Original Article

DOI: 10.5582/irdr.2024.01064

Risk associated circulating biomarkers S100A3 identified in congenital heart disease-associated pulmonary arterial hypertension

Weian Zhao[1,§], Yijing Chen[2,§], Zhongsu Yu[3], Zhouping Wang[2], Renke He[2], Simian Cai[2], Zhongzhong Chen[2,*], Liangping Cheng[3,*]

[1] Department of Anesthesiology Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing, China;
[2] Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China;
[3] Department of Gastroenterology Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing, China.

SUMMARY: With improved survival rates among congenital heart disease (CHD) patients, pulmonary arterial hypertension (PAH) linked to CHD becomes more prevalent in both children and adults. PAH remains a significant contributor to morbidity and mortality in this population. Although genome-wide association studies (GWAS) have identified potential genetic variants with PAH risk and prognosis, the identification of circulating biomarkers with causal roles in CHD-PAH remains unclear. We employed the summary data-based Mendelian randomization (SMR) method, integrating expression profile data from the Gene Expression Omnibus (GEO) database related to CHD-PAH. This approach aimed to pinpoint genes causally associated with risk of CHD-PAH. We used a two-sample Mendelian randomization (MR) approach to efficiently screen for circulating proteins affecting CHD-PAH, leveraging publicly available genetic data from the UK biobank Pharma Proteomics Project (UKB-PPP) (54,219 UKB participants). Genetic determinants (cis-SNPs) of circulating proteins were used as instruments, and MR analyses assessed the influence of these proteins on CHD-PAH susceptibility in the largest PAH GWAS (2085 cases and 9659 controls). We conducted colocalization analyses to ensure shared genetic signals between circulating proteins and PAH and performed immune cell infiltration analysis to understand immune regulatory mechanisms in CHD-PAH. We found that a 1 SD increase in circulating S100 calcium binding protein A3 (S100A3) levels correlated with a reduced PAH risk (OR: 0.073, 95% CI: 0.020-0.267; $p = 0.00799$). Sensitivity analyses including various cis-SNPs, provided consistent estimates for *S100A3* (inverse variance weighted (IVW) OR: 0.085, 95% CI: 0.032-0.225; $p = 7.5 \times 10^{-7}$ and MR-Egger OR: 0.212, 95% CI: 0.013-3.376; $p = 0.387$). Colocalization analyses confirmed a shared genetic signal for *S100A3* and PAH, with a posterior probability of 99.9%. Transcriptomic investigations further highlighted *S100A3*'s protective role in CHD-PAH. Our study using SMR and GEO data identified *S100A3* as a gene associated with a reduced risk of PAH in CHD patients. Elevated circulating levels of S100A3 were linked to a reduced PAH risk, and transcriptomic evidence further supported its protective function in CHD-PAH.

Keywords: congenital heart disease (CHD), pulmonary arterial hypertension (PAH), Mendelian randomization (MR), S100 calcium binding protein A3 (*S100A3*)

1. Introduction

Congenital heart disease (CHD) is the most common congenital malformation, occurring in approximately 8 per 1,000 live births (*1*). Complex CHDs encompass a set of more severe conditions that are potentially influenced by multifactorial inheritance (*2,3*). Pulmonary arterial hypertension (PAH) associated with congenital heart disease (CHD-PAH) is a severe and complex condition that affects a significant number of patients worldwide and worsens their prognosis and quality of life. PAH is characterized by elevated pulmonary arterial pressure and vascular resistance, which can lead to right ventricular failure and ultimately, death. The survival rate for untreated PAH is dismal, with a median survival of 2.8 years from the time of diagnosis (*4*). Despite advancements in medical and surgical treatments, the management of CHD-PAH remains challenging due to its heterogeneous nature and the complexity of its underlying pathophysiological mechanisms (*1,5,6*).

Current therapeutic strategies for CHD-PAH include pharmacological treatments, such as prostacyclin analogs,

endothelin receptor antagonists, and phosphodiesterase-5 inhibitors, which improve symptoms and slow disease progression (*6,7*) but are not curative and often cause significant side effects. Surgical interventions, like defect repair or lung transplantation, carry high risks and are not suitable for everyone (*8-10*). This underscores the urgent need for novel therapeutics.

Recent studies have highlighted the involvement of circulating proteins in the pathogenesis and progression of PAH, offering promise as biomarkers for early diagnosis, prognosis, and therapeutic targets (*11*). For instance, increased levels of soluble intercellular adhesion molecule-1 (ICAM-1) have been associated with PAH severity in pediatric CHD cases (*12*). MicroRNA studies, like the identification of microRNA-27b dysregulation, also suggest roles in CHD-PAH mechanism (*13*). Despite these discoveries, the exact role of many circulating proteins in CHD-PAH remains unclear, emphasizing the need for further investigation.

Traditional statistical methods used in observational studies to investigate the association between circulating proteins and CHD-PAH risk are often limited by confounding factors and reverse causation. Mendelian randomization (MR) is a powerful epidemiological method that utilizes genetic variants as instrumental variables to infer causal relationships between exposures (*e.g.*, circulating proteins) and outcomes (*e.g.*, CHD-PAH) while minimizing confounding and reversing causation (*14*). Common variants, which can be identified through genome-wide association studies (GWAS), typically have small genetic effects (*15*) but are often used as instrumental variables. By leveraging data from genome-wide association studies (GWAS) and protein quantitative trait loci (pQTL), MR can provide robust evidence for the causal role of specific proteins in disease development.

CHD is recognized as the most prevalent birth defect (*16*) and is often associated with various complications (*2*). This study employs a two-sample MR approach to identify CHD-PAH-related proteins, validated through Bayesian colocalization and supported by lung tissue gene expression analysis. We also explore the immune regulatory mechanisms associated with CHD-PAH through analysis of immune cell infiltration. Our primary objective is to uncover novel circulating proteins that contribute to CHD-PAH pathogenesis and could serve as potential therapeutic targets.

2. Materials and Methods

2.1. Study design and data sources

We applied a two-sample MR design to identify circulating proteins associated with risk of CHD-PAH (Figure 1A). To achieve this, we used summary data from the largest genome-wide association study (GWAS) on PAH conducted among individuals of European descent (*17*), as well as protein quantitative trait loci (pQTL) GWASs from the Pharma Proteomics Project by Sun *et al.* (*18*) in the UK Biobank. Detailed methods of protein assays are described in this study (*18*).

