

RESEARCH ARTICLE

Open Access



Transcriptome-wide association study identifies new susceptibility genes and pathways for spondyloarthritis

Xiaochen Su¹, Anfa Chen¹, Menghao Teng¹, Wenchen Ji¹ and Yingang Zhang^{1*}

Abstract

Background Spondyloarthritis (SpA) is a group of multifactorial bone diseases influenced by genetic factors, the environment and lifestyle. However, current studies have found a limited number of SpA-related genes, and the genetic and pathogenic mechanisms of SpA are still unclear.

Methods A tissue-specific transcriptome-wide association study (TWAS) of SpA was performed using GWAS (including 3966 SpA patients and 448,298 controls) summary data and gene expression weights of whole blood and skeletal muscle. The SpA-associated genes identified by TWAS were further compared with the differentially expressed genes (DEGs) identified in the SpA gene expression profile acquired from the Gene Expression Omnibus database (GEO, GSE58667). Finally, functional enrichment and annotation analyses of the identified genes were performed.

Results The TWAS detected 499 suggestive genes associated with SpA in whole blood and skeletal muscle, such as *CTNNA1* ($P_{SM} = 3.04 \times 10^{-2}$, $P_{WB} = 9.58 \times 10^{-3}$). The gene expression profile of SpA identified 20 candidate genes that overlapped in the TWAS data, such as *MCM4* ($P_{TWAS} = 1.32 \times 10^{-2}$, $P_{DEG} = 2.75 \times 10^{-2}$) and *KIAA1109* ($P_{TWAS} = 3.71 \times 10^{-2}$, $P_{DEG} = 4.67 \times 10^{-2}$). Enrichment analysis of the genes identified by TWAS identified 93 significant GO terms and 33 KEGG pathways, such as mitochondrion organization (GO: 0007005) and axon guidance (hsa04360).

Conclusion We identified multiple candidate genes that were genetically related to SpA. Our study may provide novel clues regarding the genetic mechanism, diagnosis, and treatment of SpA.

Highlights

1. This is the first study to identify susceptibility genes associated with spondyloarthritis (SpA), revealing 499 suggestive genes by TWAS analysis.
2. The TWAS results were further verified by comparing the gene expression profiles.
3. This study provides novel clues for understanding the genetic mechanism of SpA, especially focusing on aberrant gene expression in the pathogenesis of SpA.

Keywords Spondyloarthritis, Transcriptome-wide association study, Differentially expressed gene, Pathway enrichment analysis, Gene ontology

*Correspondence:

Yingang Zhang

zyingang@mail.xjtu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Spondyloarthritis (SpA) is a group of related but phenotypically distinct disorders, including psoriatic arthritis, arthritis related to inflammatory bowel disease, reactive arthritis, juvenile idiopathic arthritis, and ankylosing spondylitis [1, 2]. The symptoms of SpA include inflammation of the axial skeleton, asymmetric peripheral arthritis of the lower limbs, enthesitis, and extra-articular manifestations, and all subtypes of SpA share a common genetic background [3, 4]. The prevalence of SpA shows considerable differences among ethnic groups and populations that can vary from 0.01% (Japan) to 2.5% (Alaska), and the annual estimated incidence of SpA is 62.5/100,000. The worldwide prevalence of SpA has seriously affected the function of the musculoskeletal system and reduced the quality of life [4, 5]. Recently, an increasing number of studies have focused on the genetic mechanisms of SpA [6, 7]. Through familial aggregation, previous studies have estimated that genetic risk factors contribute to 80–90% of the susceptibility to SpA such as ankylosing spondylitis [7].

Genome-wide association studies (GWASs) have been considered to be one of the primary approaches for determining genetic links to diseases [8, 9]. By a GWAS approach, Díaz-Peña et al. [10] revealed the potential involvement of mechanisms and pathways that were previously unsuspected in SpA, particularly those regarding aminopeptidases or the *IL23/IL17* pathways. However, the GWAS is only recommended for evaluating the risk of disease. Because most GWAS-identified single nucleotide polymorphisms (SNPs) are located in the noncoding regions of the genome, so the interpretation of those variants at the gene expression level is limited [11]. Expression quantitative trait loci (eQTL) analysis is a way to identify genes related to variations in gene expression [12]. Therefore, integrating GWAS and eQTL analysis may help more effectively to identify candidate genes associated with disease. In the previous study, researcher integrated publicly available GWAS summary data and eQTL reference data sets to evaluate gene-trait relationships, an approach referred to as the transcriptome-wide association study (TWAS) [13]. Different from GWAS, TWAS can drastically reduce the comparisons in statistical analysis and enhance the ability to detect candidate genes of target diseases [14].

In recent years, TWAS has been widely used to identify genetic loci associated with target diseases. For example, Liao et al. [15] identified 9 transcriptome-wide significant hits associated with attention deficit/hyperactivity disorder, of which 6 genes were not implicated in the original GWAS. In addition, Mancuso N et al. identified 217 genes at 84 independent 1 Mb regions associated with prostate cancer risk through TWAS analysis, which could

provide novel risk loci and prioritize putative causal genes at known risk loci associated with prostate cancer [16]. However, only few GWAS and no TWAS analyses focus on SpA. Therefore, we identify candidate genes to reveal the genetic and pathogenic mechanisms of SpA.

Materials and methods

SpA GWAS summary data set

The large-scale SpA GWAS summary data was obtained from a published study [17]. In short, the data set contained 3966 diagnosed SpA patients and 448,298 controls of European ethnicity from the UK Biobank (<https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20002>). Accessed 19 March 2021) (Additional file 1) [17]. The UK Biobank participants were genotyped using the Affymetrix UK Biobank AXIOM or UK Biobank AXIOM array and imputed against approximately 90 million genetic variations from the Haplotype Reference Consortium, 1000 Genomes and the UK 10 K Project [17]. After filtering, the data set contained 9,113,133 imputed variants. The IMPUTE4 program was used to perform the imputation (<http://jmarchini.org/software/>). Detailed information on the subjects, genotyping, imputation, and quality control can be found in a published study (Additional file 2) [17]. In addition, we performed additional quality controls on the above GWAS data: (1) removal of single nucleic acid polymorphisms (SNPs) without rsID; (2) uniform alignment with the hg19 human reference genome.

