTL association peak for *LINC00299* are highly overlapping. The two genetic signals are consistently distributed across the genomic interval, exhibiting clear colocalization characteristics.

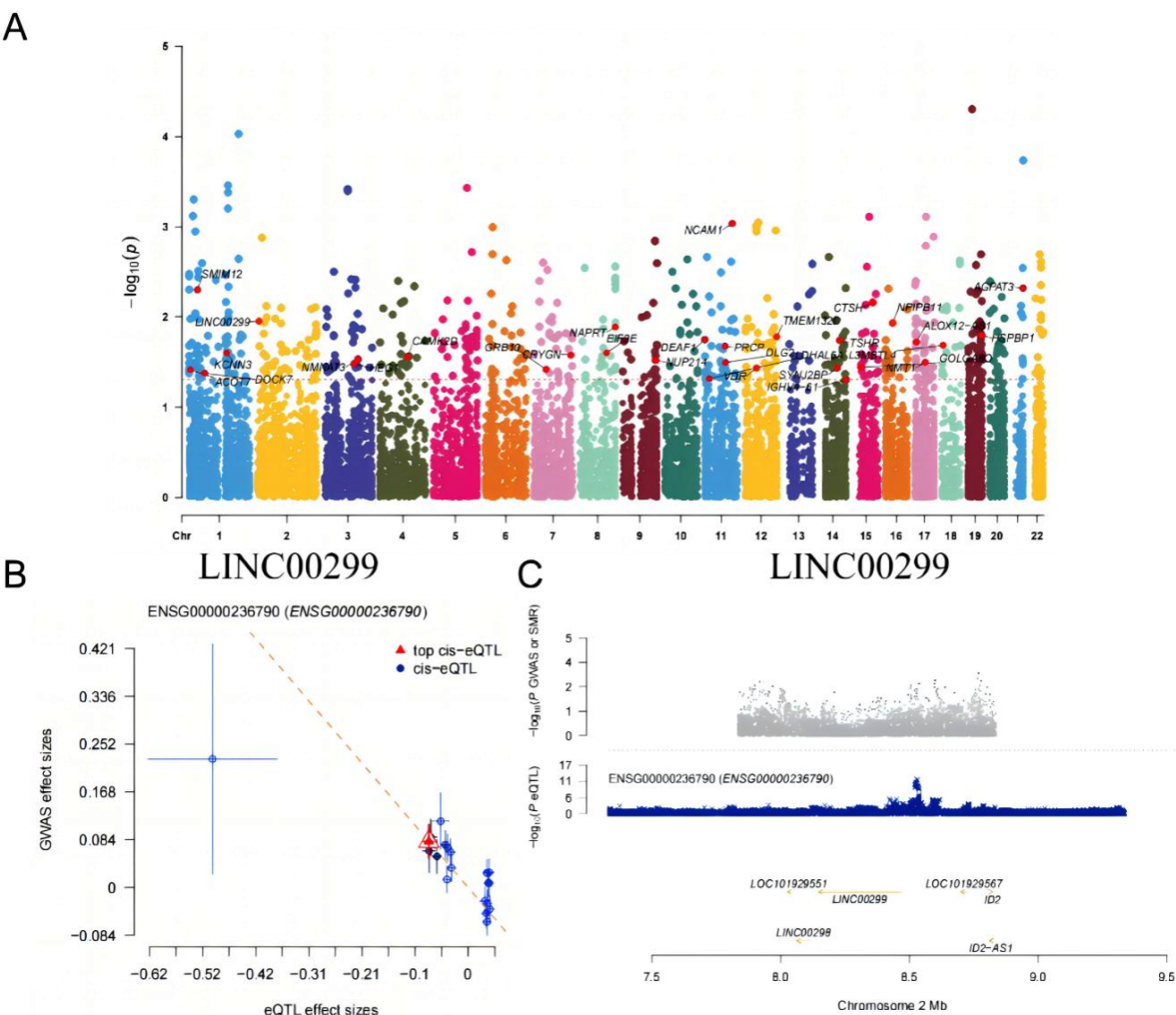


Figure 1. Chromosomal distribution of scoliosis-associated variants and cis-eQTL colocalization features at the *LINC00299* locus. (A) Manhattan plot showing the genome-wide distribution of scoliosis-associated single nucleotide polymorphisms across chromosomes; (B) Scatter plot of cis-eQTL versus GWAS effect sizes for *LINC00299*, with each point representing a variant located within the cis-regulatory region of *LINC00299*; (C) Fine-scale colocalization map of GWAS and cis-eQTL signals at the *LINC00299* locus.