First, the GSE113439 dataset was analyzed to identify 1,082 differential mRNA genes associated with congenital heart disease-related pulmonary arterial hypertension (CHD-PAH) in patients. After ID conversion, 1,052 genes remained, with 893 upregulated genes and 159 downregulated genes (Supplemental Table S1, *https://www.irdrjournal.com/action/getSupplementalData.php?ID=237*). Through literature search, GWAS data related to pulmonary arterial hypertension, GCST007228, was identified, which summarized data from four studies: UK National Institute of Health Research BioResource (NIHRBR) for Rare Diseases study, US National Biological Sample and Data Repository for Pulmonary Arterial Hypertension/ PAH Biobank (PAHB) study, Paris Pulmonary Hypertension Allele-Associated Risk cohort (PHAAR) study, and British Heart Foundation Pulmonary Arterial Hypertension GWAS (BHFPAH) study, as outcome data. Exposure data were obtained from deCODE plus UKB with duplicates removed, pQTL_data, $p < 5e^{-8}$, clump = 500 kb and $R^2 = 0.1$ (Supplemental Table S2, *https://www.irdrjournal.com/action/getSupplementalData.php?ID=238*); both standard two-sample method pQTL analysis and SMR method pQTL analysis were conducted. Additionally, SMR's eQTL analysis was performed for all genes. The results of the standard two-sample method pQTL analysis were corrected for p-value FDR, yielding 102 positive genes; the SMR method's eQTL and pQTL results selected genes with p_SMR < 0.05, amounting to 1,213 and 142 genes, respectively. Combining this with transcriptome data from congenital heart disease with pulmonary arterial hypertension found in the GEO database, a single regulatory gene was identified through meta-analysis (Figure 1B): *S100A3* (S100 calcium binding protein A3), and colocalization was performed using UKB protein data. The intersection of genes obtained from SMR_ eQTL [rs185078626; (Link to NCBI SNP) (*https://www.ncbi.nlm.nih.gov/snp/?term=rs185078626*)] and SMR_pQTL [rs1005436; (Link to NCBI SNP) (*https://www.ncbi.nlm.nih.gov/snp/?term=rs1005436*)] both map on *S100A3*, but they are not the same SNP; further colocalization analysis using UKB protein data revealed that only rs1005436 showed colocalization (Figure 1C), with a posterior probability of 0.99. This study reveals that the downregulation of S100A3 protein expression is closely associated with risk of PAH, consistent with the differential gene results from GEO (the expression of *S100A3* is shown in Figure 1D, $\log_2(FC) = -1.97$, adj.P.Val = 0.001). Enrichment analysis of the downregulated genes revealed that the Calcium-dependent protein binding pathway within the GO molecular function (MF) was regulated.

2.2. Ethical approval

No separate ethical approval was required due to the use of publicly available data.

2.3. Instrument selection and validation

To select genetic variants for studying the effect of CHD-PAH, we considered SNPs within the *S100A3* gene (GRCh37/hg19 chromosome 1 position 153519805–153521734) that associated with expression of the gene (*i.e.*, protein quantitative trait loci (pQTL)) in blood at genome-wide significance ($p < 5\times10^{-8}$). To ensure that the variants used as instruments in MR are not highly correlated with each other, we then ranked them in order of the P values of their associations with UKB_PPP_pQTL (*https://registry.opendata.aws/ukbppp/*) (*19*) and pruned to linkage disequilibrium correlation $R^2 = 0.1$ and distance threshold 500 kilobases (Table 1).

2.4. MR

The MR approach was based on three key assumptions (*19*). First, the genetic variants used were associated with the risk factors. With the advent of large-scale modern GWASs, genetic variants associating with exposure can be identified in large datasets (*20*). Second, the genetic variants must not be associated with

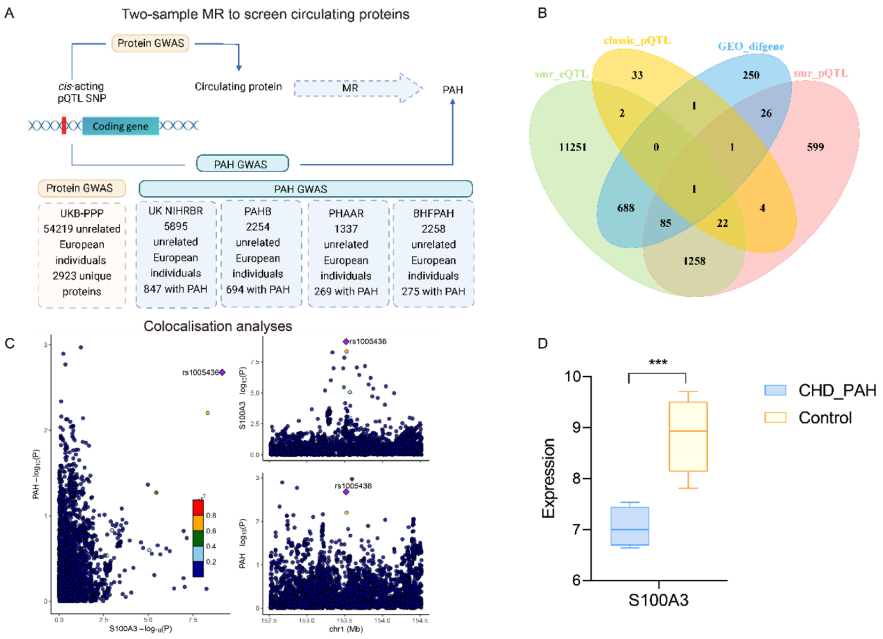

Figure 1. Overall study design. (A) Employing the classical two-sample MR and SMR methods, we obtained all gene eQTLs and pQTLs from the database and, in combination with GEO data, identified circulating protein targets for CHD-PAH; **(B)** one gene *S100A3* was identified in four data analyses; **(C)** colocalisation analyses for the target rs1005436; **(D)** *S100A3* expression significantly lower in CHD-PAH patients. *Abbreviations*: MR, Mendelian randomisation; GWAS, genome-wide association study; pQTL, protein quantitative trait loci; SNP, single nucleotide polymorphism; *S100A3*, S100 calcium binding protein A3; NIHRBR, UK National Institute of Health Research BioResource for Rare Diseases study; PAHB, US National Biological Sample and Data Repository for Pulmonary Arterial Hypertension/PAH Biobank study; PHAAR, Paris Pulmonary Hypertension Allele-Associated Risk cohort study; BHFPAH, British Heart Foundation Pulmonary Arterial Hypertension GWAS study; CHD-PAH, Congenital Heart Disease-associated pulmonary arterial hypertension; GEO dataset, GSE113439 in Gene Expression Omnibus. eQTLs, expression quantitative trait loci; GWAS, genome-wide association study; HEIDI, heterogeneity in dependent instruments; IVW-MR, inverse-variance-weighted Mendelian randomization; MR-PRESSO, Mendelian Randomization Pleiotropy RESidual Sum and Outlier; MAF, minor allele frequency; SMR, summary-data-based Mendelian randomization.