TWAS of SpA

The TWAS of SpA was carried out by functional summary-based imputation software (FUSION <http://gusevlab.org/projects/fusion/>). FUSION is a new approach to identify genes whose expression is significantly associated with complex traits in individuals without directly measuring the expression level by integrating GWAS summary data and precalculated gene expression weights of different tissues [18]. The whole blood and skeletal muscle tissues were also used in previous biological studies of SpA [19]. In this study, we used the pre-calculated gene expression weights of whole blood and skeletal muscle by the FUSION prediction models. Then, the calculated gene expression weights were combined with GWAS statistics to impute the association statistics between the gene expression level and SpA. FUSION software and the gene expression weight panels for skeletal muscle and whole blood were downloaded from the FUSION website (<http://gusevlab.org/projects/fusion/>). For TWAS results, a significant association threshold of $P < 5.38 \times 10^{-6}$ (0.05/9300) after strict Bonferroni correction was adopted. P values between $P < 5.38 \times 10^{-6}$ and 0.05 were considered suggestive of significance [20, 21].

Gene expression profile of SpA

The gene expression profile of SpA was acquired from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/GEO/>, accession number: GSE58667). Briefly, DNA microarray-based gene expression levels were examined in peripheral blood from 11 patients with juvenile SpA (jSpA) patients and 4 healthy controls from Croatia, along with a bioinformatics analysis of the retrieved data, and the carefully selected differentially expressed genes (DEGs) of all participants in the study were analyzed by qRT-PCR [22]. The GEO2R tool was used to identify the DEGs. GEO2R presents a simple interface that allows users to perform sophisticated R-based analysis of GEO data to identify and visualize differential gene expression [23]. Genes were identified as differentially expressed when the following two conditions were met: P value < 0.05 by the moderated t statistic and $|\log_2FC| > 1$ [22].

Gene annotation analysis and functional enrichment analysis

In this study, we used the Metascape (<https://metascape.org/gp/index.html>) tool to perform Gene Ontology (GO) and pathway enrichment analyses of candidate genes identified by the TWAS and gene expression profiles [24]. The significant GO terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were screened out by comparing the results of TWAS analysis and the gene expression profiles. Then, the functions of significant genes shared by TWAS analysis and gene expression profiles were annotated, prioritized, visualized, and

interpreted by the Functional Mapping and Annotation of GWAS (FUMA <https://fuma.ctglab.nl/>) tool [25].

Results

TWAS results of SpA

TWAS analysis identified 390 suggestive genes in skeletal muscle and 137 suggestive genes in whole blood, with P values between 5.38×10^{-6} and 0.05 (Fig. 1, Additional file 3: Table S1). In addition, 28 genes were identified in both skeletal muscle and whole blood, including cadherin-associated protein, alpha-like 1 (*CTNNA1*) ($P_{SM} = 3.04 \times 10^{-2}$, $P_{WB} = 9.60 \times 10^{-3}$), *AC000078.5* ($P_{SM} = 1.17 \times 10^{-2}$, $P_{WB} = 1.00 \times 10^{-4}$ 0.0001), *RP11-165J3.6* ($P_{SM} = 1.86 \times 10^{-2}$, $P_{WB} = 1.84 \times 10^{-3}$) and *ZNF100* ($P_{SM} = 3.00 \times 10^{-3}$, $P_{WB} = 2.80 \times 10^{-3}$). Table 1 presents detailed information on the top significant genes identified by TWAS, including the TWAS P value, TWAS Z score, and number of SNPs in the locus (NSNP).

Validating TWAS results by gene expression profiles of SpA

By comparing the genes identified by TWAS analysis and the gene expression profile of SpA, we screened 20 DEGs, such as *MCM4* ($P_{TWAS} = 1.32 \times 10^{-2}$, $P_{DEG} = 2.75 \times 10^{-2}$), *KIAA1109* ($P_{TWAS} = 3.71 \times 10^{-2}$, $P_{DEG} = 4.67 \times 10^{-2}$) and *SFMBT2* ($P_{TWAS} = 2.94 \times 10^{-2}$, $P_{DEG} = 2.45 \times 10^{-2}$) (Table 2). Using FUMA software, we found that 20 candidate genes were differentially expressed in the musculoskeletal system. In addition, the expression of common genes was significantly downregulated in muscle-skeletal tissue ($-\log_{10} P$ value > 4 , Fig. 2). The distribution of

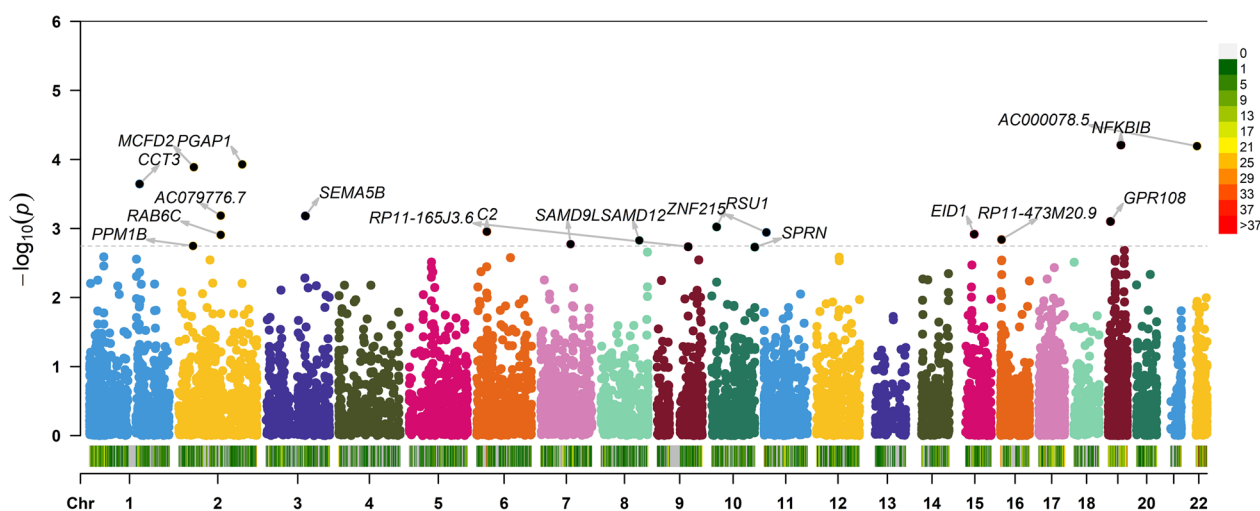


Fig. 1 Manhattan plot showing TWAS-identified genes. Note: Manhattan plot showing TWAS-identified genes and significantly expressed genes associated with SpA (annotated points). Each point represents a single gene, and the physical position (chromosome localization) is plotted on the x-axis, while the $-\log_{10}(P)$ value of the association between gene and SpA is plotted on the y-axis. TWAS: Transcriptome-wide association study; SpA: spondyloarthritis