eQTL- GWAS effect characteristics of core candidate genes

Through integrative colocalization analysis of scoliosis GWAS signals and cis-eQTL signals, this study identified four candidate genes with potential pathogenic relevance: *NCAM1*, *AGPAT3*, *SMIM12*, and *CTSH*.

For *NCAM1* (Figure 2A), *AGPAT3* (Figure 2B), and *SMIM12* (Figure 2C), the cis-eQTL effect sizes showed a stable positive

correlation with the GWAS disease effect sizes, indicating that the direction of genetic variant effects on gene expression was consistent with the direction of variant effects on scoliosis risk. In contrast, *CTSH* (Figure 2D) exhibited a clear negative correlation between eQTL effect sizes and GWAS effect sizes. Moreover, the top cis-eQTL (marked by red triangles) for each gene fell exactly on the overall fitted trend line, suggesting that altered expression levels of these four genes provide reliable genetic evidence for an association with the risk of scoliosis.

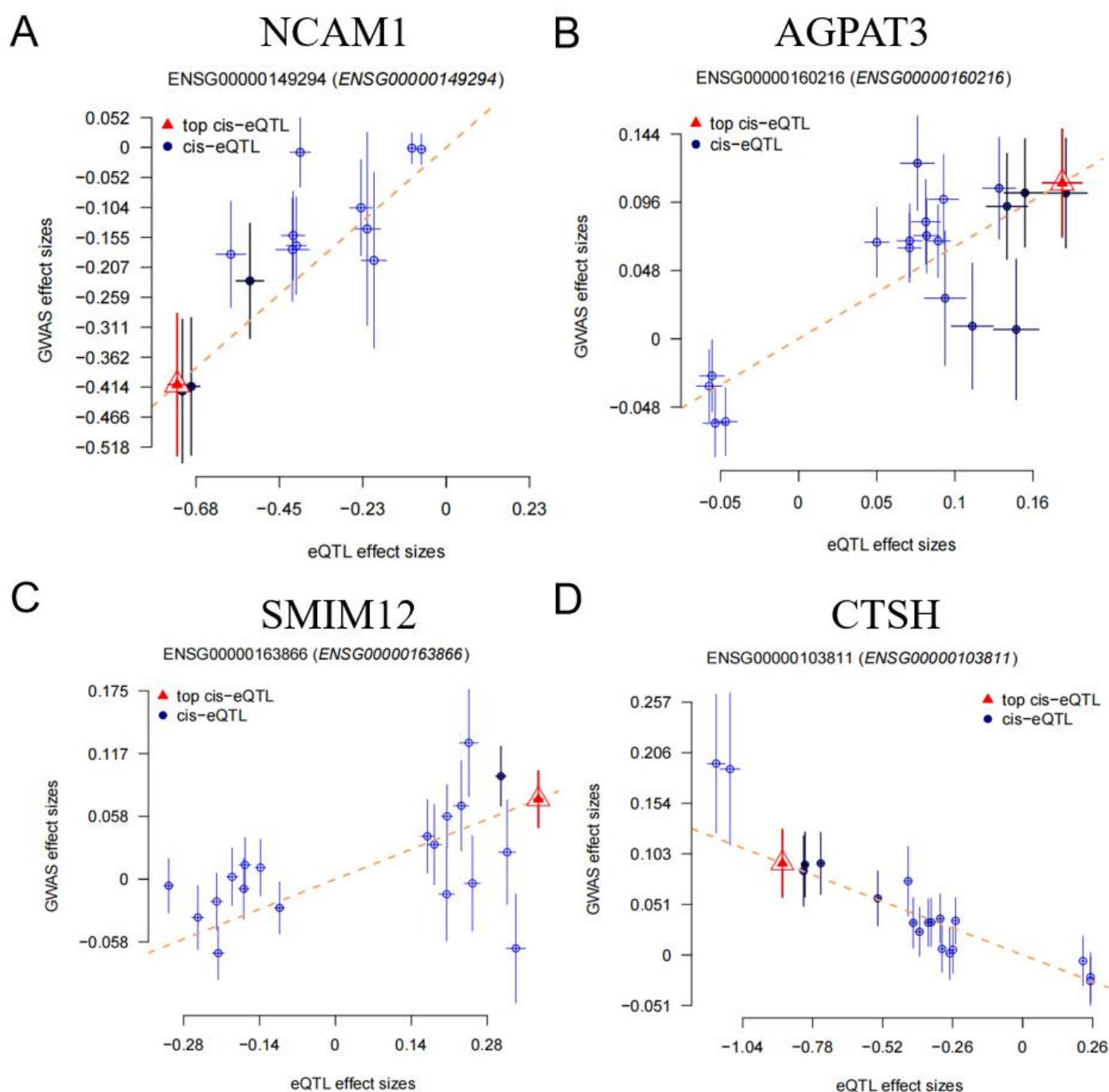


Figure 2. eQTL-GWAS effects of scoliosis-associated variants. (A) NCAM1 gene; (B) AGPAT3 gene; (C) SMIM12 gene; (D) CTSH gene.

Functional annotation of key genes associated with scoliosis

Functional enrichment analysis was conducted on the 31 consensus genes associated with scoliosis. GO analysis identified significant enrichment across three standard functional dimensions (Figure 3A-B).