Table 1. Single nucleotide polymorphism employed as instruments for *S100A3*

rsID	Chromosome	Position (hg19)	Effect allele	Other allele	Effect allele frequency	UKB-PPP pQTL (n = 33777)					pQTL
						beta	SE	p value	R^{2a}	F statistic[b]	p value
rs1005436	1	153521932	A	G	0.128	0.07	0.011	7.18E-10	0.005	37.96	0.002
rs185078626	1	153340640	T	C	0.006	0.297	0.05	5.42E-09	0.126	34.03	0.715
rs28472359	1	153528079	C	T	0.098	0.076	0.013	4.62E-09	0.041	34.34	0.006
rs73024420	1	153497360	A	G	0.017	0.167	0.029	1.45E-08	0.015	32.12	0.067

[a]Estimates the proportion of variance in the phenotype explained by the genetic variant. [b]Measure of instrument strength. *Abbreviations*: *S100A3*, S100 calcium binding protein A3, pQTL, protein quantitative trait loci; SE, standard error. UKB-PPP, UK Biobank Pharma Proteomics Project.

confounders of the exposure–outcome relationship. A potential violation of this assumption can occur due to confounding by LD and/or population ancestry (*21*). Lastly, genetic variants must not affect the outcome, except through the exposure of interest (referred to as a lack of horizontal pleiotropy) (*22*).

Large-scale GWASs for circulating proteins (*23*) have often found that the genetic determinants of circulating proteins reside cis (in close proximity) to the encoding genes. Use of cis-acting single nucleotide polymorphisms (SNPs) for MR reduces potential horizontal pleiotropy and increases the validity of MR assumptions, because a cis-SNP strongly associated with the protein is likely to directly influence the gene's transcription and consequently the circulating protein level.

2.5. Colocalization

A threat to the validity of cis-MR analyses is confounding by linkage disequilibrium. This occurs when a variant associated with the phenotype is in linkage disequilibrium with a variant associated with the outcome, thereby producing a spurious MR association. To test the robustness of our results against such confounding, we used Bayesian colocalization analysis *via* the Coloc method (*23*) between the protein marker S100A3 and all outcomes with significant MR associations. Coloc presents evidence for five hypotheses: no causal variant for either trait, a causal variant for trait 1 but not trait 2, a causal variant for trait 2 but not trait 1, distinct causal variants underlying each trait, and a shared causal variant underlying both traits. A high posterior probability for the fifth hypothesis (> 0.8) supports the presence of a shared causal variant underlying both traits, while a high posterior probability for the fourth hypothesis (> 0.8) supports the presence of distinct causal variants underlying each trait, thus indicating confounding by linkage disequilibrium in the corresponding MR association (also referred to as horizontal pleiotropy). In the presence of a statistically significant MR association that is not a false positive finding, if the posterior probabilities for both the fourth and fifth hypotheses are < 0.8, this would suggest that the colocalization analysis is likely underpowered to discriminate whether the MR association is attributable to a shared causal variant or a confounding variant in linkage disequilibrium (*i.e.*, horizontal pleiotropy).

2.6. Sensitivity analysis

For proteins supported by MR and colocalization analyses, we conducted sensitivity analyses. In IVW (inverse variance weighted) and MR-Egger analyses, we included multiple cis-SNPs that are in weak linkage disequilibrium ($R^2 < 0.6$) with the leading cis-SNPs for candidate proteins. These analyses considered correlated variants using the MR R package (*24,25*), as the consistency of estimates could strengthen the hypothesized effects. MR-Egger allows for a y-intercept term in a random effects model. An intercept that is different from zero indicates directional horizontal pleiotropy, suggesting a violation of the third MR assumption.

2.7. Transcriptomic data in lung tissue

We first investigated *S100A3* using microarray-based transcriptomic data in CHD-PAH lung: GSE113439 (*26*). Fresh frozen lung samples were obtained from the recipient organs of 4 patients with PAH secondary to congenital heart disease (CHD) and 11 normal controls (normal lung tissue obtained from tissue flanking lung cancer resections). RNA was extracted and hybridized on Affymetrix microarrays. During the analysis, two outlier samples were identified in the normal control group and subsequently excluded to ensure the robustness and reliability of the comparative analysis.

2.8. Differential gene expression analysis and probe reannotation

We obtained preprocessed data from Gene Expression Omnibus (GEO) using the R package "GEOquery". After acquiring the expression matrix, we annotated the probesets based on the annotation profile recorded in the "AnnoProbe" package associated with the "tinyarray" package (version 2.3.2), to filter out duplicate and unannotated probes. We performed quantile normalization on the log-transformed intensities using the "normalizeBetweenArrays" function in the "limma" R package. Subsequently, we used the "limma" package to identify differentially expressed mRNAs. *P*-values were adjusted using the Benjamini-Hochberg method. Unless otherwise specified, "differentially expressed" (DE) mRNAs are defined as having an FDR < 0.05 and a $|\log_2(FC)| > 1$; $\log_2(FC) = 1$ means fold change = 2 (upregulated). $\log_2(FC) = -1$ means fold change = 0.5 (downregulated). $|\log_2(FC)|$ means absolute $\log_2$ fold change exceeds 1, meaning FC > 2 or FC < 0.5.

2.9. Functional and pathway enrichment analysis

We conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis on the filtered differentially expressed genes (DEGs) using the R package "clusterProfiler" (V4.10.0) (*27*) and visualized the results. KEGG pathways and GO terms were considered statistically significant with a cut-off value of $p < 0.05$. The GO enrichment analysis consists of three components: molecular functions (MFs), biological processes (BPs), and cellular components (CCs).