Table 1 The common genes identified by TWAS analysis

Gene	CHR	Skeletal muscle			Whole blood		
		NSNP	TWAS.Z	TWAS.P _{SM}	NSNP	TWAS.Z	TWAS.P _{WB}
AC000078.5	22	551	-2.5198	6.42 × 10 ⁻²	514	-3.9967	6.42 × 10 ⁻⁵
RP11-165J3.6	9	493	-2.3532	1.86 × 10 ⁻²	473	-3.1153	1.86 × 10 ⁻³
ZNF100	19	298	2.9700	2.98 × 10 ⁻³	295	2.9885	2.80 × 10 ⁻³
ZNF493	19	286	2.8573	4.27 × 10 ⁻³	283	2.8464	4.42 × 10 ⁻³
ZNF429	19	299	2.8573	4.27 × 10 ⁻³	304	2.7375	6.19 × 10 ⁻³
STK17B	2	371	-2.1650	3.04 × 10 ⁻²	376	-2.7341	6.26 × 10 ⁻³
ZC3H3	8	360	3.0619	2.20 × 10 ⁻³	353	2.6965	7.01 × 10 ⁻³
ZNF738	19	285	2.2961	2.17 × 10 ⁻²	276	2.6729	7.52 × 10 ⁻³
CTNNAL1	9	622	2.1645	3.04 × 10 ⁻²	633	2.5905	9.58 × 10 ⁻³
GGTA1P	9	442	2.6587	7.84 × 10 ⁻³	445	-2.5057	1.22 × 10 ⁻²
RP11-254F7.2	2	579	-2.4359	1.49 × 10 ⁻²	555	-2.5043	1.23 × 10 ⁻²
ARFGAP3	22	531	2.1865	2.88 × 10 ⁻²	532	-2.4671	1.36 × 10 ⁻²
KLHL12	1	490	-2.7270	6.39 × 10 ⁻³	471	2.4202	1.55 × 10 ⁻²
SRR	17	409	-2.2354	2.54 × 10 ⁻²	406	-2.4130	1.58 × 10 ⁻²
RP11-611E13.2	12	575	2.4076	1.61 × 10 ⁻²	763	2.4076	1.61 × 10 ⁻²
NT5C3B	17	409	-2.3002	2.14 × 10 ⁻²	399	-2.3818	1.72 × 10 ⁻²
MBLAC1	7	281	-2.0665	3.88 × 10 ⁻²	265	-2.3050	2.12 × 10 ⁻²
BNIP1	5	579	-2.1862	2.88 × 10 ⁻²	566	-2.2589	2.39 × 10 ⁻²
LRRC61	7	348	2.0121	4.42 × 10 ⁻²	344	2.2517	2.43 × 10 ⁻²
CCDC125	5	235	-2.2450	2.48 × 10 ⁻²	250	-2.2383	2.52 × 10 ⁻²
CHD1 L	1	407	-2.5622	1.04 × 10 ⁻²	427	2.2280	2.59 × 10 ⁻²
RP11-218M22.1	12	457	-2.0853	3.70 × 10 ⁻²	452	-2.2221	2.63 × 10 ⁻²
HLA-DQA1	6	214	-2.0896	3.67 × 10 ⁻²	210	-2.1745	2.97 × 10 ⁻²
ZNF205	16	385	-2.0686	3.86 × 10 ⁻²	363	-2.1201	3.40 × 10 ⁻²
POLR2J3	7	228	-2.2285	2.59 × 10 ⁻²	228	-2.0773	3.78 × 10 ⁻²
RAPGEFL1	17	295	-2.1276	3.34 × 10 ⁻²	290	-2.0455	4.08 × 10 ⁻²
NUDCD3	7	352	-2.4406	1.47 × 10 ⁻²	371	-2.0091	4.45 × 10 ⁻²
AC091729.9	7	400	2.0595	3.95 × 10 ⁻²	378	2.0062	4.48 × 10 ⁻²

The SpA GWAS summary data set and the pre-calculated reference weights of gene expression profiles in whole blood and skeletal muscle were used for TWAS analysis of SpA. Each TWAS.P value was calculated by TWAS analysis (<http://gusevlab.org/projects/fusion/>)

TWAS: Transcriptome-wide association study; GWAS: genome-wide association study; SpA: spondyloarthritis; TWAS.P_{SM}: TWAS P_{skeletal-muscle} value; TWAS.P_{WB}: TWAS P_{Whole-Blood} value; TWAS.Z: TWAS Z score; NSNP: number of SNPs in the locus; FUSION, functional summary-based imputation

Table 2 The common genes identified by both TWAS analysis and gene expression profiles of SpA

Tissue	Gene	CHR	P _{TWAS}	TWAS.Z	P _{DEG}	Regulation
Skeletal muscle	MCM4	8	3.37 × 10 ⁻²	-2.1232	1.61 × 10 ⁻²	Up
	ATRNL1	10	4.34 × 10 ⁻²	2.0195	2.56 × 10 ⁻²	Down
	CWF19L2	11	4.06 × 10 ⁻²	2.0481	4.69 × 10 ⁻²	Down
	CYB5A	18	1.83 × 10 ⁻²	2.3589	3.82 × 10 ⁻²	Down
	ZNF738	19	2.17 × 10 ⁻²	2.2961	3.07 × 10 ⁻²	Down
	KIAA1109	4	3.71 × 10 ⁻²	2.0843	4.67 × 10 ⁻²	Down
Whole blood	SFMBT2	10	2.94 × 10 ⁻²	-2.1775	2.45 × 10 ⁻²	Up
	FRA10AC1	10	4.00 × 10 ⁻²	2.0535	1.49 × 10 ⁻²	Down
	SAMD10	20	2.24 × 10 ⁻²	-2.2829	2.20 × 10 ⁻²	Up
	SMG5	1	4.30 × 10 ⁻³	-2.8549	5.00 × 10 ⁻³	Up

Each P_{TWAS} value was calculated by TWAS Analysis. Each P_{DEG} value was the DEG derived from published studies

TWAS: Transcriptome-wide association study; DEG: differentially expressed gene; P_{TWAS}: P_{Transcriptome-Wide Association Study} value; P_{DEG}: P_{Differentially Expressed Gene} value; NSNP: number of SNPs in the locus

gene expression values in the gene expression profiles was visualized in the corresponding volcano plot (Fig. 3).