In the BP category, these genes were mainly clustered in pathways related to energy coenzyme biosynthesis and neural signal transduction. The most significantly enriched terms included NAD biosynthetic process, nicotinamide nucleotide biosynthetic process, pyridine nucleotide biosynthetic process, pyridine-containing compound biosynthetic process, Rho protein signal transduction, and regulation of cold-induced thermogenesis.

For the CC category, prominent enrichment was found in

synaptic complexes, neuronal structures and specialized cell membrane domains, such as basal part of cell, postsynaptic density, asymmetric synapse, juxtaparanode region of axon, postsynaptic specialization, and neuron-to-neuron synapse.

In the MF group, enriched functions were primarily involved in catalytic activity, ion channel regulation and molecular binding. Key significant terms covered exopeptidase activity, dipeptidyl-peptidase activity, bile acid binding, myristoyltransferase activity, nuclear export signal receptor activity, and calcium-activated potassium channel activity. Full detailed statistics for all GO enrichment results are listed in **Table 2**.

Further KEGG pathway enrichment analysis revealed that these genes were significantly enriched in nicotinate and nicotinamide metabolism, insulin secretion, glucagon signaling pathway, Hypoxia-Inducible Factor-1 (HIF-1) signaling pathway, cofactor biosynthesis, and the renin-angiotensin system (Figure 3C).