2.10. Immune infiltration analysis

To assess the proportion of infiltrating immune cells in CHD-PAH gene expression profiles, we utilized the CIBERSORT (*28*) bioinformatics algorithm. This method quantifies immune infiltration by referencing the LM22 dataset, which comprises 22 immune cell subtypes and 1,000 permutations. We analyzed and visualized correlations for these immune cells with R's "corrplot" package. Differences between CHD-PAH and control samples were depicted using violin plots created with the "vioplot" package in R. Additionally, we explored the relationships between *S100A3* expression and the levels of infiltrating immune cells using Spearman's rank correlation analysis in R, with the findings graphically presented through the "ggplot2" package.

2.11. Software and preregistration

The MR analyses in this paper were conducted using the "TwoSampleMR"(version 0.6.8), "MRPRESSO"(version1.0), "locuscomparer"(version 1.0.0) R packages, as well as the SMR & HEIDI methods and software tool (*29,30*). Some data from genome-wide association studies were extracted from the OpenGWAS platform (*17*) This study was not preregistered.

3. Results

3.1. Instrument selection and validation

First, we did classic two-sample MR. After clumping, we identified 4 variants to serve as the genetic instrument for *S100A3* (Figure 2, A-C), in the MR analysis of 2,923 proteins, we found that an increase of one standard deviation (SD) in circulating S100 calcium-binding protein A3 (S100A3) was associated with a reduced risk of PAH [odds ratio (OR): 0.08, 95% confidence interval (CI): 0.020-0.267; $p = 7.51\times10^{-7}$], used FDR for *p*-value adjustment) (Figure 2A). Then we used the SMR & HEIDI methods and software tool also found lead variant predicted that a one SD increase in serum S100A3 protein levels would significantly reduce the occurrence of PAH, and the average F statistic across all variants was

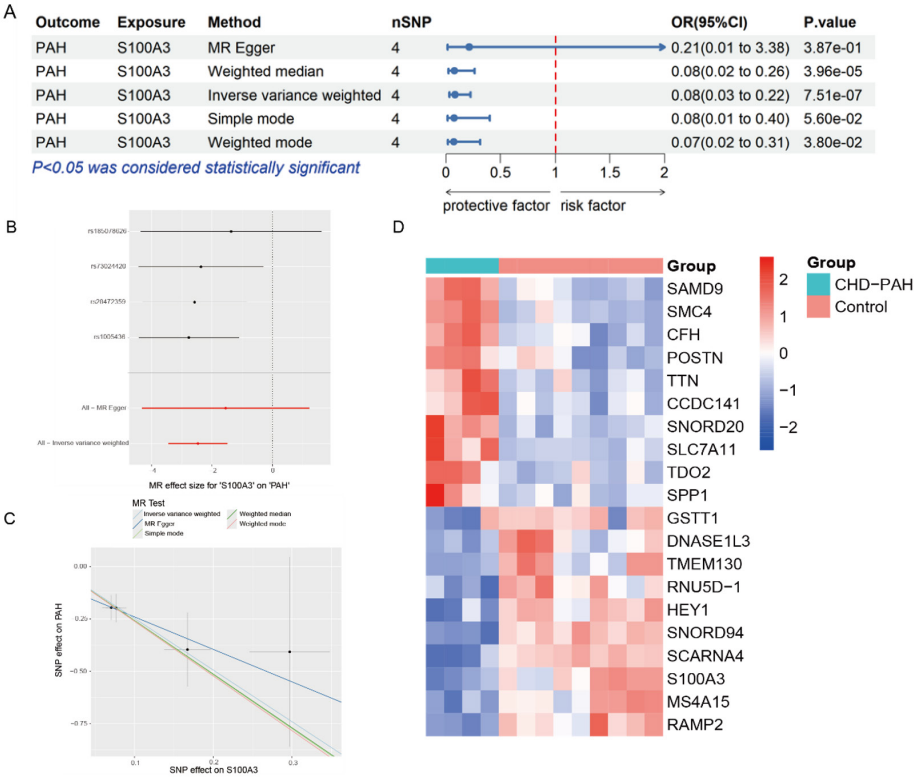

Figure 2. The association between genetically proxied serum S100A3 protein levels and pulmonary arterial hypertension related to congenital heart disease. (A) showed odds ratios are calculated per standard deviation increase in the exposure. Using the classical two-sample MR method, we obtained all potential regulatory SNPs upstream and downstream of the OID31352 marker probe encoding the S100A3 protein on human chromosome 1 and conducted MR analysis with PAH GWAS to identify positive instrumental variables, thereby confirming its potential as a therapeutic target for CHD-PAH treatment-related causal inference. **(B)** Four positive instrument SNP showed that rs1005436 was a good target. **(C)** Five methods were used for the examination, and the inverse variance weighted result was significant, suggesting that after accounting for confounding, the down regulation of S100A3 protein is causally associated with the occurrence of PAH. **(D)** displays the expression profiles of the top ten differentially expressed mRNAs, it shows that *S100A3* was the top third down regulated mRNA in congenital heart disease associated pulmonary arterial hypertension patients. *Abbreviations*: MR, Mendelian randomization; *S100A3*, S100 calcium binding protein A3; CHD-PAH, congenital heart disease associated pulmonary arterial hypertension; HEIDI, heterogeneity in dependent instruments; IVW-MR, inverse-variance-weighted Mendelian randomization; MR-PRESSO, Mendelian Randomization Pleiotropy RESidual Sum and Outlier; MAF, minor allele frequency; SMR, summary-data-based Mendelian randomization.

34 (Table 1), indicating a low risk of weak instrument bias. The positive control analysis identified a MR association in the expected direction between genetically proxied *S100A3* and pulmonary arterial hypertension ($p < 0.001$) (Figure 2, B and C). Finally, we obtained consistent results in the transcriptome data expression profiles (Figure 1D). The heat map displays the expression profiles of the top ten differentially expressed mRNAs in the lung tissue of children with pulmonary arterial hypertension associated with congenital heart disease (Figure 2D).

