ORIGINAL ARTICLE

Unveiling novel susceptibility genes for sarcoidosis by a cross-tissue transcriptome-wide association study

XIQIAO SUN

Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

ABSTRACT

Background: Sarcoidosis is a systemic granulomatous disease with heterogeneous clinical manifestations and unclear pathogenesis. Although genome-wide association studies (GWAS) have identified several immune-related loci, the functional interpretation of these signals remains limited. Integrative transcriptome-wide approaches may uncover novel susceptibility genes and provide mechanistic insights.

Objectives: The primary objective of this study was to identify novel susceptibility genes for sarcoidosis and provide mechanistic insights into its pathogenesis through a comprehensive, cross-tissue transcriptome-wide approach.

Methods: We conducted a cross-tissue transcriptome-wide association study (TWAS) using the UTMOST framework, followed by single-tissue TWAS via FUSION. Candidate genes identified in both analyses were further evaluated using Multi-marker Analysis of Genomic Annotation (MAGMA), Mendelian randomization (MR), and Bayesian colocalization. Functional characterization was explored through gene–chemical–disease associations from the Comparative Toxicogenomics Database (CTD) and phenome-wide association studies (PheWAS) in the UK Biobank.

Results: Cross-tissue TWAS identified 48 genes. Integrative analyses prioritized three novel susceptibility genes: RNF215, PLCL1, and RFTN2, and validated a previously reported susceptibility gene: FAM117B. MR and colocalization convergent genetic evidence of RNF215 (risk-increasing), FAM117B and RFTN2 (protective), and tissue-dependent effects for PLCL1. CTD analyses identified interactions of these genes with environmental chemicals including bisphenol A and tetrachlorodibenzodioxin, while PheWAS analyses revealed pleiotropic associations for RNF215, PLCL1, and RFTN2, whereas FAM117B did not show significant phenotype associations.

Received: 7 October 2025 | Accepted: 7 March 2026

Correspondence: Xiqiao Sun, MD / Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China /
E-mail: sunxq27@mail2.sysu.edu.cn
ORCID: 0009-0003-3173-1367

Conclusions: This comprehensive integrative study identifies four biologically plausible susceptibility genes for sarcoidosis, expanding the genetic architecture of sarcoidosis and suggests potential targets for mechanistic and therapeutic investigation.

Key words: sarcoidosis, transcriptome-wide association study, gene expression, susceptibility genes

Introduction

Sarcoidosis is a systemic granulomatous disorder of unknown etiology, affecting multiple organs—most frequently the lungs and lymph nodes—with variable prevalence across different populations and regions. Although generally considered rare, its global disease burden is increasing, particularly in regions such as Andean Latin America and among elderly males, as evidenced by the Global Burden of Disease (GBD) Study (2021) reporting 43 million prevalence cases and 40 million DALYs worldwide (1–3). Its annual incidence ranges from 10–15 cases per 100,000 in Northern Europe and North America to below 1 per 100 000 in Eastern Asia and the Southern Hemisphere (4,5). The clinical presentation of sarcoidosis is highly heterogeneous and may include pulmonary symptoms such as cough and dyspnea (>90% cases), as well as systemic manifestations like fever, fatigue, weight loss, and extrapulmonary involvement including skin, eyes, heart, nervous system, liver, spleen, and kidneys (6–8). The disease course may range from spontaneous remission to chronic progression, which can culminate in organ fibrosis or failure, significantly impairing quality of life—and carrying a mortality risk of approximately 1–7% (7). Recurrent serious infections and related morbidity further exacerbate patient outcomes, with sarcoidosis patients showing nearly a threefold increased risk of serious infection compared with the general population (9). Genetic predisposition contributes substantially to sarcoidosis susceptibility, as highlighted by GWAS identifying associations near immune-related loci. Notably, a susceptibility locus at chromosome 11q13 (CCDC88B) has been validated in European populations (10,11). Additional GWAS have implicated genes like ANXA11, IL23R, and others involved in T-cell signaling and autophagy pathways, underscoring the crucial role of immune regulation in pathogenesis (12,13). However, GWAS primarily identifies non-coding variants and provides limited insight into causal genes or biological mechanisms (14). Furthermore, most GWAS signals reside in regulatory regions, making it difficult to pinpoint which gene is functionally affected (15). These limitations highlight the need for integrative approaches that can link genetic variation to gene expression and disease phenotype more directly. Given the complexity of sarcoidosis—its pathogenesis involving genetic predisposition, immune dysregulation, and environmental triggers—there is a compelling need for integrative genomic approaches that bridge variants to gene expression and disease outcomes. Transcriptome-wide association studies have emerged as a powerful approach to connect genome-wide association study findings with gene function by integrating gene expression quantitative trait loci (eQTL) data with genome-wide association study summary statistics (16,17). This strategy allows the identification of genes whose genetically regulated expression levels are associated with disease risk, thereby providing functional insights beyond the interpretation of individual single nucleotide polymorphisms (SNPs). Among the available methods, the unified test for molecular signatures (UTMOST) utilizes expression data across multiple tissues to enhance statistical power and capture shared regulatory effects while maintaining sensitivity to tissue-specific signals (18). Compared with single-tissue approaches, cross-tissue transcriptome-wide association studies have shown greater accuracy in gene prioritization, which is particularly valuable in complex disorders such as sarcoidosis, where both systemic and organ-specific mechanisms are involved. In this study,

we conducted a comprehensive integrative analysis to identify candidate susceptibility genes for sarcoidosis. We first performed cross-tissue TWAS using the UTMOST, followed by single-tissue TWAS using the Functional Summary-based Imputation (FUSION). Genes reaching significance in both analyses were further evaluated using gene-level association testing via Multi-marker Analysis of Genomic Annotation (MAGMA), and causality was assessed through Mendelian randomization (MR). To determine whether shared causal variants underlie both gene expression and disease risk, we conducted Bayesian colocalization analysis. Furthermore, we explored chemical compounds associated with the candidate genes using the Comparative Toxicogenomics Database (CTD), and carried out a phenome-wide association study (PheWAS) to examine potential pleiotropic effects across diverse phenotypes. This integrative pipeline provides a systematic framework for discovering and characterizing novel risk genes for sarcoidosis.

Materials and methods

Figure 1 illustrates the analysis process.