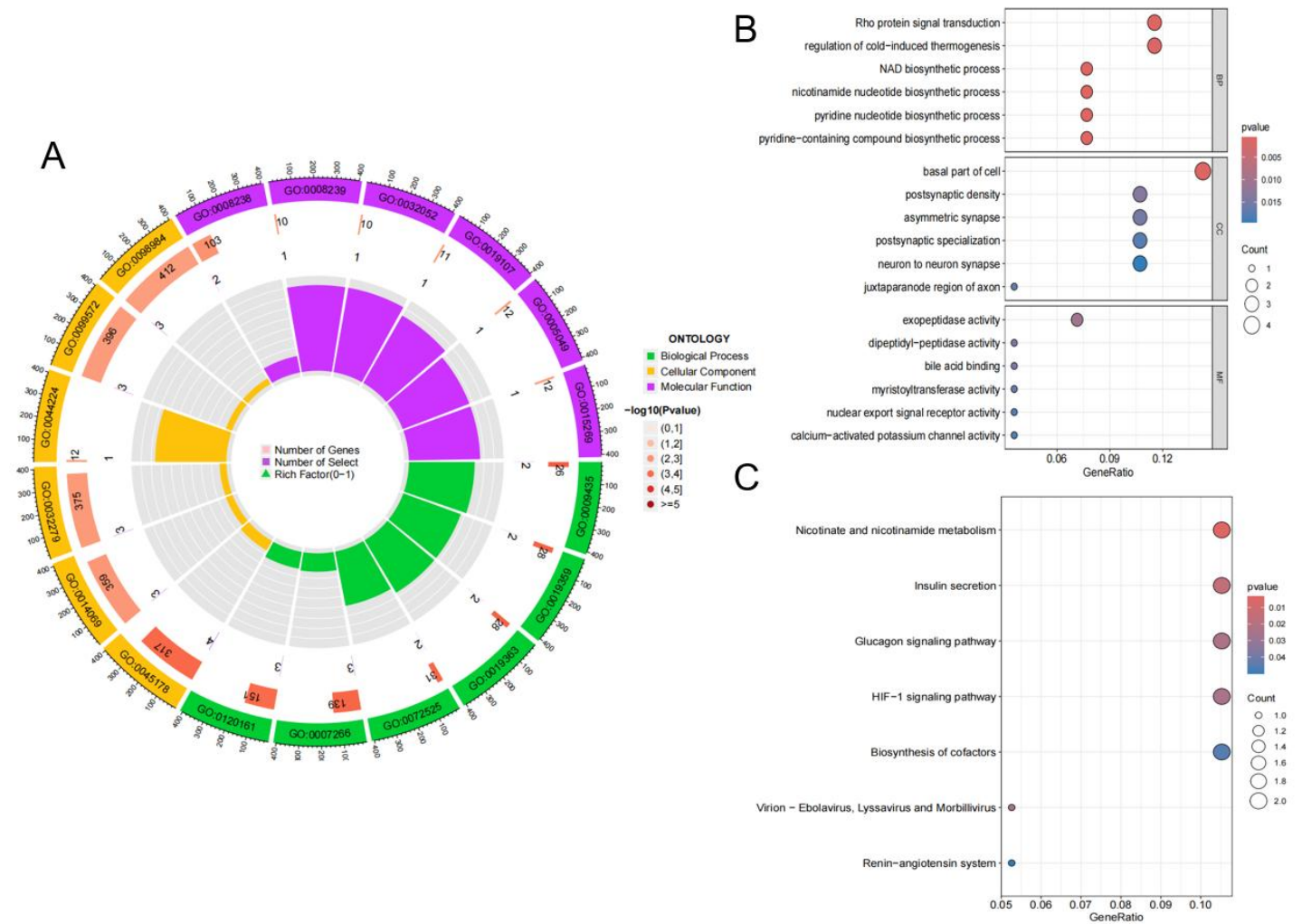


Figure 3. GO annotation and KEGG pathway enrichment analysis of key genes associated with scoliosis. (A) Circular plot of GO enrichment, categorized by Biological Process (green), Cellular Component (yellow), and Molecular Function (purple). The inner ring indicates the number of genes, while the outer rings represent the enrichment ratio and statistical significance (-log10 P-value); (B) Bubble plot of GO functional enrichment, showing the proportion of genes and statistical significance for each GO term; (C) KEGG pathway enrichment results, where bubble size denotes the number of genes involved in each pathway and color intensity represents the significance level of enrichment.