In two separate pQTL GWASs (deCODE+UKB, pQTL_data with duplicate proteins removed, filtered for $p < 5 \times 10^{-8}$, clumped at 500 kb with $R^2 = 0.1$) and one independent eQTL GWAS (eQTLGen_Consortium blood cis_eqtl) (Supplemental Table S3, *https://www.irdrjournal.com/action/getSupplementalData.php?ID=239*), we conducted an MR scan to identify proteins associated with CHD-PAH. After Benjamini-Hochberg correction, one candidate protein remained: circulating S100A3. A genetically determined increase of one SD in plasma S100A3 was associated with an average reduction of 63% in the risk of developing CHD-PAH (OR (95% CI) from smr_eQTL was 0.683 (0.495 to 0.872), $p = 0.371$; OR (95% CI) from smr_pQTL was 0.063 (0.040 to 0.085), $p = 0.004$) (Table 2).

3.2. Colocalization analysis

We performed a colocalization analysis between the GWAS of the candidate protein (S100A3) from Sun *et al.* (18), and the PAH GWAS to assess potential confounding due to linkage disequilibrium (LD). The SNP site rs1005436, which was significant in both the classical two-sample pQTL and SMR_pQTL (Supplemental Table S4, *https://www.irdrjournal.com/action/getSupplementalData.php?ID=240* and S5, *https://www.irdrjournal.com/action/getSupplementalData.php?ID=241*), showed excellent colocalization with PAH, with a posterior probability of 99.9% in the colocalization analyses (Figure 1C), indicating the presence of a shared signal.

3.3. Lung tissue transcriptomic data

Using microarray-based transcriptomic data from whole lung samples (GSE113439), we confirmed that *S100A3* was significantly downregulated in the transcriptome sequencing results of fresh frozen lung tissues from patients with pulmonary arterial hypertension associated with congenital heart disease compared to the control group ($\log_2(FC) = -1.97$, $p = 6.98 \times 10^{-5}$, adj.P.Val = 0.001). Further analysis of all 159 downregulated genes through GO and KEGG annotation revealed that the Calcium-dependent protein binding pathway within the molecular function (MF) category of GO was regulated (Figure 3A and 3B). Using the SMR method to obtain the risk factor corelation of CHD-PAH and PAH, the downregulation of S100A3 protein expression among the top-ranked genes reflects a causal association between these regulatory sites and the occurrence of PAH (Figure 3C), which is consistent with the data results from the GEO expression profiles (Figure1D, Figure 2D, and Figure 3C).

3.4. Immune infiltration analysis

The results of immune cell infiltration analysis showed that, compared to the control group, the expression of memory B cells, $CD8^+$ T cells, and monocyte cells was significantly increased in CHD-PAH patients (Figure 4A). Further analysis revealed that in CHD-PAH patients, the expression level of the *S100A3* gene was positively correlated with $CD8^+$ T cells, memory B cells, and NK cells (Figure 4B and 4C). These findings suggest that *S100A3* may play a complex role in immune regulation by modulating the functions of these key immune cells, thereby influencing the progression of CHD-PAH.

4. Discussion

This study addresses the essential challenge of identifying circulating proteins linked to the risk of CHD-PAH. Using a two-sample MR approach that combines GWAS and pQTL analyses, the research aimed to identify proteins associated with CHD-PAH. Bayesian colocalization analysis was employed to confirm the robustness of the findings, and differential gene expression analysis in lung tissue was conducted to further substantiate the results. The study's significant contribution is the identification of S100A3 as a key protein associated with CHD-PAH, supported by differential expression and functional pathway enrichment analyses, which elucidates potential biological mechanisms underlying the disease.

Several studies support the findings of our research, which identified S100A3 as a significant protein

Table 2. The instrument SNPs that affect pulmonary arterial hypertension at both the transcriptional and protein levels, identified through the SMR method, were mapped to the gene *S100A3*

	Gene	Method	topSNP	nsnp_HEIDI	*p*.value	OR	OR (95% CI)
mRNA	*S100A3*	smr_eQTL	rs185078626	7	0.371	0.683	0.683(0.495 to 0.872)
Protein		smr_pQTL	rs1005436	7	0.004	0.063	0.063(0.040 to 0.085)

Abbreviations: SNPs, SNP, single nucleotide polymorphism; SMR, Summary-data-based Mendelian Randomization; *S100A3*, S100 calcium binding protein A3.