GWAS data source

Summary-level genome-wide association study (GWAS) data for sarcoidosis were obtained from the FinnGen consortium (release R12, November 2024; https://r12.finngen.fi/). The dataset included 497,722 individuals of European ancestry, comprising 5,411 cases and 492,311 controls. Sarcoidosis

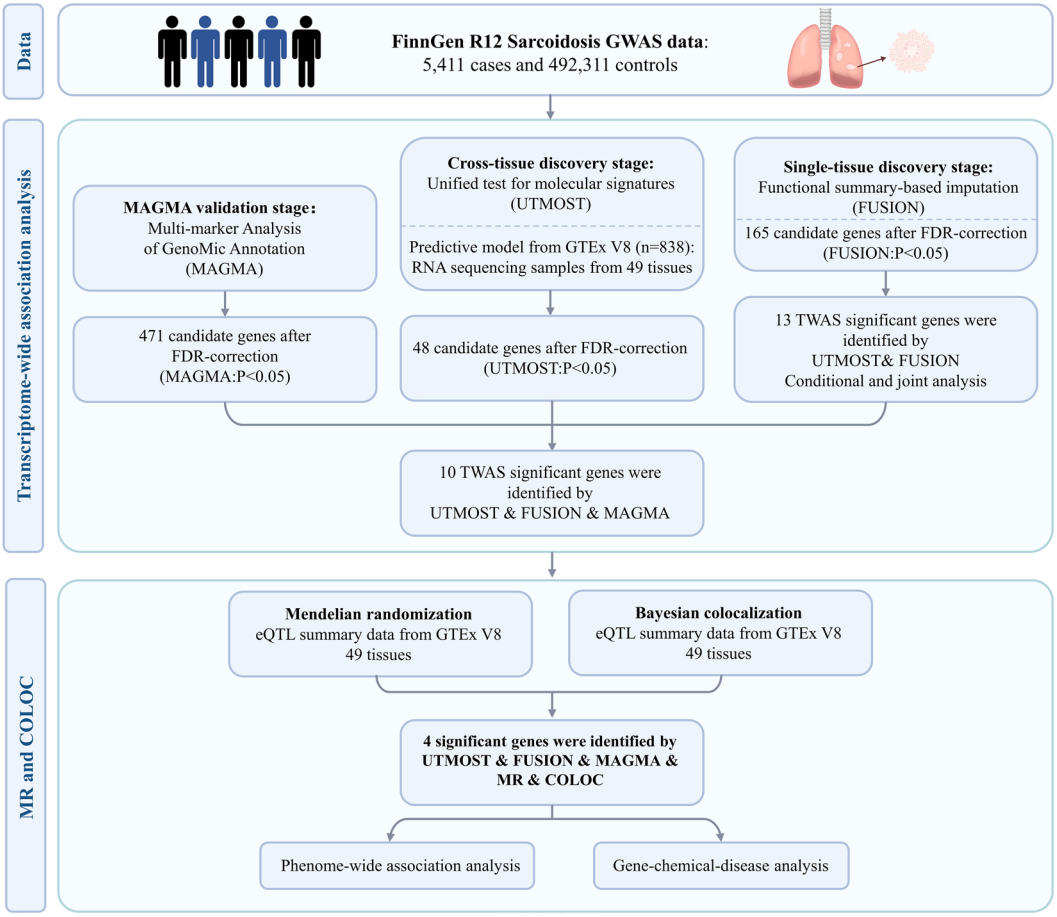

Figure 1. Flowchart of the study design.

cases were defined based on ICD-coded endpoints in the FinnGen datasets (ICD-10, D86). Detailed information on organ-specific involvement, such as pulmonary versus extrapulmonary manifestations (e.g., cardiac, cutaneous or ocular sarcoidosis), was not available in the summary-level data used in this study (19).

eQTL files source

Expression quantitative trait loci (eQTL) reference data were obtained from the Genotype-Tissue Expression (GTEx) Project, version 8 (https://gtexportal.org/home/) (20). This dataset includes genotype and gene expression profiles from 838 postmortem donors across 49 human tissues. Sample sizes varied by tissue, ranging from 73 samples in the renal cortex to 706 samples in skeletal muscle.

TWAS analyses in cross-tissue

Cross-tissue transcriptome-wide association studies (TWAS) were conducted using UTMOST (https://github.com/Joker-Jerome/UTMOST) to investigate gene–sarcoidosis associations at the organismal level. UTMOST applies penalized multivariate regression to build cross-tissue expression imputation models, accounting for heterogeneity in eQTL directions and effect sizes. This approach enhances imputation accuracy and improves gene discovery, particularly in tissues enriched for trait heritability (18). In total, 22,695 genes were analyzed for their associations with sarcoidosis across various tissues. Gene–trait associations across tissues were integrated using the generalized Berk–Jones (GBJ) test (21), incorporating covariance from single-tissue statistics. Statistical significance was defined as false discovery rate (FDR) < 0.05 (22). BH correction was applied to the aggregated p-values across all tested genes and all 49 tissues included in the output summary file; genes for which test statistics could not be computed by the UTMOST pipeline were not included in FDR correction.

TWAS analyses in single tissue

Single-tissue TWAS was performed using FUSION (http://gusevlab.org/projects/fusion/), which integrates sarcoidosis GWAS data with GTEx v8 eQTL reference data across 49 tissues (16). Linkage disequilibrium (LD) between prediction models and SNPs at each GWAS locus was estimated using samples from the 1000 Genomes European reference population. FUSION evaluates multiple predictive models for estimating SNP-derived gene expression weights, selecting the model with the best predictive performance, the number of tested genes per tissue ranged from approximately 22,000 to 23,000, depending on the tissue-specific reference expression models. TWAS associations were then obtained by integrating GWAS Z-scores with these gene weights. Candidate genes were retained if they met both criteria: (1) FDR < 0.05 in cross-tissue TWAS and (2) FDR < 0.05 in at least one tissue in single-tissue TWAS (23). For single-tissue TWAS, results were first aggregated across the 49 tissues, each comprising results from all 22 autosomes, and BH correction was then applied to the combined set of gene-level p-values. To refine associations, we applied the conditional and joint (COJO) module in FUSION to identify independent genetic signals at each locus while accounting for LD (16,24). LD was estimated using the European ancestry panel from the 1000 Genomes Project. A window size of 100 kb was applied. Jointly significant genes were identified through an iterative conditional analysis that selects independent genetic signals based on marginal Z statistics and LD r^2, following the default settings of the FUSION.post_process.R script. All analyses were performed using the default parameters recommended by the FUSION developers.

Gene analysis

Gene-level analyses were conducted using MAGMA (version 1.10) with default parameters to aggregate SNP-level association statistics into gene scores, thereby quantifying each gene's association with sarcoidosis. A total of 20,137 genes were included in this analysis, allowing for a comprehensive assessment of gene associations. BH correction was directly applied to the gene-level p-values reported in the MAGMA output files. Gene location files and European reference data were obtained from the MAGMA website (https://cncr.nl/research/magma/) (25,26).

Mendelian randomization and bayesian colocalization

Mendelian randomization (MR) analyses were performed using the "TwoSampleMR" R package (27) to assess potential causal relationships between genetically predicted gene expression and sarcoidosis. Cis-eQTL variants within 1 Mb of each gene were used as instrumental variables (IVs), gene expression served as the exposure, and sarcoidosis GWAS summary statistics were used as the outcome. Independent IVs were selected as SNPs reaching genome-wide significance ($p < 5 \times 10^{-8}$), with LD clumping applied ($r^2 < 0.001$, 10,000 kb window) (28). SNPs with minor allele frequency < 0.01 were excluded. When only a single independent IV was available, causal effects were estimated using the Wald ratio method, with significance set at $p < 0.05$. Bayesian colocalization analysis was conducted using the "coloc" R package (29) to determine whether GWAS and eQTL signals at a given locus shared a common causal variant. Colocalization analysis was performed with default priors ($p^1 = 1\times10^{-4}$, $p^2 = 1\times10^{-4}$, $p^{12} = 1\times10^{-5}$). For each gene, variants located within the gene body and ±1 Mb flanking regions were included. The same genomic window and prior settings were applied consistently across all genes and tissues. Five posterior probabilities (PP) were estimated: H0, no association; H1, only trait 1 associated; H2, only trait 2 associated; H3, both traits associated with different causal variants; and H4, both traits associated with the same causal variant. Following the literature, we defined colocalization when PPH4 > 0.8 and moderate colocalization when PPH4 > 0.5.