Table 2. GO enrichment analysis results of the 31 common genes associated with scoliosis

ID			Description	P value	Enrichment factor
BP	GO:0009435		NAD biosynthetic process	5.7e-04	56.17

BP	GO:0019359	nicotinamide nucleotide biosynthetic process	6.7e-04	52.16
BP	GO:0019363	pyridine nucleotide biosynthetic process	6.7e-04	52.16
BP	GO:0072525	pyridine-containing compound biosynthetic process	8.2e-04	47.11
BP	GO:0007266	Rho protein signal transduction	8.8e-04	15.76
BP	GO:0120161	regulation of cold-induced thermogenesis	1.1e-03	14.51
CC	GO:0045178	basal part of cell	9.5e-04	9.00
CC	GO:0014069	postsynaptic density	1.4e-02	5.96
CC	GO:0032279	asymmetric synapse	1.5e-02	5.70
CC	GO:0044224	juxtaparanode region of axon	1.7e-02	59.40
CC	GO:0099572	postsynaptic specialization	1.8e-02	5.40
CC	GO:0098984	neuron to neuron synapse	2.0e-02	5.19
MF	GO:0008238	exopeptidase activity	1.0e-02	12.99
MF	GO:0008239	dipeptidyl-peptidase activity	1.5e-02	66.92
MF	GO:0032052	bile acid binding	1.5e-02	66.92
MF	GO:0019107	myristoyltransferase activity	1.6e-02	60.83
MF	GO:0005049	nuclear export signal receptor activity	1.8e-02	55.76
MF	GO:0015269	calcium-activated potassium channel activity	1.8e-02	55.76

Discussion

By integrating multi-omics data from family-based WES, cis-eQTL, and GWAS, combined with SMR-HEIDI colocalization analysis and family-based variant validation, this study systematically dissected the genetic susceptibility mechanisms underlying scoliosis. We identified 31 common genes that link family-derived rare variants with population-level expression regulation signals, constructing a convergent genetic map connecting transcriptional regulation to scoliosis susceptibility^[22-24]. Among these, the non-coding RNA *LINC00299* and the protein-coding genes *NCAMI*, *AGPAT3*, *SMIM12*, and *CTSH* demonstrated key functional and regulatory potential, providing new insights into the pathogenesis of scoliosis.

The strong colocalization signal observed for *LINC00299* suggests that its expression level is inversely correlated with scoliosis risk, indicating an important role for this long non-coding RNA as a regulatory element in spinal morphogenesis^[25,26]. As a key non-coding regulatory gene identified in this study, dysregulation of *LINC00299* may interfere with the gene regulatory network governing normal spinal development, positioning it as a potential critical regulatory node mediating genetic susceptibility to scoliosis.

Beyond *LINC00299*, four protein-coding genes were also confirmed to participate in core biological processes relevant to scoliosis. *NCAMI* mediates neuronal adhesion and synaptic signaling, influencing muscle tone balance and the maintenance of spinal symmetry through the regulation of neuronal development and neuromuscular junction function^[27-29]. *AGPAT3* is involved in phospholipid biosynthesis and fatty acid remodeling, playing an important role in maintaining chondrocyte membrane integrity and regulating normal skeletal morphogenesis^[30-32]. *SMIM12* is a mitochondrial membrane protein that

regulates cellular energy metabolism and oxidative stress homeostasis; its expression imbalance may induce muscle metabolic disturbances and bilateral muscle tension asymmetry^[33]. *CTSH* encodes a lysosomal cysteine protease that participates in extracellular matrix degradation and bone remodeling, thereby regulating vertebral bone metabolism and spinal structural stability^[34,35].

Together, these genes cover multiple key biological dimensions including neural regulation, lipid and energy metabolism, muscle function, and bone remodeling, collectively forming a multi-system pathogenic regulatory network spanning the metabolic-neural-skeletal axis. This provides new molecular evidence and research directions for deciphering the complex multi-dimensional molecular pathogenesis of scoliosis. Recent genetic studies have identified susceptibility genes for scoliosis through genome-wide association analysis, and bidirectional Mendelian randomization investigations have further revealed causal associations of inflammatory cytokines and gut microbiota with scoliosis, supporting that scoliosis is regulated by multiple genetic and molecular factors^[36-38].