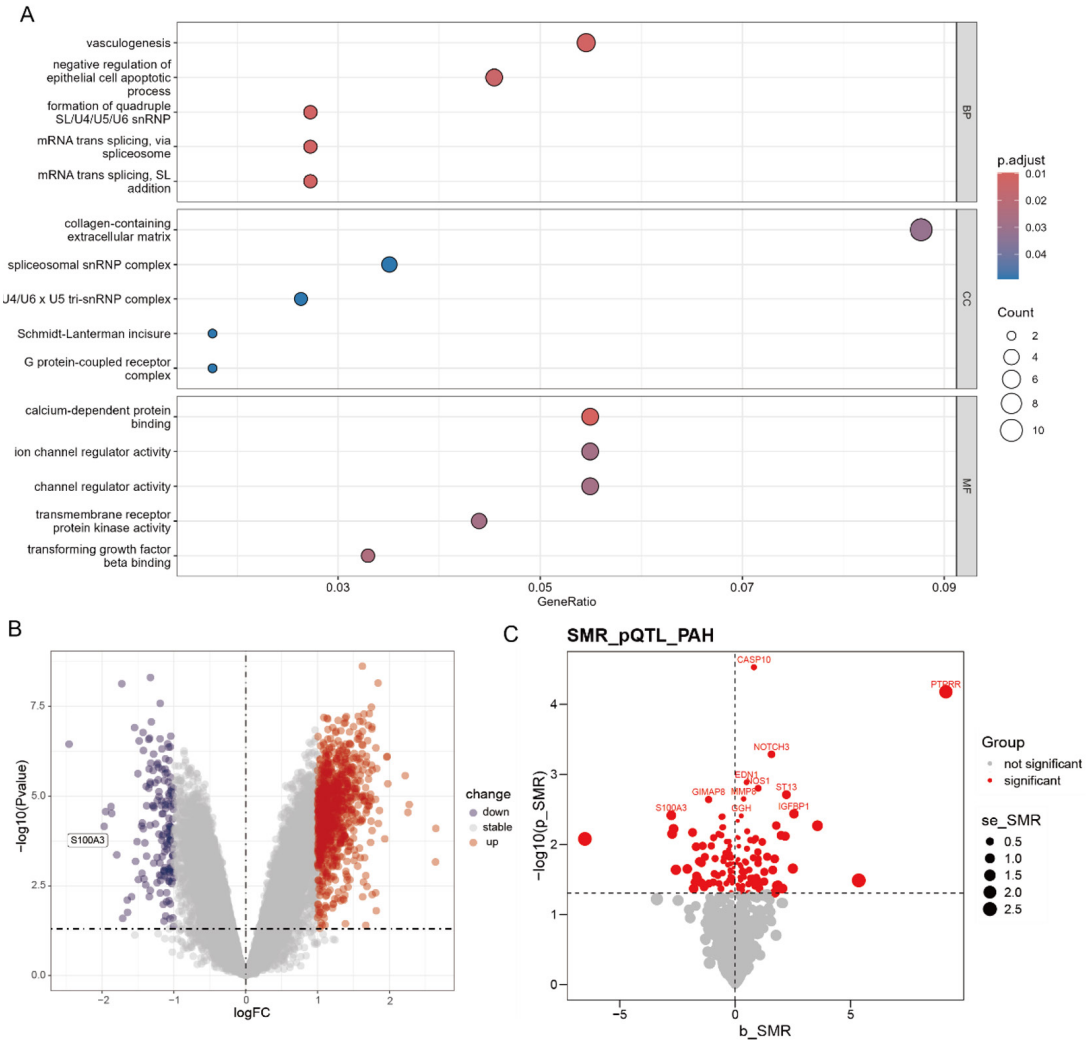

Figure 3. Analysis of the *S100A3* gene expression in CHD-PAH patients and involved pathways. (A) Enrichment analysis was performed on the downregulated genes, among which the Calcium-dependent protein binding pathway within the molecular function (MF) category of GO was found to be regulated. **(B)** and **(C)** The pQTL obtained from SMR analysis and present the volcano plot for positive exposure, showing that the downregulation of S100A3 protein expression is closely associated with the risk of PAH, consistent with the differential gene results from GEO datasect (logFC = -1.97, adj.P.Val = 0.001). *Abbreviations*: CHD-PAH , congenital heart disease associated pulmonary arterial hypertension ; GEO, Gene Expression Omnibus; logFC, log 2 fold change.

associated with CHD-PAH. For instance, the role of *S100A3* in other pathological conditions has been explored extensively. Liu *et al.* observed that *S100A3* activation is involved in tumorigenesis in colorectal cancer, suggesting its critical role in cellular processes and disease progression (*31*). This aligns with our findings where *S100A3* levels were significantly altered in CHD-PAH patients, indicating its broader relevance in various diseases. Furthermore, the study by Tao *et al.* on hepatocellular carcinoma (HCC) demonstrated that *S100A3* is implicated in tumor aggressiveness and that modulating its expression could be a potential therapeutic strategy (*32*). This supports our hypothesis that *S100A3* could be a therapeutic target in CHD-PAH as well. Additionally, a study by Gianni *et al.* highlighted the interaction of *S100A3* with RARα and PML-RARα in breast and lung cancer cells, affecting the stability and activity of these receptors (*33*). This interaction underscores the importance of *S100A3* in regulating cellular functions, which could be extrapolated to its role in CHD-PAH pathophysiology. Contrarily, some studies have reported different roles for *S100A3*. For instance, the work of Fritz *et al.* on the structural properties of *S100A3* revealed its unique calcium and zinc-binding properties, which are crucial for its function in hair cuticle formation (*34*). While this study focuses on a different biological context, it highlights the multifunctional nature of *S100A3*, which could explain its involvement in diverse pathologies, including CHD-PAH. Moreover, the research by Minato *et al.* on the evolution of *S100A3* and *PAD3* genes in mammals provided insights into the adaptive significance of these genes in hair formation (*35*). Although this study does not directly relate to CHD-PAH, it underscores the evolutionary importance of *S100A3*, suggesting its critical role in mammalian physiology, which may

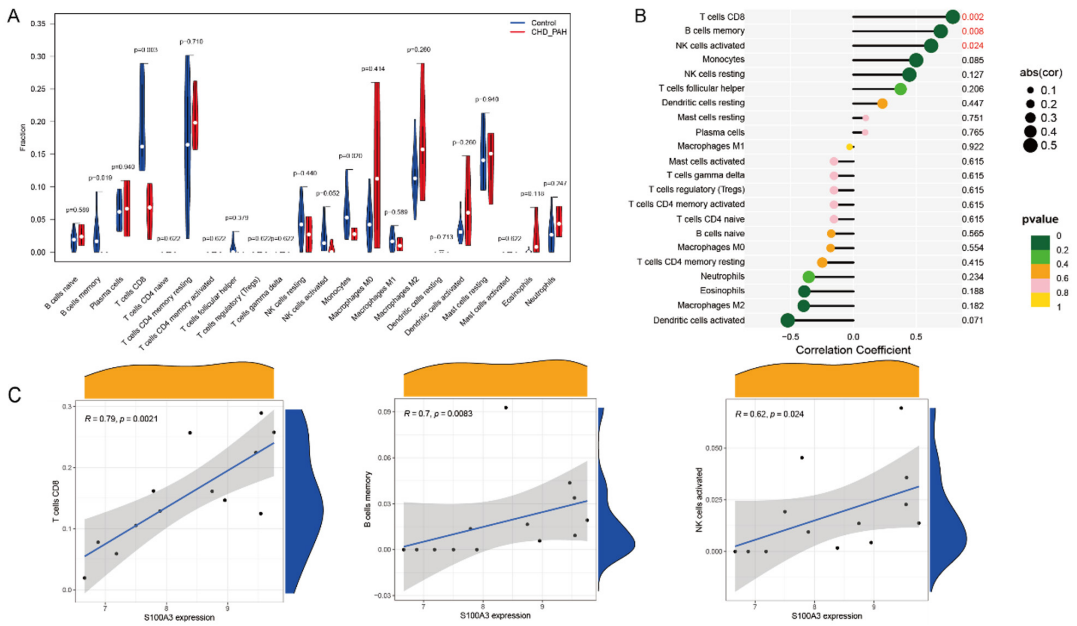

Figure 4. Analysis of the correlation between *S100A3* gene expression and immune cell profiles in CHD-PAH patients. (A) Shows the differential expression of various immune cell types between CHD-PAH patients and normal controls, with significant differences indicated by p-values. **(B)** A lollipop plot illustrating the strength and statistical significance of correlations between *S100A3* expression and different immune cell types. **(C)** Scatter plots and trend lines reveal the correlation between *S100A3* expression and specific immune cell types (*e.g.*, CD8+ T cells and NK cells), with each plot indicating the correlation coefficient (R) and *p* value.