Gene-chemical-disease analysis

To explore chemical drugs associated with the candidate genes (RNF215, PLCL1, FAM117B, and RFTN2), searches were performed in the Comparative Toxicogenomics Database (https://ctdbase.org/) (30) for the candidate genes mentioned above. The CTD is a comprehensive public database that compiles extensive data on chemical-gene/protein interactions, as well as associations between chemicals and diseases, and between genes and diseases. This resource enables the prediction of the environmental impact on diseases by curating data on gene-chemical-disease interactions.

PheWAS analysis

To explore the broader phenotypic consequences of candidate genes associated with sarcoidosis, we conducted a phenome-wide association study (PheWAS) using the AstraZeneca PheWAS Portal (https://azphewas.com/) (31). This platform integrates exome sequencing and phenotype data from approximately 484,111 UK Biobank participants, enabling systematic evaluation of associations between genetic variants and 15,042 binary and quantitative phenotypes.

Results

TWAS analyses in cross-tissue

In the cross-tissue association analysis using UTMOST, we first evaluated 49 individual tissues separately. We then performed cross-tissue analyses with the generalized Berk–Jones (GBJ) method and applied p-value adjustments. In total, 48 genes ($P_{FDR} < 0.05$) were identified through UTMOST for further investigation (Table S1,2).

TWAS analyses in single-tissue

The single-tissue TWAS using FUSION identified 165 genes that reached significance ($P_{FDR} < 0.05$) in at least one tissue (Table S3). Among them, 13 genes also achieved significance in the cross-tissue TWAS, comprising 12 protein-coding genes (RNF215, RFTN2, PLCL1, SF3A1, MOB4, ICA1L, ANKRD36BP2, WDR12, NBEAL1, SEC14L2, FAM117B, and SF3B1) and one non-coding gene (ZNF14 pseudogene). To determine independent association signals, COJO analysis was performed on the candidate genes identified in the FUISON. Significant overlaps between cross-tissue and single-tissue results were observed on chromosomes 1, 2, and 22. Accordingly, COJO analysis was conducted on the FUSION results within their respective tissues for these three chromosomes (Table S4). After adjusting for linkage disequilibrium, 12 genes remained jointly significant. In contrast, the signal for SF3A1 was no longer significant after conditioning, suggesting that its association may be driven by nearby genes in LD (Figure 2).

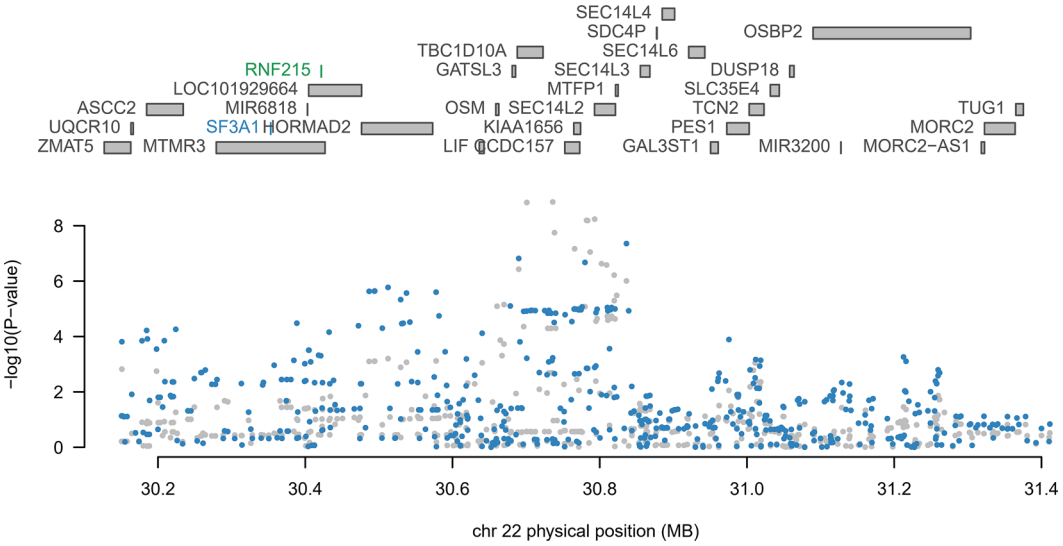

Figure 2. Conditional and joint analysis (COJO) analysis of SF3A1 in adipose visceral omentum tissue was performed using FUISON result to identify independent association signals. The scatter plot shows genetic variants across the chr22 region, gene annotations within the region are displayed at the top. Genes jointly significant after conditioning are highlighted in green, while marginally associated genes are shown in blue.

Gene analysis of MAGMA

To further prioritize candidate genes associated with sarcoidosis, we conducted a gene-level association analysis using MAGMA. After FDR correction, 471 genes reached statistical significance ($P_{FDR} < 0.05$) (Table S5). Among these, 10 genes overlapped with those identified in cross-tissue and single-tissue TWAS, including RFTN2, RNF215, PLCL1, SF3A1, MOB4, ICA1L, NBEAL1, SEC14L2, FAM117B, and SF3B1, reinforcing their potential roles in sarcoidosis susceptibility (Figure 3, Table S6).

Mendelian randomization

To assess the potential causal role of gene expression in sarcoidosis, we performed Mendelian randomization (MR) analyses using the Wald ratio method, focusing on five candidate genes (RNF215, PLCL1, FAM117B, RFTN2, and MIER1) with available cis-eQTLs. All five genes demonstrated statistically significant associations in at least one tissue ($p < 0.05$), suggesting possible causal effects (Figure 4, Table S7).