Functional enrichment analysis revealed that GO annotations of the candidate genes were mainly enriched in core functions such as neuronal synapse structure, energy coenzyme biosynthesis, and ion channel regulation, confirming their involvement in neural development and energy metabolism regulation at the levels of cellular component, biological process, and molecular function. KEGG pathway enrichment further highlighted significant enrichment in nicotinate and nicotinamide metabolism, insulin and glucagon signaling pathways, HIF-1 signaling, and the renin-angiotensin system, suggesting extensive interplay between endocrine-metabolic regulation and skeletal growth. The key genes identified in this study may link systemic metabolic regulation with spinal skeletal development by modulating metabolic homeostasis, energy substrate synthesis,

and neuromuscular signaling. These findings indicate that scoliosis is not a purely orthopedic disorder but rather a multi-organ developmental condition involving coordinated neural and metabolic regulation^[34,39].

Several limitations of this study should be acknowledged. First, the GWAS summary data were obtained from the FinnGen database, which predominantly includes individuals of Finnish European ancestry, and the eQTL data were also derived from whole-blood samples of European populations. Therefore, the generalizability of our conclusions to other ethnic groups requires further validation^[40]. Second, although SMR analysis can effectively identify causal relationships between gene expression and disease, it cannot completely exclude bias due to horizontal pleiotropy, and the power of the HEIDI test depends on the number of instrumental variables and the accuracy of the linkage disequilibrium structure^[10]. Third, the functional roles of the key genes identified in this study are primarily based on bioinformatics annotations and lack experimental validation in vitro or in vivo; their specific pathogenic mechanisms await further investigation. Finally, scoliosis exhibits high clinical heterogeneity, and the GWAS data used in this study were not stratified by different clinical subtypes (e.g., idiopathic, congenital, etc.), which may affect the fine mapping of our results.

Conclusion

This study integrated family-based multi-omics data and SMR-HEIDI colocalization analysis to identify 31 scoliosis susceptibility-related consensus genes, with *LINC00299* and four protein-coding genes emerging as critical regulatory and functional hubs. Our findings reveal that scoliosis arises from a complex neural – metabolic – skeletal interactive network, rather than being a simple orthopedic disease. This family-driven integrative genomic analysis expands the understanding of scoliosis genetic mechanisms, and provides novel candidate targets and a foundational framework for future mechanistic research and translational development.

Reference

- Liang ZT, Tang H, Li J, et al. IL6ST participates in the development of adolescent idiopathic scoliosis by regulating bone marrow mesenchymal stem cells. *Communications Biology*. 2025,8(1):1369-1369.
- Jia Q, Zhang B, Wang H, Zheng W. Effectiveness of physical therapeutic scoliosis exercise (PSSE) intervention for adolescent idiopathic scoliosis: a systematic review and meta-analysis. *BMC Musculoskeletal Disorders*. 2025,26(1):947-947.
- Tuncay IO, Lee EK, Gustafson A, et al. Whole genome sequencing in adolescent idiopathic scoliosis cohort implicates multiple biological pathways. *npj Genomic Medicine*. 2025,10(1):67-67.
- Zhao M, Zhou M, Han Y, et al. Evaluating the readability and quality of AI-generated scoliosis education materials: a comparative analysis of five language models. *Scientific Reports*. 2025,15(1):35454-35454.
- Yao S, Dingxin S, Jinqiao W, et al. Role of erythritol in coronary heart disease, ischemic stroke, and venous thromboembolism: A Mendelian randomization analysis. *Medicine*. 2025,104(43):e45187-e45187.
- Tan Y, Hu X, Zhang J, et al. Causal role of immune cells in allergic rhinitis: A bidirectional Mendelian randomization study. *Brazilian journal of otorhinolaryngology*. 2025,92(1):101726.
- Yun JJ, Wang JQ, Wang JH, et al. Causal Effect of Blood Cell Perturbation Phenotypes on Multiple Sclerosis via Immune Mediation: A Mendelian Randomization Study. *Cellular and molecular neurobiology*. 2025,45(1):93.
- Urmo V, Annique C, HarmJan W, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature genetics*. 2021,53(9):1300-1310.
- Kurki MI, Karjalainen J, Palta P, et al. Author Correction: FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023,615(7952):E19.
- Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016,48(5):481-487.
- Yan L. ggvenn: Draw Venn Diagram by 'ggplot2'. Version 0.1.16. R Foundation for Statistical Computing; 2025.
- Wickham H, François R, Henry L, et al. dplyr: A Grammar of Data Manipulation. Version 1.1.4. R Foundation for Statistical Computing; 2023.
- Wickham H, Vaughan D, Girlich M. tidyr: Tidy Messy Data. Version 1.3.1. R Foundation for Statistical Computing; 2024.
- LiLin-Yin. CMplot: Circle Manhattan Plot. Version 4.5.1. R Foundation for Statistical Computing; 2024.
- Ooms J. magick: Advanced Graphics and Image-Processing in R. Version 2.8.6. R Foundation for Statistical Computing; 2025.
- Snow G. TeachingDemos: Demonstrations for Teaching and Learning. Version 2.13. R Foundation for Statistical Computing; 2024.
- Zhou L, Chen W, Li J, Chen M. Integrated bioinformatics analysis reveals cross-talking hub genes and therapeutic agents between sepsis and acute myocardial infarction. *Ann Med*. 2026,58(1):2645276.
- Yu G. clusterProfiler: A Universal Enrichment Tool for Interpreting Omics Data. Version 4.19.3. R Foundation for Statistical Computing; 2025.
- Zhu Y, Qu X, Pan J, et al. Identification of the immune-related diagnostic biomarkers between Graves' disease and thyroid carcinoma based on comprehensive bioinformatics analysis and machine learning. *Autoimmunity*. 2026,59(1):2631208.