extend to cardiovascular diseases. In conclusion, the existing literature supports our findings regarding the role of *S100A3* in CHD-PAH. The consistent observation of *S100A3*'s involvement in various diseases and its regulatory functions in cellular processes reinforce its potential as a biomarker and therapeutic target. However, further studies are necessary to elucidate the precise mechanisms by which *S100A3* influences CHD-PAH and to explore its potential in clinical applications. Despite the limited specific literature on *S100A3* in the context of CHD-PAH, the discussion can still be framed around the identified results and general mechanisms of S100 proteins and their role in cardiovascular diseases.

Our study identified S100A3 as a significant protein associated with CHD-PAH, suggesting a protective role against PAH. S100 proteins, including S100A3, are known to be involved in various cellular processes such as cell cycle progression and differentiation, which are critical in maintaining vascular homeostasis (*31,33,35,36*). The downregulation of *S100A3* in patients with CHD-PAH, as observed in our transcriptomic analysis, indicates a potential disruption in these cellular processes, contributing to the pathogenesis of PAH. The significant association between S100A3 levels and PAH risk, supported by both MR and colocalization analyses, underscores the potential mechanistic role of S100A3 in modulating pulmonary vascular remodeling. This remodeling is a hallmark of PAH and involves the proliferation and migration of pulmonary arterial smooth muscle cells (PASMCs) and pulmonary vascular endothelial cells (PVECs). The protective effect of higher circulating S100A3 levels could be mediated through its influence on PASMCs, possibly by inhibiting their proliferation and promoting apoptosis, thereby preventing vascular remodeling and subsequent PAH development. Furthermore, the genetic association of the rs1005436 SNP with PAH, as demonstrated in our pQTL and SMR_pQTL analyses, highlights the importance of genetic factors in regulating *S100A3* expression and its downstream effects on pulmonary vasculature. The high posterior probability of colocalization at this SNP locus suggests a shared genetic basis for S100A3 expression and PAH risk, providing a potential target for therapeutic intervention.

Our findings are further supported by the GO and KEGG pathway analyses, which revealed that calcium-dependent protein binding pathways are significantly regulated in CHD-PAH patients. S100A3 is a small calcium-binding protein (molecular weight ~10-12 kDa) belonging to the S100 protein family, characterized by tissue- or cell type-specific expression patterns. It contains two EF-hand calcium-binding sites, and calcium binding induces conformational changes that facilitate interactions with ligands or specific receptors. *S100A3* typically forms homodimers, heterodimers, and higher-order oligomers, and is involved in various cellular processes, including cell cycle regulation, proliferation, differentiation, migration, metabolism, cytoskeletal dynamics, signal transduction, and cell death (*36*). The dysregulation of these pathways in CHD-PAH patients could contribute to the observed downregulation of *S100A3* and its protective effects against PAH. Our study

provides compelling evidence for the involvement of *S100A3* in the pathogenesis of CHD-PAH. The protective association of higher *S100A3* levels with reduced PAH risk, supported by genetic and transcriptomic analyses, suggests that S100A3 could be a potential biomarker and therapeutic target for PAH. Further research is needed to elucidate the precise molecular mechanisms by which *S100A3* exerts its protective effects and to explore its potential in clinical applications.

In our study, using a two-sample MR approach combined with GWAS and pQTL analysis, we identified circulating S100A3 protein as significantly associated with a reduced risk of CHD-PAH. This finding is novel and robust, supported by Bayesian colocalization analysis and differential gene expression in lung tissue, which consistently showed downregulation of *S100A3* in CHD-PAH patients. Previous studies have highlighted the complexity and high risk associated with pulmonary arterial hypertension in congenital heart disease (PAH-CHD), emphasizing the need for precise biomarkers and therapeutic targets (*7,33*). The association of *S100A3* with CHD-PAH provides new insights into the pathophysiological mechanisms underlying this condition and opens potential avenues for targeted therapy. Unlike earlier studies that primarily focused on clinical and hemodynamic parameters (*37,38*), our approach integrates genetic and proteomic data, offering a more comprehensive understanding of the disease. Additionally, our results remained significant after FDR adjustment, and the SMR & HEIDI methods further validated the findings, demonstrating the reliability of our results. The identification of rs1005436 SNP as a key genetic locus associated with PAH risk through colocalization analysis underscores its potential as a genetic marker for early diagnosis and risk stratification in CHD-PAH patients.

In reflecting upon the limitations of this study, several aspects warrant consideration. First, the study primarily relies on computational methods and lacks integration with wet lab experiments, which could provide additional validation and insights into the biological mechanisms underlying the associations identified. Second, the sample size used in the analysis may be relatively small, potentially limiting the generalizability of the findings. Furthermore, the absence of clinical validation analyses means that the practical applicability of the identified proteins in clinical settings remains uncertain. Additionally, the use of multiple datasets introduces the possibility of batch effects, which could influence the results and interpretations.

In summary, this study successfully identifies circulating proteins associated with the risk of CHD-PAH, with a particular emphasis on the S100A3 protein. The integration of MR, colocalization analysis, and differential gene expression analysis provides robust evidence supporting these findings. Looking ahead, these results pave the way for further research to explore the biological mechanisms and potential therapeutic targets for CHD-PAH. Future studies should aim to incorporate larger sample sizes, clinical validation, and experimental approaches to enhance the understanding and applicability of these findings in clinical practice.

Acknowledgements

We thank Professor Wentao Ni from Peking University People's Hospital for providing the Mendelian randomization tutorial, code, and writing guidance. We thank Dr. Jianming Zeng (University of Macau), and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes.