Increased expression of RNF215 in testis (OR, 1.57; 95% CI, 1.35–1.83) and MIER1 in artery aorta (OR, 1.27; 95% CI, 1.10–1.46) showed positive

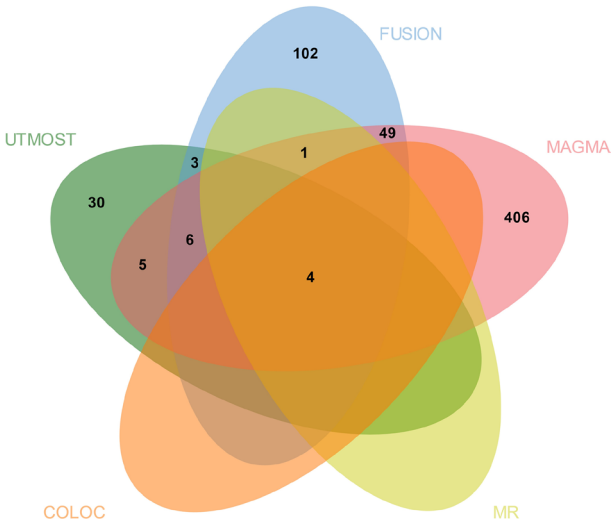

Figure 3. Venn diagram showing 4 common significant genes identified by cross-tissue TWAS (UTMOST), single-tissue TWAS (FUSION), gene-based association analysis (MAGMA), Mendelian randomization (MR), and colocalization analysis (COLOC). Genes were considered significant based on the corresponding method-specific significance thresholds.

associations with sarcoidosis risk. In contrast, higher expression of FAM117B in pancreas, spleen, and whole blood (OR, 0.75; 95% CI, 0.68–0.82), as well as RFTN2 in ovary (OR, 0.67; 95% CI, 0.57–0.78), was

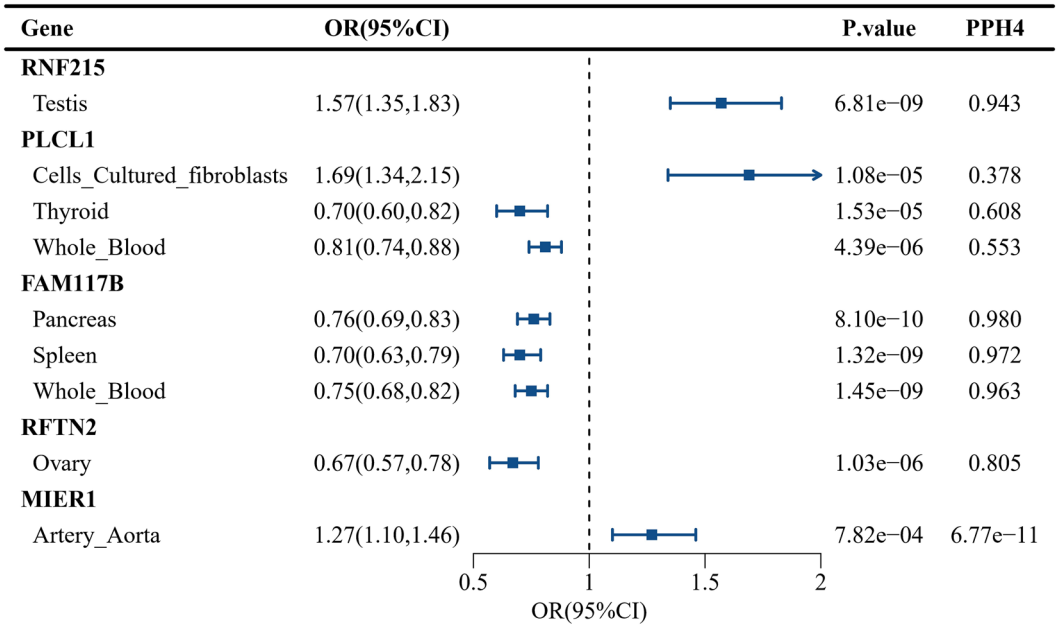

Figure 4. Mendelian Randomization forest plot showing the estimated causal effects of genetically predicted gene expression on sarcoidosis risk across different tissues. Effect estimates are presented as odds ratio (OR) with 95% confidence interval (CI), and PPH4 represents the posterior probability of a shared causal variant between gene expression and sarcoidosis based on colocalization analysis.

negatively associated with disease susceptibility. Interestingly, PLCL1 exhibited bidirectional effects, with its expression in cultured fibroblasts (OR, 1.69; 95% CI, 1.34–2.15) positively associated with sarcoidosis, while its expression in the thyroid and whole blood (OR, 0.81; 95% CI, 0.74–0.88) was negatively associated with the disease.

Colocalization of eQTL and GWAS associations

To further evaluate whether the observed associations between gene expression and sarcoidosis were driven by shared causal variants, we performed colocalization analyses. The results demonstrated strong evidence of colocalization (PPH4 > 0.8) for RNF215, FAM117B, and RFTN2, supporting a shared genetic basis between eQTL signals and sarcoidosis risk. For PLCL1, colocalization evidence was moderate in the thyroid and whole blood (PPH4 > 0.5), whereas no colocalization was observed in cultured fibroblasts (PPH4 < 0.5). In contrast, MIER1 did not show evidence of colocalization across tested tissues (PPH4 < 0.5), suggesting its MR associations may be influenced by linkage disequilibrium with nearby variants rather than shared causal variants (Figure 5). To facilitate comparison across analytical frameworks, we summarized the convergent genetic evidence for the four prioritized genes across TWAS, MAGMA, colocalization, and MR analyses in Table 1.

Gene-chemical-disease analysis results

Gene-chemical-disease analyses show that the gene RNF215 has significant interactions with multiple chemicals, with tetrachlorodibenzodioxin and bisphenol A being the primary interactors. The gene FAM117B likewise shows strong interactions with bisphenol A and tetrachlorodibenzodioxin. In the PLCL1 gene, interactions with benzo (a)pyrene and tetrachlorodibenzodioxin are most prominent; additionally, associations with several drugs such as acetaminophen and atrazine were found. Finally, the gene RFTN2 displays main interactions with valproic acid and bisphenol A, suggesting that the interplay between these genes and various chemicals may