20. Kanehisa M, Furumichi M, Sato Y, et al. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 2021,49(D1):D545-d551.
21. Chen Z, Shen Z, Zhang Z, et al. RNA-Associated Co-expression Network Identifies Novel Biomarkers for Digestive System Cancer. *Front Genet.* 2021,12:659788.
22. Ren W, Zhang C, Wang H, et al. Identification of galangin as a therapeutic candidate for primary biliary cholangitis via systematic druggable genome-wide Mendelian randomization analysis and experimental validation. *Frontiers in Pharmacology.* 2025,16:1674693-1674693.
23. Wei Q, Yu S, Luo Y, et al. Exploring the causal relationship between plasma proteins and postherpetic neuralgia: a Mendelian randomization study. *Frontiers in Neurology.* 2025,16:1575941-1575941.
24. Zheng X, Hong Y, Wei W. Mendelian randomization and nomogram-based prediction of hepatocellular carcinoma risk in patients with hepatitis B cirrhosis. *PeerJ.* 2025,13:e20179.
25. Xiaodong L, Na Z, Caiping W, et al. LINC00299 polymorphisms rs891058, rs13395467, and rs13398375 reduce the risk of allergic rhinitis among the Chinese Han population. *Gene.* 2023,889:147794-147794.
26. Chang M, Liu G, Wang Y, et al. Long non-coding RNA LINC00299 knockdown inhibits ox-LDL-induced T/G HA-VSMC injury by regulating miR-135a-5p/XBP1 axis in atherosclerosis. *Panminerva Med.* 2022,64(1):38-47.
27. Shaw AD, Tiwari Y, Kaplan W, et al. Characterisation of genetic variation in ST8SIA2 and its interaction region in NCAM1 in patients with bipolar disorder. *PLoS ONE.* 2017,9(3):e92556.
28. Xu DH, Du JK, Liu SY, et al. Upregulation of KLK8 contributes to CUMS-induced hippocampal neuronal apoptosis by cleaving NCAM1. *Cell Death Dis.* 2023,14(4):278.
29. Chen Y, Liu N, Yang Y, et al. NCAM1 modulates the proliferation and migration of pulmonary arterial smooth muscle cells in pulmonary hypertension. *Respir Res.* 2024,25(1):435.
30. Su S, Zeng Y, Chen J, Dong X. AGPAT3 Regulates Immune Microenvironment in Osteosarcoma via Lysophosphatidic Acid Metabolism. *Oncology research.* 2025,34(1):27.
31. Zhou H, Fick K, Patel V, et al. AGPAT3 deficiency impairs adipocyte differentiation and leads to a lean phenotype in mice. *Am J Physiol Endocrinol Metab.* 2024,327(1):E69-e80.
32. Liu C, Hu C, Cheng J, et al. AGPAT3 reshapes tumor cell vulnerability to IFN γ -mediated ferroptosis and enhances immunotherapy efficacy through lipid remodeling. *J Immunother Cancer.* 2026,14(3).
33. Zhang M, Ning J, Liu J, et al. Peripheral blood immune landscape and NXPE3 as a novel biomarker for hypertensive intracerebral hemorrhage risk prediction and targeted therapy. *Imeta.* 2025,4(3):e70030.
34. Lin Y, Li J, Liang S, et al. Pan-cancer Analysis Reveals m6A Variation and Cell-specific Regulatory Network in Different Cancer Types. *Genomics, proteomics & bioinformatics.* 2024 Oct 15;22(4):qzae052.
35. Wu PY, Yu YL, Zhao WR, Zhou B. Identifying and Validating Extracellular Matrix-Related Gene CTSH in Diabetic Foot Ulcer Using Bioinformatics and Machine Learning. *J Inflamm Res.* 2024,17:5871-5887.
36. Dai Z, Wu Z, Xu L, et al. Genome-wide association study identifies GAK and KLF12 associated with curve severity of adolescent idiopathic scoliosis. *PeerJ.* 2026,14:e20638.
37. Mardan M, Mamat M, Yasin P, et al. Investigating the causal links between inflammatory cytokines and scoliosis through bidirectional Mendelian randomization analysis. *JOR spine.* 2024,7(4):e70019.
38. Bowen L, Heng J, Yuan G, Xuhui Z. Causal effects of gut microbiota on scoliosis: A bidirectional two-sample mendelian randomization study. *Heliyon.* 2023,9(11):e21654-e21654.
39. Ma H, Zhan S, Tang X, Yi P. Single-Segment Spinal Decompression and Fusion With Selective Nerve Root Block in Adult Degenerative Scoliosis: A Retrospective Comparative Study. *Orthop Surg.* 2026.
40. C MM, Charles R. A scientometric review of genome-wide association studies. *Communications biology.* 2019,2(1):9.