Gene annotation analysis and functional enrichment

In this study, pathway and process enrichment analyses were carried out based on the following ontology resources: GO biological processes, KEGG pathways, GO molecular functions, Reactome gene sets, canonical pathways, and CORUM. The 528 genes identified by TWAS in the two tissues were successfully submitted to Metascape to perform gene enrichment analysis. The Metascape tool identified 93 GO terms that were enriched in SpA, such as mitochondrion organization (GO:0007005), generation of precursor metabolites and energy (GO:0006091), and histone H4-R3 methylation (GO:0043985). We also identified 33 KEGG pathways in SpA, such as axon guidance (hsa04360) and purine metabolism (hsa00230)

(Additional file 4: Table S2). The significant terms were then hierarchically clustered, a subset of representative terms was selected, and the results were converted into a network layout (Fig. 4). Combined with Fig. 4 and Additional file 4: Table S2, the results indicated that the most significant biological pathways were antigen processing and presentation, the adaptive immune system, and mitochondrion organization. The GO Chord, Sankey and dot plots illustrating the top representative and overrepresented GO terms belonging to SpA are shown in Fig. 5A and B.

Discussion

SpA is a group of complex trait diseases, and genetic susceptibility accounts for a large part of its pathogenesis. Previous studies have identified a large number of genetic loci associated with SpA, but their ability to interpret the

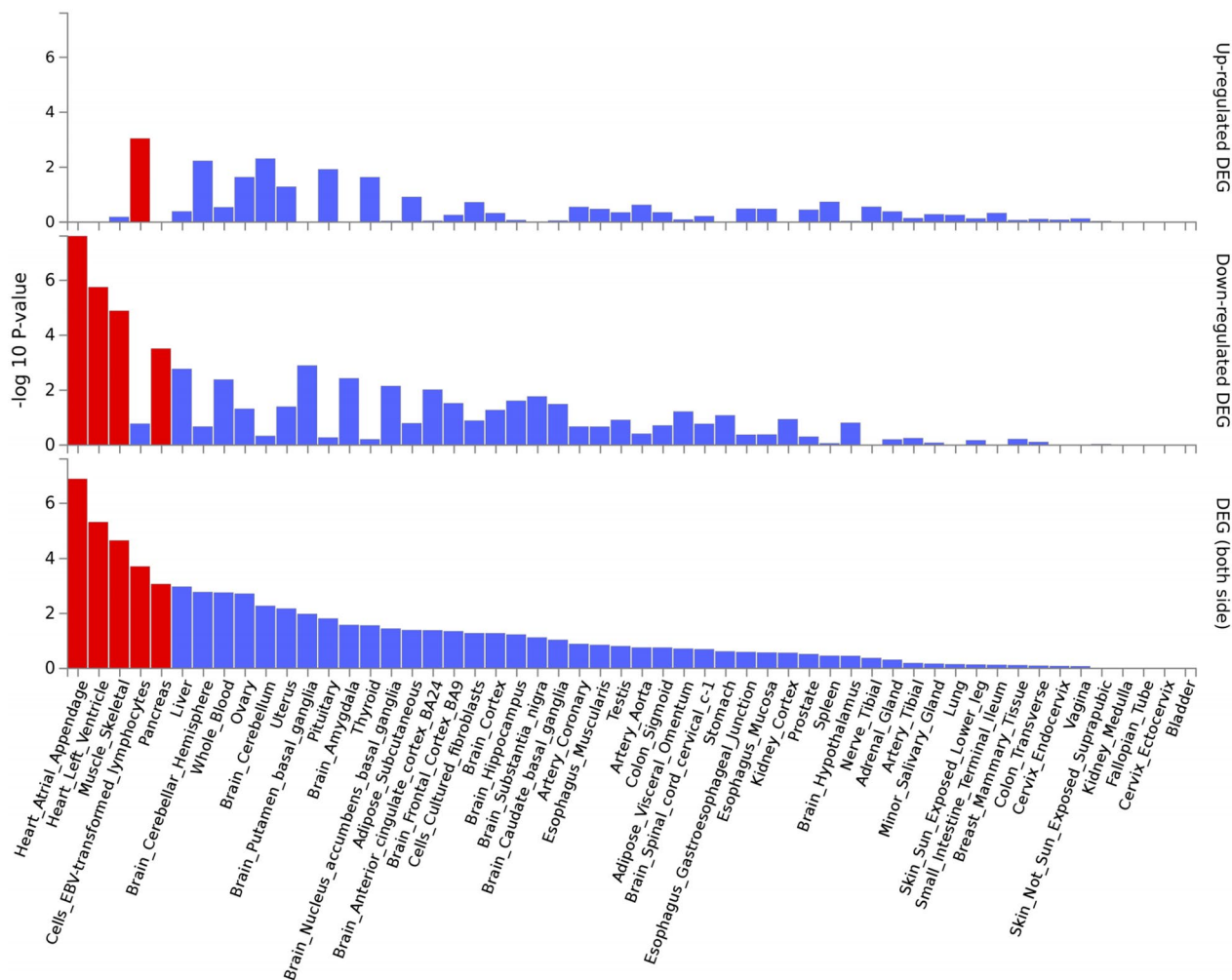


Fig. 2 The expression of common spondyloarthritis (SpA) genes at different tissue sites. Note: Image of the expression of common genes identified by both TWAS analysis and gene expression profiling. The bars in red show significant differential expression. TWAS, Transcriptome-wide association study; SpA, spondyloarthritis

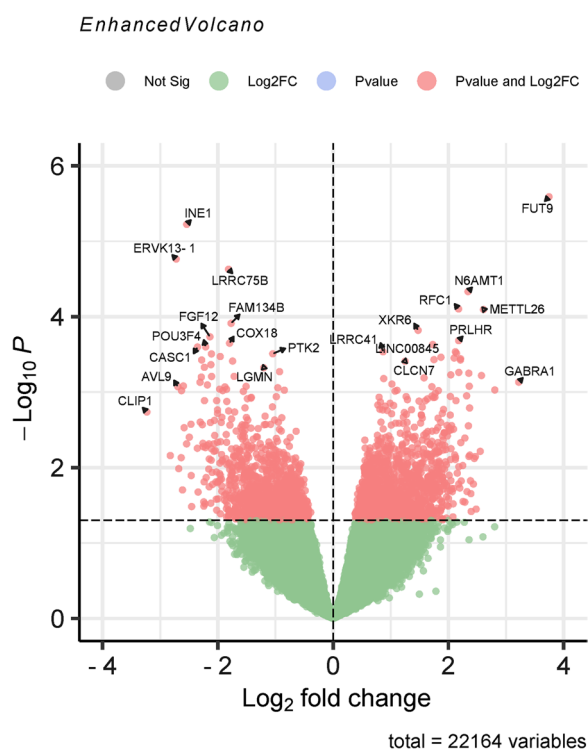


Fig. 3 Volcano plot of gene expression profiles for spondyloarthritis (SpA). Note: The results of gene expression profiling were output in a volcano map. Genes marked in red were differentially expressed according to the following two conditions: P value of < 0.05 by the moderated t statistic and $|\log_2FC| > 1$. DEG: Differentially expressed gene; SpA: spondyloarthritis

relationship between the significant genes and SpA was limited [10]. To learn more about the potential genetic mechanisms of SpA, we conducted a TWAS analysis based on the large-scale GWAS summary data set acquired from a European cohort including 3966 SpA patients and 448,298 controls. Our results revealed a total of 499 potential disease-associated genes, including 390 suggestive genes identified in skeletal muscle and 137 suggestive genes identified in whole blood. This is the first study to identify genes associated with SpA by TWAS analysis.