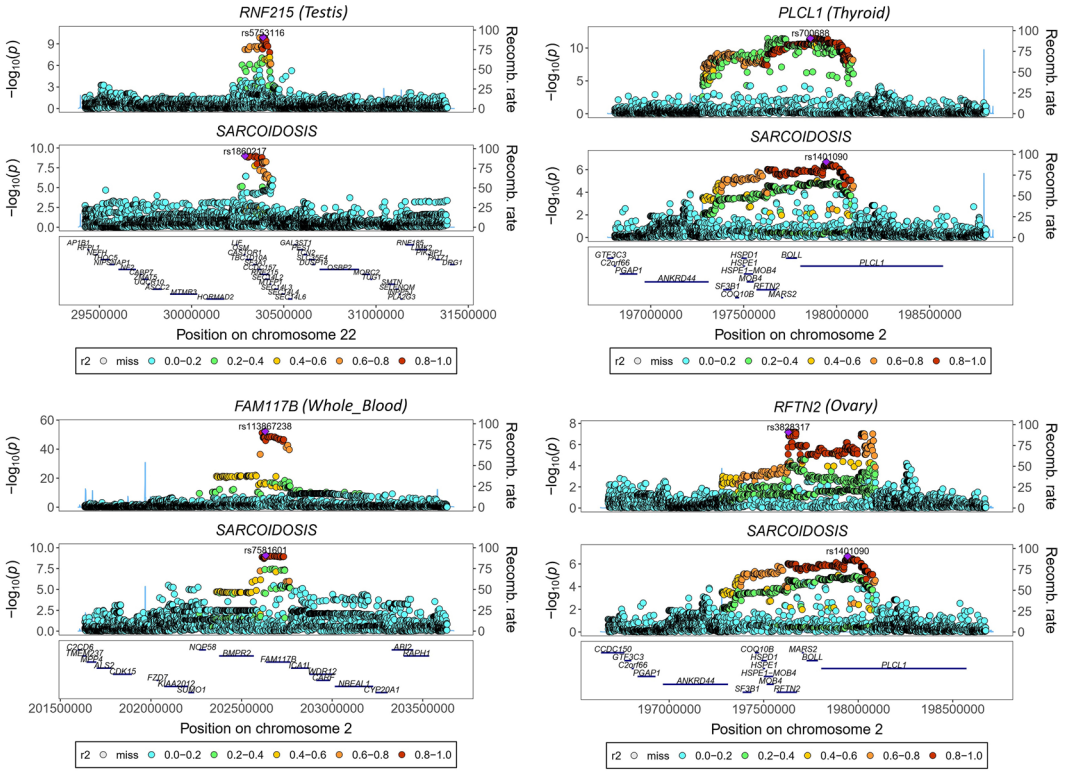

Figure 5. Stacked regional plots showing genetic colocalization between sarcoidosis and eQTL signals from four candidate genes. Upper panels display eQTL association signals, with colors indicating linkage disequilibrium (LD) r^2 values relative to the lead variant. Lower panels show sarcoidosis GWAS association signals.

Table 1. Summary of convergent genetic evidence for prioritized sarcoidosis susceptibility genes. The table summarizes cross-tissue TWAS results (UTMOST), the most significant single-tissue TWAS signals (FUSION), gene-based association results (MAGMA), Mendelian randomization (MR) estimates, and colocalization probabilities (PPH4) for the four prioritized genes.

Gene	UTMOST (cross-tissue) P_FDR	FUSION (top tissue) P_FDR	MAGMA (gene-based) P_FDR	MR (most significant tissue; OR [95% CI], P)	PPH4
RNF215	2.87E-06	7.24E-05 (Brain_Caudate_basal_ganglia)	4.95E-05	1.57 (1.35,1.83), 6.81E-09	0.943
PLCL1	3.97E-05	0.012 (Thyroid)	4.85E-05	0.81 (0.74,0.88), 4.39E-06	0.553
FAM117B	2.19E-05	6.05E-05 (Pancreas)	6.74E-08	0.76 (0.69,0.83), 8.10E-10	0.980
RFTN2	6.05E-05	0.016 (Pituitary)	5.78E-05	0.67 (0.57,0.78), 1.03E-06	0.805

represent potential environmental contributors to the pathogenesis of sarcoidosis (Figure 6).

Furthermore, the investigation of the Gene-Chemical-Disease interactions identified through CTD indicate that RNF215, PLCL1, FAM117B, and RFTN2 genes show significant interactions with various chemicals and are associated with a range of diseases including weight loss, hepatomegaly, chemical and drug-induced liver injury, inflammation, necrosis, hyperplasia, cognitive disorders, and more, highlighting their potential roles in the pathological mechanisms of nodular diseases (Tables S8–S11).

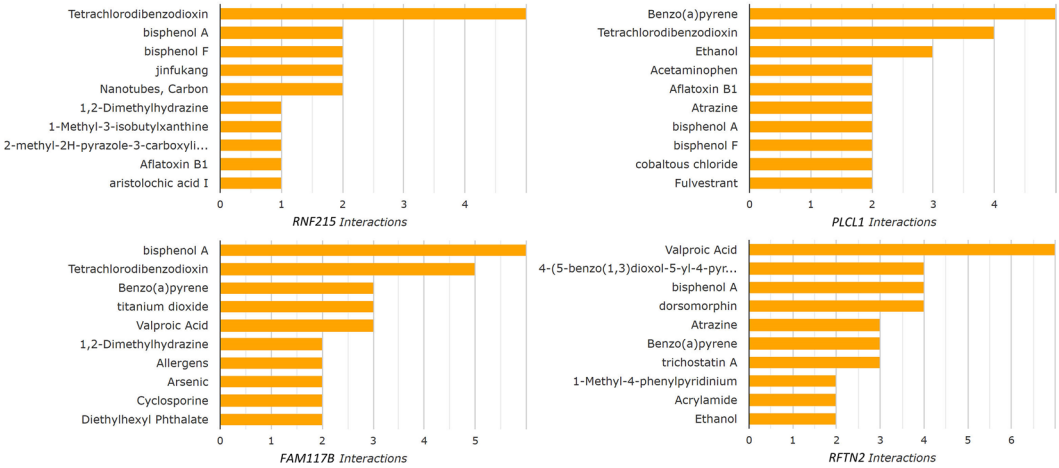

Figure 6. Bar plots showing the top chemical compounds interacting with the four candidate genes, based on data from the Comparative Toxicogenomics Database (CTD).

PheWAS analysis

In a phenome-wide association study (PheWAS) of four genes (RNF215, PLCL1, FAM117B, and RFTN2), RNF215, PLCL1, and RFTN2 all exhibited statistically significant associations with a variety of clinical phenotypes. Specifically, RNF215 was associated with red blood cell distribution width and platelet count, suggesting its potential role in inflammation and blood health. PLCL1 was significantly associated with allergic rhinitis and respiratory diseases, implying its potential impact on immune regulation. RFTN2 showed strong associations with multiple phenotypes related to immunity, the urogenital system, and cardiovascular health, indicating its potentially critical role in these physiological processes and diseases. In contrast to these three, the analysis of FAM117B did not reach statistical significance ($p < 1e\text{-}08$) and did not show an association with any clinical phenotypes, indicating that this gene may not have had a significant impact on phenotypic variation in this analysis (Figure 7).

Discussion

This study employed a cross-tissue TWAS approach, integrating sarcoidosis GWAS summary data with predictive models from the GTEx v8 dataset across 49 tissues, to identify genetic associations with the disease. By combining intersection evidence from UTMOST, FUSION, MAGMA, MR, and colocalization, we prioritized four candidate susceptibility genes: RNF215, PLCL1, FAM117B, and RFTN2. MR and colocalization analyses provided statistical support for putative causal relationships for several of these genes, with RNF215 showing risk effects, FAM117B and RFTN2 exhibiting protective associations, and PLCL1 demonstrating bidirectional effects depending on tissue context. Additionally, functional annotation by the Comparative Toxicogenomics Database (CTD) revealed that these genes interact with a range of environmental chemicals, including bisphenol A and tetrachlorodibenzodioxin, suggesting possible gene–environment interactions in sarcoidosis pathogenesis. PheWAS findings further highlighted pleiotropic effects, linking these genes to diverse traits such as immune-related disorders, respiratory diseases, hematological parameters, and cardiovascular phenotypes, underscoring their potential relevance beyond sarcoidosis. Sarcoidosis is a clinically heterogeneous multisystem disease, with variable patterns of organ involvement, most commonly affecting the lungs and intrathoracic lymph nodes but also involving extrapulmonary organs such as the skin, eyes and heart. In the present study, sarcoidosis was defined using ICD-based endpoints in large-scale GWAS datasets, which did not provide information on organ-specific manifestations or disease subtypes. As a result, our

Figure 7. Manhattan plots displaying phenome-wide association study (PheWAS) results for four candidate genes in the UK Biobank. Purple and green points alternate by phenotype category to facilitate visual distinction between adjacent trait groups. The horizontal dotted lines indicate significance thresholds at $P = 5\times10^{-8}$ (Significant Threshold) and $P = 5\times10^{-6}$ (Suggestive Threshold).