Author contributions

Xuanchen Li, Zhenrui Lin: Conceptualization, data curation, formal analysis, visualization, manuscript drafting. Xiaosheng Chen, Zhixiang Zhu, Rui Zhang: Data evaluation, Data collection, Data filtering. Weijun Wang, WenYu Zhou, Yi He, Lei Yang: Data analysis, result visualization, manuscript editing. Lihe Jiang, Guodong Huang and Qian Liang: Conceptualization, investigation, data curation, supervision, writing manuscript and review.

Funding

This work was supported by the Natural Science Foundation of Guangdong Province (2024A1515013174), Sanming Project of Medicine in Shenzhen (SZSM202211003), Shenzhen Key Medical Discipline Construction Fund (SZXK022), the National Natural Science Foundation Youth Project of China (82203807), Shenzhen-Hong Kong Jointly Funded Project, Shenzhen Science and Technology Program (SGDX20230116093645007), Shenzhen Medical Research Fund (B2303005), Development and Reform Commission of Shenzhen Municipality (S2002Q84500835), Shenzhen Basic Research Project (JCYJ20220530150810024), (JCYJ20240813141020027).

Acknowledgements

The authors would like to express sincere thanks to all participating schools and students of the study and all the rehabilitation therapists from Shenzhen Youth Spine Health Center for their help with the

scoliosis screening and data collection. Thanks to/Supported by Shenzhen Portion of Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone, project No. HTHZQSW-S-KCCYB-2023060.

SUPPORTING MATERIALS

Additional supplementary information is available for download and review in the supplementary information section located on the right-hand side of this article's HTML page.

Supplementary Table 1. xlsx