The TWAS identified several genes associated with SpA, such as *SFMBT2*, *MCM4*, *KIAA1109*, and *CTNNAL1*. The *SFMBT2* protein is a member of the polycomb group (PcG) of proteins. Hussain et al. [26] found that *SFMBT2* interference altered the expression of key metabolic genes in chondrocytes; *SOX9* and *COL2A1* were decreased, whereas *MMP13* and *ADAMTS4* were significantly increased. Some studies have shown that upregulation or downregulation of these genes, which are altered by the *SFMBT2* gene, can lead to cartilage degeneration and further cause SpA [27, 28]. The *ATRNL1* gene, which was identified in this study, can regulate the

expression of *SOX9* and is significantly highly expressed in the cartilage tissues of patients with osteoarthritis [29]. In summary, *SFMBT2* and *ATRNL1* may be associated with the genetic mechanisms of SpA pathogenesis. *SFMBT2* and *ATRNL1* may be candidate genes related to SpA, and the genetic mechanisms linking these genes and SpA need further research.

Another important candidate gene identified in this study was *MCM4*. The *MCM4* protein is a DNA replication licensing factor that is essential for DNA replication initiation and elongation in eukaryotic cells. In other words, the *MCM4* gene acts as an essential regulator of the cell cycle. *MCM4* can cause many diseases by regulating the cell cycle and inducing apoptosis [30, 31]. In addition, previous studies have shown that the pathogenesis of SpA includes thinning of the cartilage and cartilage degeneration, which involves chondrocyte apoptosis and proteoglycan loss [31].

In addition, *KIAA1109* was identified in both the TWAS and the gene expression profile of SpA. A study showed that a 480 kb block on chromosome 4q27 encompassing the *KIAA1109/Tenr/IL-2/IL-21* gene cluster is associated with rheumatoid arthritis [32]. Zhernakova et al. [33] found that the *KIAA1109/Tenr/IL-2/IL-21* gene cluster is involved in susceptibility to multiple autoimmune diseases, suggesting that this locus is a general risk factor for diseases such as rheumatoid arthritis and celiac disease. SpA is a subset of seronegative rheumatic-related autoimmune diseases, and Bowes et al. found significant evidence associating SpA susceptibility with the *IL-2* and *IL-21* genes [34, 35]. To the best of our knowledge, no researchers studied whether *KIAA1109* has a direct effect on SpA. Therefore, this is the first study exploring the genetic correlation between *KIAA1109* and SpA.

CTNNAL1 is ubiquitously expressed in many tissues, including skeletal muscle, the pancreas, and the heart [36]. By comparing psoriasis patients who did not have psoriatic arthritis and patients with psoriatic arthritis, Patrick et al. [37] identified significant loci overlapping the regulatory elements that encompass genes that are differentially expressed in differentiated osteoblasts, including genes that participate in Wnt signaling, such as *RUNX1*, *FUT8*, and *CTNNAL1*. The proteins in Wnt/ β -catenin signaling play essential roles in the development of SpA. Xie et al. [38] found that Wnt proteins were essential in normal bone homeostasis, particularly in osteoblastic new bone formation. Therefore, Wnt proteins may also play roles in new bone formation in ankylosing spondylitis, and various Wnt signaling molecules were shown to be involved in maintaining bone mass [39]. In summary, these results provide new clues for future studies on the genetic mechanism of *CTNNAL1*.