analyses could not distinguish between pulmonary and extrapulmonary sarcoidosis or assess tissue-specific clinical correlations. Accordingly, the genes prioritized by our integrative TWAS and post-GWAS framework should be interpreted as candidates associated with overall sarcoidosis susceptibility, rather than with specific organ involvement. Although cross-tissue signals may indicate shared regulatory mechanisms, future studies with detailed clinical phenotyping and tissue-resolved transcriptomics are needed to validate direct gene–organ associations. To our knowledge, the potential roles of RNF215, PLCL1, and RFTN2 in sarcoidosis have not been previously documented, representing novel findings, while FAM117B is a gene previously reported by GWAS research, and this integrated analysis has validated its role. By integrating COJO analysis, we mitigated LD-related confounding, while MAGMA provided additional gene-level evidence, reinforcing the biological plausibility of our results. Furthermore, TWAS methods (UTMOST and FUSION) bridged noncoding GWAS loci to function by using eQTL data to connect regulatory

variants with gene expression. The identification of these four genes highlights the utility of this integrative framework in refining genetic discoveries, elucidating the genetic architecture of sarcoidosis, and addressing prior gaps in understanding its molecular mechanisms.

RNF215 encodes a RING-type E3 ubiquitin ligase implicated in the regulation of antiviral and inflammatory signaling (32). Experimental work suggests that RING-type E3 ligases, including RNF215, may influence core innate signaling modules such as TBK1, IKKε, and TRAF3, thereby influencing type I interferon and NF-κB outputs and fine-tuning responses to nucleic acid–sensing pathways (33). Importantly, RNF215 has been reported to negatively regulate type I IFN production, and its potential effects on NF-κB activity have been linked to K63-linked ubiquitination of RelA/p65, a post-translational modification known to modulate NF-κB signaling dynamics (32,34). In the context of granulomatous inflammation, altered interferon and NF-κB signaling tone has been proposed to contribute to sustained activation of macrophages and T cells, which are central to granuloma formation and persistence. (35,36). In our study, integrative statistical analyses implicate genetically proxied RNF215 expression in sarcoidosis susceptibility. While these findings do not establish a direct mechanistic role, they suggest that variation in RNF215-mediated immune regulation may influence disease risk through modulation of NF-κB/IFN–related pathways, warranting further functional investigation.Tissue labeling (testis) in GTEx likely reflects shared cis-eQTL architecture rather than causal involvement of gonadal tissue per se, a known caveat in cross-tissue TWAS/MR where informative eQTLs may arise in tissues that best capture the regulatory signal (37). PLCL1 (also known as PRIP-1) is a phospholipase C-like protein that binds IP3 but lacks catalytic activity; it functions as a scaffold/regulator that shapes phosphoinositide and Ca^{2+} signaling (38,39). These pathways are broadly involved in receptor-proximal signaling and cytoskeletal regulation across multiple cell types, including immune and stromal cells. Recent disease-model data show that PLCL1 can influence inflammatory signaling processes, including regulation of NLRP3 inflammasome activity in fibroblast-like synoviocytes in rheumatoid arthritis, indicating a potential link to inflammatory pathways that are also relevant to granulomatous diseases, albeit in a different pathological context (40). In our analyses, genetically proxied PLCL1 expression showed directionally heterogeneous associations across tissues, with a risk-increasing effect observed in fibroblasts and protective associations in thyroid and whole blood, the latter displaying moderate colocalization evidence. These findings highlight the complexity of interpreting tissue-specific TWAS signals and may reflect context-dependent regulatory effects rather than discrete organ-specific mechanisms. One possible explanation is that PLCL1's scaffold function interacts with distinct molecular partners in different cellular environments, leading to divergent downstream signaling consequences. While speculative, this hypothesis aligns with PLCL1's known role as a non-catalytic scaffold in phosphoinositide/Ca^{2+}pathways (40,41). The PheWAS associations with allergic rhinitis and respiratory-related traits provide additional phenotypic context, suggesting potential relevance to mucosal/airway inflammation, but do not establish a direct mechanistic link to sarcoidosis. Overall, our integrative analyses implicate PLCL1 in sarcoidosis risk, but the tissue-specific effects highlight the need for future studies to define its cell-type–specific roles in inflammatory and granulomatous contexts.

FAM117B is comparatively understudied functionally but has reproducible human genetic support: recent large-scale GWAS meta-analyses of sarcoidosis identified FAM117B among genome-wide significant loci, with directions of effect concordant with our MR results (higher expression associated with lower risk) (13,42). These convergent findings strengthen the evidence for a protective role of FAM117B in sarcoidosis susceptibility. While the molecular role of FAM117B remains unclear, its replicated association suggests that it may be involved in regulatory processes that influence immune homeostasis or cellular stress responses relevant to chronic inflammation. The observation of protective associations across immune-relevant tissues, including spleen and whole blood, is compatible with a systemic immune-regulatory effect rather than a strictly organ-specific mechanism (11,13). Further functional studies, such as targeted perturbation experiments in relevant immune cell types, will be necessary to elucidate the cellular pathways through which FAM117B