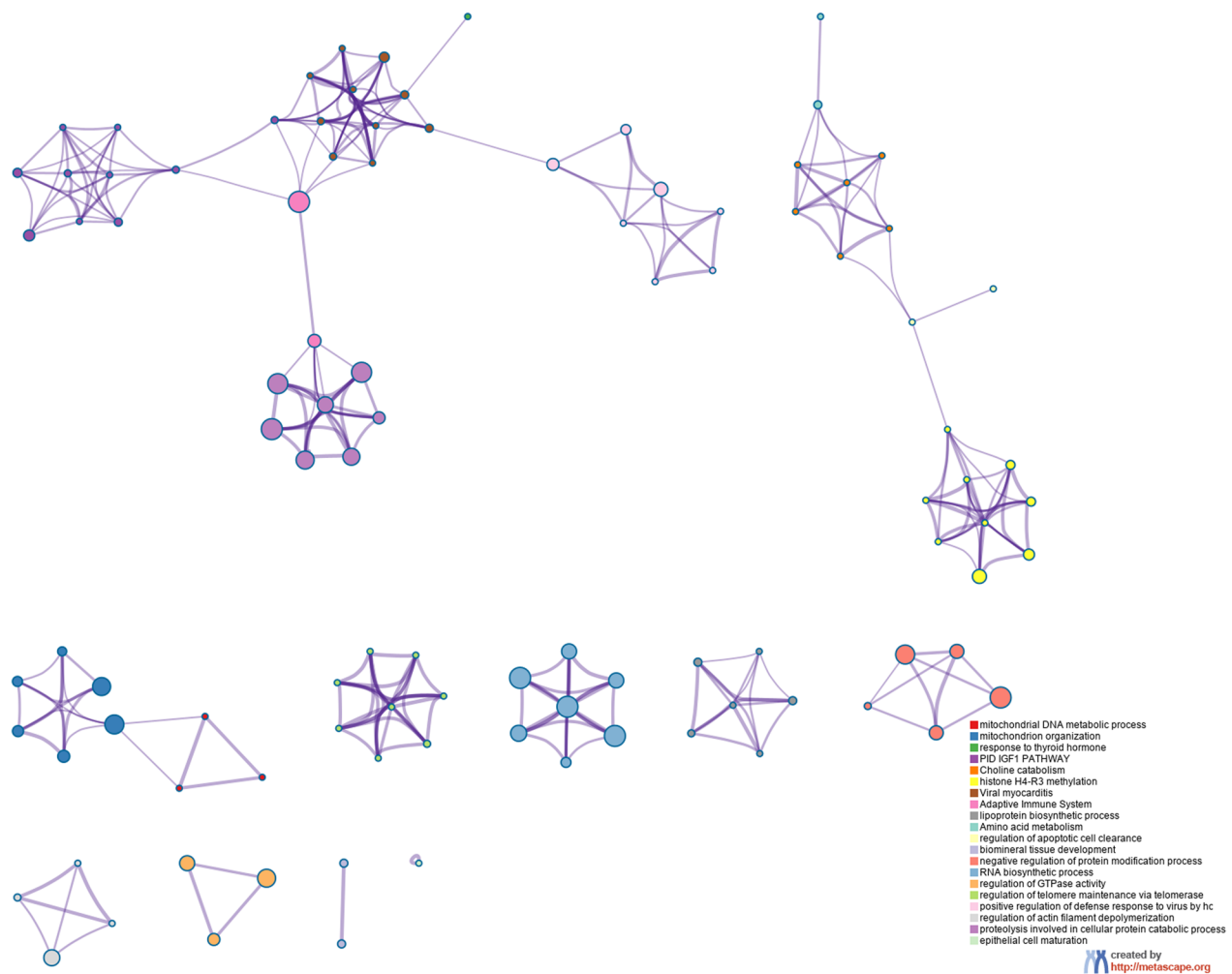


Fig. 4 The network layout of representative Gene ontology (GO) terms. Note: The network layout of representative GO terms under hierarchical clustering. In the network, each circle node represents a term, where node size is proportional to the number of input genes associated with that term, and node color represents cluster identity (i.e., nodes of the same color belong to the same cluster). Terms with a similarity score >0.3 are linked by an edge (the thickness of the edge represents the similarity score). GO: Gene ontology

In our study, GO enrichment analysis and KEGG pathway analysis were also conducted to explore the functions of candidate genes and how they are distributed in SpA. For example, mitochondrion organization (GO: 0007005) was identified by both TWAS analysis and gene expression profiling. Cytochrome c is primarily known for its role in the mitochondria as a key participant in the life-supporting function of ATP synthesis [40]. Recently, researchers found that cytochrome c can interact with protease, which leads to the activation of the apoptosis protease activation factor [41]. It has been demonstrated that this biological signal is responsible for apoptosis and activation of the inflammatory process during the pathogenesis of psoriatic arthritis [41]. Overall, these findings suggested that

abnormal mitochondrial organization may play a role in the pathogenesis of SpA.

Axon guidance (KEGG: hsa04360) was also identified as enriched in SpA. Recently, the semaphorin family was originally identified as axonal guidance molecules, and semaphorins affect the pathogenesis of multiple types of arthritis by regulating immunity, angiogenesis, bone remodeling, apoptosis, cell migration and invasion [42, 43]. In addition, the semaphorin family can regulate the biological pathway of *TNF-α/ADAMTS-4*, and blocking semaphorins can decrease cartilage and bone destruction, cell infiltration into the synovium, and the production of *TNF-α* and *IL-6* [44]. The TWAS analysis and gene expression profile of SpA identified axon guidance

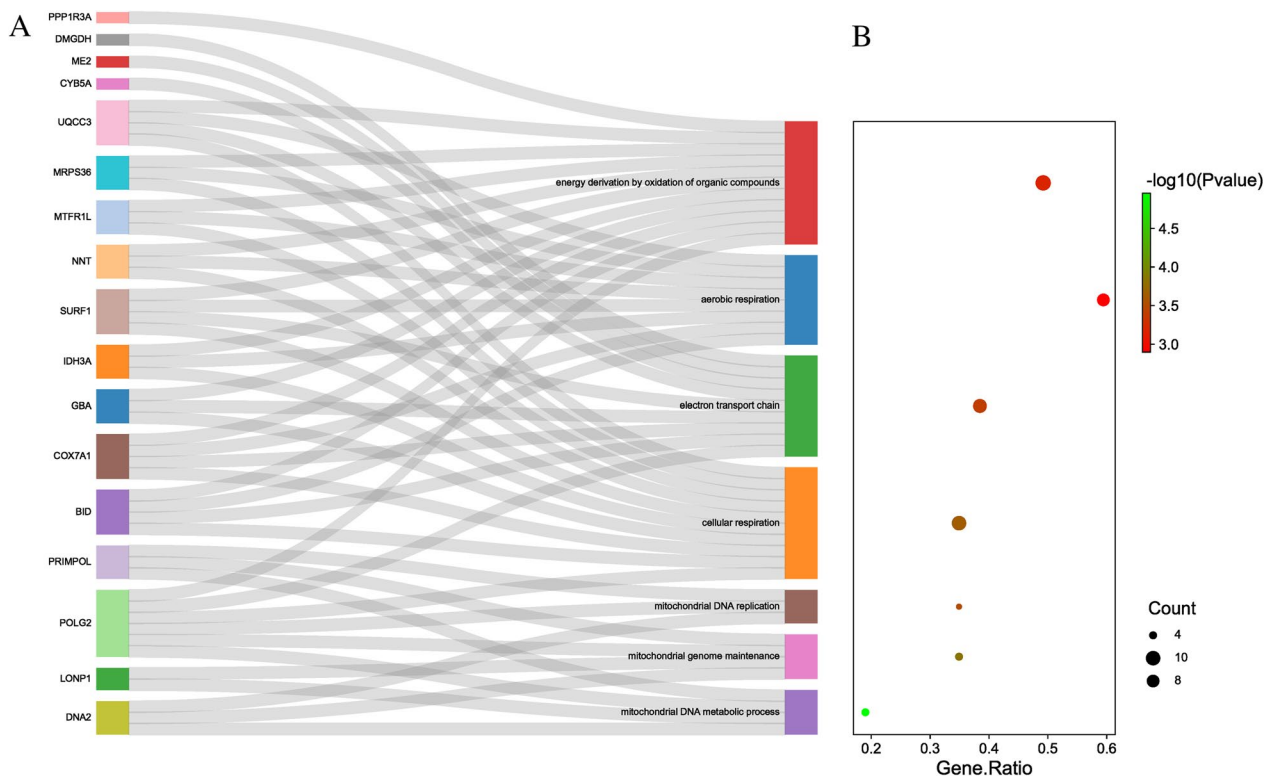


Fig. 5 Top representative and overrepresented Gene ontology (GO) terms and related genes. **A** Note: The GO Chord plot of top overrepresented GO terms belonging to the Biological Process subontology for spondyloarthritis (SpA). The genes are linked to their assigned terms via colored ribbons. Genes are ordered according to the observed log FC, which is displayed in descending and ascending intensity of red and blue squares displayed next to the selected genes. GO: Gene ontology; SpA: spondyloarthritis; log FC: log fold change. **B** Note: Sankey plot showing the relationships between the genes and overrepresented GO terms. The dot plot shows the ratio between the genes found to be involved in GO terms and the total number of genes included in each GO term (FDR $P \leq 0.05$). GO: Gene ontology

as a susceptibility pathway associated with SpA, which was consistent with existing research.

In addition, we found that some common genes identified in skeletal muscle and whole blood tissue showed different risk directions. For example, the Z values of *GGTA1P* in skeletal muscle and whole blood were opposite each other. This difference potentially can be explained in two aspects. First, they may be attributed to the biological variations in gene expression across different tissue types. For instance, studies have reported that in patients with type 2 diabetes, *GLUT4* is selectively downregulated only in adipose tissue, while its expression level remains relatively high in skeletal muscle [45]. This also explains why our TWAS results (sampled from skeletal muscle and peripheral blood) do not perfectly overlap with the expression profile results (sampled from intervertebral discs). Second, it has been reported that TWAS exhibits bias in expression profiles from non-trait-related tissues [13]. Therefore, further investigation of the functional roles and mechanisms of these genes across different tissue types is necessary to better explain these disparities.

After strict Bonferroni correction, the significance threshold of the TWAS results was $P < 5.38 \times 10^{-6}$. P values between $P < 5.38 \times 10^{-6}$ and 0.05 were considered suggestive of significance. Unfortunately, according to our results, the TWAS results showed only suggestive associations with SpA. Because reference panel size affects the P value. We should focus more on effective size (Z score) instead of P value. Although $P < 5.38 \times 10^{-6}$ and 0.05 were considered to be suggestive of significance, these genes are still valuable.

One of the strengths of TWAS lies in its ability to accurately prioritize genes likely to be causal while excluding non-causal genes [13]. Our study aimed to enhance the accuracy of our results through three approaches. Firstly, we used multi-tissues TWAS in order to produce more accurate results than single-tissue counterparts [18]. Secondly, we performed TWAS analysis based on the latest GWAS summary data for SpA. The large sample size of the GWAS summary data ensured the precision of our research findings. Finally, we validated the candidate genes by comparing them

with the gene expression profile, which substantially improved the credibility of our TWAS analysis results.

This study also has some limitations. First, the GWAS summary data are based on those with European ancestry and may not apply to other ancestry studies. Therefore, caution should be taken when applying our results to other populations. Further TWAS analysis of other populations is needed to verify our results. Second, to validate the TWAS results, we compared the significant genes identified by TWAS analysis of SpA with the gene expression profile of jSpA, but jSpA is one subtype of SpA. Our results should be interpreted with caution. Further biological studies should be conducted to confirm our findings.

Conclusions

In summary, based on GWAS summary data, a TWAS analysis identified novel and common susceptibility genes for SpA. This study not only provides novel clues for understanding the genetic mechanism of SpA but also provides a basis for further experiments. In addition, beyond specific mechanistic findings for SpA, this work outlines a systematic approach for identifying functional mediators of complex-trait diseases.

Abbreviation

DEG	Differentially expressed genes
eQTL	Expression quantitative trait loci
FUSION	Functional summary-based imputation
FUMA	Functional Mapping and Annotation of Genome-wide Association Study
GEO	Gene expression omnibus
GO	Gene ontology
GWAS	Genome-wide association study
KEGG	Kyoto Encyclopedia of Genes and Genomes
TWAS	Transcriptome-wide association study
SNP	Single nucleotide polymorphisms
SpA	Spondyloarthritis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13018-023-04029-4>.

Additional file 1: Risk Loci in GWAS Dataset of SpA.

Additional file 2: Quality Control Steps.

Additional file 3: Table S1 Top genes selected by transcriptome-wide association study (TWAS) analysis

Additional file 4: Table S2 GO terms identified by Metascape enriched for SpA

Acknowledgements

Not applicable.

Author contributions

ZYG had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. SXC designed this study and wrote this manuscript. CAF and TMH were responsible for data

collection. Data analysis and interpretation were performed by SXC and TMH. Revision of the manuscript and data analysis were performed by SXC and JWC. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (81371987 and 81171761), Natural Science Basic Research Foundation of Shaanxi (2021SF-247), Postdoctoral Science Foundation of China (2020M670073ZX) and Shaanxi Provincial Administration of Traditional Chinese Medicine (2021-04-ZZ-003). Financial support had no impact on the outcomes of this study.

Availability of data and materials

The datasets analyzed during the current study are available from the Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/gds>) accession number: GSE58667. The GWAS summary data of SpA obtained from UK biobank (https://nealelab.github.io/UKBB_ldsc/downloads.html).

Declarations

Ethics approval and consent to participate

The source of the data was a publicly available database, and no human participants were involved; hence, ethical parameters are not applicable.

Consent for publication

Not applicable.

Competing interests

All authors declare that they have no competing interests.

Author details

¹Department of Orthopaedics of the First Affiliated Hospital, Medical School, Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China.

Received: 4 July 2022 Accepted: 19 July 2023

Published online: 04 September 2023

References

- Gulino GR, Van Mechelen M, Lories R. Cellular and molecular diversity in spondyloarthritis. *Semin Immunol.* 2021;58:101521.
- Sharip A, Kunz J. Understanding the pathogenesis of spondyloarthritis. *Biomolecules.* 2020;10(10):1461.
- Mauro D, Thomas R, Guggino G, et al. Ankylosing spondylitis: an autoimmune or autoinflammatory disease? *Nat Rev Rheumatol.* 2021;17(7):387–404.
- Wang R, Ward MM. Epidemiology of axial spondyloarthritis: an update. *Curr Opin Rheumatol.* 2018;30(2):137–43.
- Chandran V, Leung YY. Psoriatic arthritis: current status. *Best Pract Res Clin Rheumatol.* 2021;35(2):101703.
- Rahmati S, Li Q, Rahman P, et al. Insights into the pathogenesis of psoriatic arthritis from genetic studies. *Semin Immunopathol.* 2021;43(2):221–34.
- Wang H, Guo J, Jiang J, et al. New genes associated with rheumatoid arthritis identified by gene expression profiling. *Int J Immunogenet.* 2017;44(3):107–13.
- Tam V, Patel N, Turcotte M, et al. Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* 2019;20(8):467–84.
- Saurabh R, Fouodo CJK, König IR, et al. A survey of genome-wide association studies, polygenic scores and UK Biobank highlights resources for autoimmune disease genetics. *Front Immunol.* 2022;13:972107.
- Díaz-Peña R, Castro-Santos P, Durán J, et al. The genetics of spondyloarthritis. *J Pers Med.* 2020;10(4):151.
- Friedman NP, Banich MT, Keller MC. Twin studies to GWAS: there and back again. *Trends Cogn Sci.* 2021;25(10):855–69.
- Sheng X, Qiu C, Liu H, et al. Systematic integrated analysis of genetic and epigenetic variation in diabetic kidney disease. *Proc Natl Acad Sci USA.* 2020;117(46):29013–24.