influences sarcoidosis risk and to bridge the gap between statistical association and biological mechanism. RFTN2 encodes raftlin 2, a lipid raft–associated protein and paralog of raftlin (RFTN1), which has been more extensively characterized. RFTN1 is established as a major raft protein that organizes membrane microdomains to regulate B-cell receptor (BCR) signaling, including tyrosine phosphorylation and calcium mobilization events (43). RFTN1 also contributes to Toll-like receptor (TLR) trafficking, for example by mediating clathrin-dependent endocytosis of TLR4 in response to lipopolysaccharide (44) and facilitating poly (I:C) delivery to TLR3-positive endosomes to trigger interferon-β production (45). Although direct mechanistic studies of RFTN2 are currently lacking, its high sequence homology to RFTN1 and detectable expression in immune cells (46) suggest that it may participate in related membrane-associated regulatory processes. In granulomatous disorders, appropriate regulation of receptor trafficking and signaling attenuation has been proposed as a mechanism to limit sustained activation of macrophages and dendritic cells (47). Our MR and colocalization results (higher RFTN2 expression associated with lower sarcoidosis risk) are therefore consistent with, although they do not yet prove, the hypothesis that variation in RFTN2 expression could influence immune signaling dynamics through effects on membrane organization or receptor routing.This interpretation should be regarded as hypothesis-generating, and further experimental studies will be required to clarify the functional role of RFTN2 in immune regulation and granulomatous disease. Collectively, our integrative analysis implicates dysregulation of ubiquitination (RNF215), phosphoinositide/Ca^{2+} scaffolding (PLCL1), and membrane raft organization/TLR trafficking (RFTN2) in sarcoidosis susceptibility, alongside a previously reported locus (FAM117B) awaiting functional characterization. Taken together, these findings suggest that the genetic architecture of sarcoidosis may involve perturbations in shared immune signaling processes, including pathways related to macrophage and T-cell activation and inflammatory regulation. The pleiotropic effects observed in the PheWAS further suggest that these genes do not operate in a disease-specific vacuum but rather influence broader immunological pathways that can manifest across a spectrum of human traits. Despite these insights, several limitations of our study must be acknowledged. First, while TWAS and colocalization provide strong evidence for a gene-level association, they are ultimately computational inferences based on cis-eQTLs. The causal gene at a locus, and indeed the specific causal transcript isoform, can sometimes be misidentified. Second, the predictive models from GTEx are derived from bulk tissue, which may mask critical cell-type–specific expression effects, particularly in a disease dominated by granulomatous inflammation. The tissue-context-dependent effects we observed for PLCL1 underscore this complexity. Third, as most genetic data used derive from individuals of European ancestry, the generalizability to other populations remains uncertain. In addition, several MR analyses relied on single cis-eQTL instruments and Wald ratio estimates. While cis-based instruments reduce confounding, single-variant MR is inherently limited in its ability to assess horizontal pleiotropy and does not permit sensitivity analyses such as MR-Egger. Therefore, MR results in this study should be interpreted as supportive genetic evidence rather than definitive causal inference. To build upon these findings, future studies will be required to both replicate and experimentally validate the roles of the identified genes in sarcoidosis-relevant contexts. Replication in independent GWAS datasets and the use of alternative or disease-specific eQTL resources, particularly those derived from immune cells, would help assess the robustness of the integrative TWAS and colocalization framework applied here. Moreover, direct transcriptomic analyses in tissues most frequently involved in sarcoidosis—such as lung parenchyma and intrathoracic lymph nodes—as well as in peripheral blood or serum-derived immune cells, may determine whether genetically predicted expression differences are reflected in active granulomatous lesions. From a translational perspective, integrating genetic signals with clinical phenotypes, circulating biomarkers, and imaging features may ultimately facilitate patient stratification and improve understanding of disease progression patterns. Bridging integrative genetic discovery with tissue-resolved functional validation will therefore be essential to translate these statistical associations into clinically meaningful insights in sarcoidosis.

Conclusion

By integrating cross-tissue TWAS with rigorous post-GWAS analyses, we identified three novel and biologically plausible susceptibility genes for sarcoidosis, and validated a previously reported susceptibility gene. These findings provide new insights into the genetic architecture of sarcoidosis and highlight potential targets for future therapeutic development.

Acknowledgments: We thank all participants and researchers involved in the FinnGen project for providing the sarcoidosis GWAS data, as well as the developers of UTMOST, FUSION, MAGMA, and other tools used in this study.

Data Availability: The sarcoidosis GWAS data were obtained from the FinnGen R12 dataset (https://storage.googleapis.com/finngen-public-data-r12/summary_stats/release/finngen_R12_D3_SARCOIDOSIS.gz). The software used in the study and their corresponding reference files can be obtained via the links provided in the Materials and methods section.

Conflict of Interest: Each author declares that he or she has no commercial associations (e.g. consultancies, stock ownership, equity interest, patent/licensing arrangement etc.) that might pose a conflict of interest in connection with the submitted article.

Ethical Approval: The studies for which we have used publicly available GWAS data have received ethical approval from their respective institutional review boards.

Declaration on the Use of AI: The author declares that no AI-assisted tools were used in the preparation of this manuscript.

References

1. Zhou M, Zhou Y, Yang X, Zhou K, Zhu X. Global, regional, and national burden of interstitial lung diseases and pulmonary sarcoidosis from 2000 to 2021: a systematic analysis of incidence, mortality, and disability-adjusted life years. Front Public Health. 2025;13:1578480. doi:10.3389/fpubh.2025.1578480
2. Cao J, Li H, Yin X, Yang J, Pu L, Yang J. Global burden of pulmonary sarcoidosis from 1990 to 2021: a comprehensive analysis based on the GBD 2021 study. Front Med (Lausanne). 2025;12:1585005. doi:10.3389/fmed.2025.1585005
3. Li X, Zhang Z. Global disease burden of interstitial lung disease and pulmonary sarcoidosis, 1990–2021, and forecast analysis over the next decade. Sarcoidosis Vasc Diffuse Lung Dis. 2025;42(3):16462. doi:10.36141/svdld.v42i3.16462
4. Neves FS, Pereira IA, Sztajnbok F, Neto NSR. Sarcoidosis: a general overview. Adv Rheumatol. 2024;64(1):57. doi:10.1186/s42358-024-00381-z
5. Arkema EV, Cozier YC. Sarcoidosis epidemiology: recent estimates of incidence, prevalence and risk factors. Curr Opin Pulm Med. 2020;26(5):527–34. doi:10.1097/MCP.0000000000000715
6. Franzen DP, Brutsche M, Nilsson J, et al. Sarcoidosis: a multisystem disease. Swiss Med Wkly. 2022;152:w30049. doi:10.4414/smw.2022.w30049
7. Coker RK, Cullen KM. Sarcoidosis: key disease aspects and update on management. Clin Med (Lond). 2025;25(3):100326. doi:10.1016/j.clinme.2025.100326
8. Mannes K, Thomas PS. Sarcoidosis: rarely a single system disorder. Breathe (Sheff). 2020;16(4):200207. doi:10.1183/20734735.0207-2020
9. Rossides M, Kullberg S, Eklund A, et al. Risk of first and recurrent serious infection in sarcoidosis: a Swedish register-based cohort study. Eur Respir J. 2020;56(3):2000767. doi:10.1183/13993003.00767-2020
10. Fischer A, Schmid B, Ellinghaus D, Nothnagel M, Gaede KI, Schürmann M, et al. A novel sarcoidosis risk locus for Europeans on chromosome 11q13.1. Am J Respir Crit Care Med. 2012;186(9):877–85. doi:10.1164/rccm.201204-0708OC
11. Calender A, Weichhart T, Valeyre D, Pacheco Y. Current insights in genetics of sarcoidosis: functional and clinical impacts. J Clin Med. 2020;9(8):2633. doi:10.3390/jcm9082633
12. Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, Gaede KI, et al. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. Nat Genet. 2008;40(9):1103–6. doi:10.1038/ng.198
13. Yuan S, Chen J, Geng J, Zhao SS, Yarmolinsky J, Arkema EV, et al. GWAS identifies genetic loci, lifestyle factors and circulating biomarkers that are risk factors for sarcoidosis. Nat Commun. 2025;16:2481. doi:10.1038/s41467-025-57829-z
14. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467–84. doi:10.1038/s41576-019-0127-1
15. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190–5. doi:10.1126/science.1222794
16. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245–52. doi:10.1038/ng.3506
17. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. Opportunities and challenges for transcriptome-wide association studies. Nat Genet. 2019;51(4):592–9. doi:10.1038/s41588-019-0385-z

18. Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, et al. A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet. 2019;51(3):568–76. doi:10.1038/s41588-019-0345-7
19. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023; 613(7944):508–18. doi:10.1038/s41586-022-05473-8
20. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580–5. doi:10.1038 /ng.2653
21. Sun R, Hui S, Bader GD, Lin X, Kraft P. Powerful gene set analysis in GWAS with the generalized Berk-Jones statistic. PLoS Genet. 2019;15(3):e1007530. doi:10.1371/journal. pgen.1007530
22. Sun X. Unveiling novel susceptibility genes for kidney stone disease by a cross-tissue transcriptome-wide association study. World J Urol. 2026;44(1):108. doi:10.1007 /s00345-025-06070-w
23. Gui J, Yang X, Tan C, Wang L, Meng L, Han Z, et al. A cross-tissue transcriptome-wide association study reveals novel susceptibility genes for migraine. J Headache Pain. 2024;25(1):94. doi:10.1186/s10194-024-01802-6
24. Liao C, Laporte AD, Spiegelman D, Akçimen F, Joober R, Dion PA, et al. Transcriptome-wide association study of attention deficit hyperactivity disorder identifies associated genes and phenotypes. Nat Commun. 2019;10(1):4450. doi:10.1038/s41467-019-12450-9
25. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219. doi:10.1371 /journal.pcbi.1004219
26. de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. Nat Rev Genet. 2016;17(6):353–64. doi:10.1038/nrg.2016.29
27. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408. doi:10.7554/eLife.34408
28. Sun X. Physical activity associated with age at menopause: a Mendelian randomization study. Medicine (Baltimore).2025;104(6):e41514.doi:10.1097/MD.000000000 0041514
29. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383. doi:10.1371 /journal.pgen.1004383
30. Davis AP, Wiegers TC, Sciaky D, Barkalow F, Strong M, Wyatt B, et al. Comparative Toxicogenomics Database's 20th anniversary: update 2025. Nucleic Acids Res. 2025;53(D1):D1328–34. doi:10.1093/nar/gkae883
31. Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. Nature. 2021;597(7877):527–32. doi:10.1038/s41586-021-03855-y
32. Wu Y, Chen D, Hu Y, Zhang S, Dong X, Liang H, et al. Ring finger protein 215 negatively regulates type I IFN production via blocking NF-κB p65 activation. J Immunol. 2022;209(10):2012–21. doi:10.4049/jimmunol.2200346
33. Zhang Y, Li LF, Munir M, Qiu HJ. RING-domain E3 ligase-mediated host-virus interactions: orchestrating immune responses by the host and antagonizing immune defense by viruses. Front Immunol. 2018;9:1083. doi:10.3389 /fimmu.2018.01083
34. Song K, Li S. The role of ubiquitination in NF-κB signaling during virus infection. Viruses. 2021;13(2):145. doi:10.3390/v13020145
35. Fallahi-Sichani M, Kirschner DE, Linderman JJ. NF-κB signaling dynamics play a key role in infection control in tuberculosis. Front Physiol. 2012;3:170. doi:10.3389/fphys .2012.00170
36. Wilson JL, Mayr HK, Weichhart T. Metabolic programming of macrophages: implications in the pathogenesis of granulomatous disease. Front Immunol. 2019;10:2265. doi:10.3389/fimmu.2019.02265
37. Li B, Ritchie MD. From GWAS to gene: transcriptome-wide association studies and other methods to functionally understand GWAS discoveries. Front Genet. 2021;12: 713230. doi:10.3389/fgene.2021.713230
38. Muter J, Brighton PJ, Lucas ES, Lacey L, Shmygol A, Quenby S, et al. Progesterone-dependent induction of phospholipase C-related catalytically inactive protein 1 (PRIP-1) in decidualizing human endometrial stromal cells. Endocrinology. 2016;157(7):2883–93. doi:10.1210 /en.2015-1914
39. Pan Z, Huang J, Song H, Xiao Y, Liu T, Zeng Y, et al. Tumor cell "slimming" regulates tumor progression through PLCL1/UCP1-mediated lipid browning. Adv Sci (Weinh). 2019;6(10):1801862. doi:10.1002/advs.201801862
40. Luo S, Li XF, Yang YL, Song B, Wu S, Niu XN, et al. PLCL1 regulates fibroblast-like synoviocytes inflammation via NLRP3 inflammasomes in rheumatoid arthritis. Adv Rheumatol. 2022;62(1):25. doi:10.1186/s42358-022-00252-5
41. Liu YZ, Wilson SG, Wang L, Liu XG, Guo YF, Li J, et al. Identification of PLCL1 gene for hip bone size variation in females in a genome-wide association study. PLoS One. 2008;3(9):e3160. doi:10.1371/journal.pone.0003160
42. Fischer A, Ellinghaus D, Nutsua M, Hofmann S, Montgomery CG, Iannuzzi MC, et al. Identification of immune-relevant factors conferring sarcoidosis genetic risk. Am J Respir Crit Care Med. 2015;192(6):727–36. doi:10.1164 /rccm.201503-0418OC
43. Saeki K, Miura Y, Aki D, Kurosaki T, Yoshimura A. The B cell-specific major raft protein, Raftlin, is necessary for the integrity of lipid raft and BCR signal transduction. EMBO J. 2003;22(12):3015–26. doi:10.1093/emboj/cdg293
44. Tatematsu M, Yoshida R, Morioka Y, Ishii N, Funami K, Watanabe A, et al. Raftlin controls lipopolysaccharide-induced TLR4 internalization and TICAM-1 signaling in a cell type-specific manner. J Immunol. 2016;196(9): 3865–76. doi:10.4049/jimmunol.1501734

45. Watanabe A, Tatematsu M, Saeki K, Shibata S, Shime H, Yoshimura A, et al. Raftlin is involved in the nucleocapture complex to induce poly (I:C)-mediated TLR3 activation. J Biol Chem. 2011;286(12):10702–11. doi:10.1074/jbc.M110.185793
46. Lin YT, Tsai MH, Su YY, Chen WC, Huang SC, Chien CY. Expression of major lipid raft protein Raftlin in chronic rhinosinusitis with nasal polyps in smoking and non-smoking patients correlated with interleukin-17 and tumor necrosis factor-α levels. Biomolecules. 2022;12(9):1316. doi:10.3390/biom12091316
47. Ivashkiv LB. Inflammatory signaling in macrophages: transitions from acute to tolerant and alternative activation states. Eur J Immunol. 2011;41(9):2477–81. doi:10.1002/eji.201141783

Annex

Table S1. The original results of TWAS analyses in cross-tissue (UTMOST)

Table S2. The results of significant genes in TWAS analyses in cross-tissue (UTMOST)

Table S3. The results of significant genes in TWAS analyses in single-tissue (FUSION)

Table S4. The results of jointly significant genes in conditional and joint analysis

Table S5. The results of significant genes in MAGMA

Table S6. Fusion results for 10 identified significant genes in UTMOST, FUSION and MAGMA

Table S7. The results of Mendelian randomization analysis and Bayesian colocalization analysis

Table S8. The RNF215-Chemical-Disease interaction identified by Comparative Toxicogenomics Database (CTD)

Table S9. The PLCL1-Chemical-Disease interaction identified by Comparative Toxicogenomics Database (CTD)

Table S10. The FAM117B-Chemical-Disease interaction identified by Comparative Toxicogenomics Database (CTD)

Table S11. The RFTN2-Chemical-Disease interaction identified by Comparative Toxicogenomics Database (CTD)