13. Wainberg M, Sinnott-Armstrong N, Mancuso N, et al. Opportunities and challenges for transcriptome-wide association studies. *Nat Genet.* 2019;51(4):592–9.
14. Pan C, Ning Y, Jia Y, et al. Transcriptome-wide association study identified candidate genes associated with gut microbiota. *Gut Pathog.* 2021;13(1):74.
15. Liao C, Laporte AD, Spiegelman D, et al. Transcriptome-wide association study of attention deficit hyperactivity disorder identifies associated genes and phenotypes. *Nat Commun.* 2019;10(1):4450.
16. Mancuso N, Gayther S, Gusev A, et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat Commun.* 2018;9(1):4079.
17. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet.* 2018;50(11):1593–9.
18. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* 2016;48(3):245–52.
19. Zhang CL, Li YC, Wu JW, et al. Expression and function of peripheral blood miRNA16a in patients with ankylosing spondylitis. *Eur Rev Med Pharmacol Sci.* 2018;22(16):5106–13.
20. Xu J, Ma J, Zeng Y, et al. Transcriptome-wide association study identifies novel genes associated with bone mineral density and lean body mass in children. *Endocrine.* 2023;79(2):400–9.
21. Qian Y, Xia J, Liu KQ, et al. Observational and genetic evidence highlight the association of human sleep behaviors with the incidence of fracture. *Commun Biol.* 2021;4(1):1339.
22. Lamot L, Borovecki F, Tambic Bukovac L, et al. Aberrant expression of shared master-key genes contributes to the immunopathogenesis in patients with juvenile spondyloarthritis. *PLOS ONE.* 2014;9(12):e115416.
23. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.* 2013;41(2):991–5.
24. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019;10(1):1523.
25. Watanabe K, Taskesen E, van Bochoven A, et al. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017;8(1):1826.
26. Hussain S, Sun M, Min Z, et al. Down-regulated in OA cartilage, SFMBT2 contributes to NF- κ B-mediated ECM degradation. *J Cell Mol Med.* 2018;22(11):5753–8.
27. Song H, Park KH. Regulation and function of SOX9 during cartilage development and regeneration. *Semin Cancer Biol.* 2020;67(Pt 1):12–23.
28. Nakamura A, Rampersaud YR, Sundararajan K, et al. Zinc finger protein-440 promotes cartilage degenerative mechanisms in human facet and knee osteoarthritis chondrocytes. *Osteoarthritis Cartil.* 2021;29(3):372–9.
29. Zhu J, Fu Q, Shao J, et al. Regulating effect of Circ_ATRNL1 on the promotion of SOX9 expression to promote chondrogenic differentiation of hAMSCs mediated by MiR-145-5p. *J Tissue Eng Regen Med.* 2021;15(5):487–502.
30. Ding X, Duan H, Luo H. Identification of core gene expression signature and key pathways in colorectal cancer. *Front Genet.* 2020;11:45.
31. Maguire A, Chen X, Wisner L, et al. Enhanced DNA repair and genomic stability identify a novel HIV-related diffuse large B-cell lymphoma signature. *Int J Cancer.* 2019;145(11):3078–88.
32. Teixeira VH, Pierlot C, Migliorini P, et al. Testing for the association of the KIAA1109/Tenr/IL2/IL21 gene region with rheumatoid arthritis in a European family-based study. *Arthritis Res Ther.* 2009;11(2):R45.
33. Zhernakova A, Alizadeh BZ, Bevova M, et al. Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. *Am J Hum Genet.* 2007;81(6):1284–8.
34. Bowes J, Ho P, Flynn E, et al. Comprehensive assessment of rheumatoid arthritis susceptibility loci in a large psoriatic arthritis cohort. *Ann Rheum Dis.* 2012;71(8):1350–4.
35. Tsukazaki H, Kaito T. The role of the IL-23/IL-17 pathway in the pathogenesis of spondyloarthritis. *Int J Mol Sci.* 2020;21(17):6401.
36. Wu D, Jiang W, Liu C, et al. CTNNL1 participates in the regulation of mucus overproduction in HDM-induced asthma mouse model through the YAP-ROCK2 pathway. *J Cell Mol Med.* 2022;26(5):1656–71.
37. Patrick MT, Stuart PE, Raja K, et al. Integrative approach to reveal cell type specificity and gene candidates for psoriatic arthritis outside the MHC. *Front Genet.* 2019;10:304.
38. Xie W, Zhou L, Li S, et al. Wnt/ β -catenin signaling plays a key role in the development of spondyloarthritis. *Ann N Y Acad Sci.* 2016;1364(1):25–31.
39. Soós B, Szentpétery Á, Raterman HG, et al. Effects of targeted therapies on bone in rheumatic and musculoskeletal diseases. *Nat Rev Rheumatol.* 2022;18(5):249–57.
40. Santucci R, Sinibaldi F, Cozza P, et al. Cytochrome c: an extreme multifunctional protein with a key role in cell fate. *Int J Biol Macromol.* 2019;136:1237–46.
41. Chimenti MS, Sunzini F, Fiorucci L, et al. Potential role of cytochrome c and tryptase in psoriasis and psoriatic arthritis pathogenesis: focus on resistance to apoptosis and oxidative stress. *Front Immunol.* 2018;9:2363.
42. Avouac J, Pezet S, Vandebeuque E, et al. Semaphorins: from angiogenesis to inflammation in rheumatoid arthritis. *Arthritis Rheumatol.* 2021;73(9):1579–88.
43. Fard D, Tamagnone L. Semaphorins in health and disease. *Cytokine Growth Factor Rev.* 2021;57:55–63.
44. Yoshida Y, Ogata A, Kang S, et al. Semaphorin 4D contributes to rheumatoid arthritis by inducing inflammatory cytokine production: pathogenic and therapeutic implications. *Arthritis Rheumatol.* 2015;67(6):1481–90.
45. Chadt A, Al-Hasani H. Glucose transporters in adipose tissue, liver, and skeletal muscle in metabolic health and disease. *Pflugers Arch.* 2020;472(9):1273–98.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