Funding: This work was supported by the National Natural Science Foundation of China (NSFC82200263), and Natural Science Foundation of Guangdong Province, China (407222037058).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Diller GP, Gatzoulis MA. Pulmonary vascular disease in adults with congenital heart disease. Circulation. 2007; 115:1039-1050.
2. Chen Z, Gao Y, Lu L, Li N, Liu P, Peng R, Liu L, Huang H, Fu Q, Hong H, Zhang J, Wang H. Rare loss-of-function variants reveal threshold and multifactorial inheritance of dextrocardia. Sci Bull (Beijing). 2023; 68:1993-1998.
3. Chen ZZ, Gao YQ, Xie H, Huang YC, Chen F, Lei YP. Transcription factors dysregulated in three complex birth defects datasets. Reproductive and Developmental Medicine. 2022; 6:79-85.
4. Fraisse A, Jais X, Schleich JM, di Filippo S, Maragnès P, Beghetti M, Gressin V, Voisin M, Dauphin C, Clerson P, Godart F, Bonnet D. Characteristics and prospective 2-year follow-up of children with pulmonary arterial hypertension in France. Arch Cardiovasc Dis. 2010; 103:66-74.
5. D'Alto M, Mahadevan *VS.* Pulmonary arterial hypertension associated with congenital heart disease. Eur Respir Rev. 2012; 21:328-337.
6. Zhao M, Liu J, Xin M, Yang K, Huang H, Zhang W, Zhang J, He S. Pulmonary arterial hypertension associated with congenital heart disease: An omics study. Front Cardiovasc Med. 2023; 10:1037357.
7. Togănel R, Benedek I, Suteu C, Blesneac C. Management of pulmonary arterial hypertension associated with congenital heart disease. Rom J Intern Med. 2007; 45:229-234.
8. Bando K, Turrentine MW, Sharp TG, Sekine Y, Aufiero TX, Sun K, Sekine E, Brown JW. Pulmonary hypertension after operations for congenital heart disease: Analysis of risk factors and management. J Thorac Cardiovasc Surg. 1996; 112:1600-1607; discussion 1607-1609.
9. Villamor E, Gonzalez-Luis GE. Pulmonary hypertension and lung damage associated with prolonged exposure to patent ductus arteriosus in very preterm infants. Am J Respir Crit Care Med. 2024; 210:254-255.

10. Suesaowalak M, Cleary JP, Chang AC. Advances in diagnosis and treatment of pulmonary arterial hypertension in neonates and children with congenital heart disease. World J Pediatr. 2010; 6:13-31.
11. Righini FM, Apostolo A, Heck PB, Farina S, Hager A, Correale M, Badagliacca R, Barbieri S, Sciomer S, Agostoni P. Exercise physiology in pulmonary hypertension patients with and without congenital heart disease. Eur J Prev Cardiol. 2019; 26:86-93.
12. Oguz MM, Oguz AD, Sanli C, Cevik A. Serum levels of soluble ICAM-1 in children with pulmonary artery hypertension. Tex Heart Inst J. 2014; 41:159-164.
13. Ma K, Zhao Q, Chen W, Zhang H, Li S, Pan X, Chen Q. Human lung microRNA profiling in pulmonary arterial hypertension secondary to congenital heart defect. Pediatr Pulmonol. 2015; 50:1214-1223.
14. Das BB, Frank B, Ivy D. Segmental pulmonary hypertension in children with congenital heart disease. Medicina (Kaunas). 2020; 56:492.
15. Chen Z, Finnell RH, Lei Y, Wang H. Progress and clinical prospect of genomic structural variants investigation. Sci Bull (Beijing). 2024; 69:705-708.
16. Chen Z, Wang Y, Lan F, Li S, Wang J. An expanded view of infertility: The challenge of the changing profiling of major birth defects in China. Biosci Trends. 2023; 17:318-321.
17. Rhodes CJ, Batai K, Bleda M, *et al.* Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. Lancet Respir Med. 2019; 7:227-238.
18. Sun BB, Chiou J, Traylor M, *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023; 622:329-338.
19. Davey Smith G, Davies NM, Dimou N, *et al.* STROBE-MR: Guidelines for strengthening the reporting of Mendelian randomization studies. *https://peerj.com/preprints/27857v1* (accessed November 14, 2024).
20. Swanson SA, Hernán MA. The challenging interpretation of instrumental variable estimates under monotonicity. Int J Epidemiol. 2018; 47:1289-1297.
21. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019; 20:467-484.
22. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. BMJ. 2018; 362:k601.
23. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet. 2021; 17:e1009440.
24. Yavorska OO, Burgess S. Mendelian Randomization: An R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017; 46:1734-1739.
25. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. Stat Med. 2016; 35:1880-1906.
26. Mura M, Cecchini MJ, Joseph M, Granton JT. Osteopontin lung gene expression is a marker of disease severity in pulmonary arterial hypertension. Respirol Carlton Vic. 2019; 24:1104-1110.
27. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021; 2:100141.
28. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015; 12:453-457.
29. Woolf B, Rajasundaram S, Cronjé HT, Yarmolinsky J, Burgess S, Gill D. A drug target for erectile dysfunction to help improve fertility, sexual activity, and wellbeing: Mendelian randomisation study. BMJ. 2023; 383:e076197.
30. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, Yang J. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016; 48:481-487.
31. Tao R, Wang ZF, Qiu W, He YF, Yan WQ, Sun WY, Li HJ. Role of S100A3 in human hepatocellular carcinoma and the anticancer effect of sodium cantharidinate. Exp Ther Med. 2017; 13:2812-2818.
32. Liu B, Sun WY, Zhi CY, Lu TC, Gao HM, Zhou JH, Yan WQ, Gao HC. Role of S100A3 in human colorectal cancer and the anticancer effect of cantharidinate. Exp Ther Med. 2013; 6:1499-1503.
33. Gianni M, Terao M, Kurosaki M, Paroni G, Brunelli L, Pastorelli R, Zanetti A, Lupi M, Acquavita A, Bolis M, Fratelli M, Rochette-Egly C, Garattini E. S100A3 a partner protein regulating the stability/activity of RARα and PML-RARα in cellular models of breast/lung cancer and acute myeloid leukemia. Oncogene. 2019; 38:2482-2500.
34. Fritz G, Mittl PRE, Vasak M, Grutter MG, Heizmann CW. The crystal structure of metal-free human EF-hand protein S100A3 at 1.7-a resolution. J Biol Chem. 2002; 277:33092-33098.
35. Minato T, Unno M, Kitano T. Evolution of S100A3 and PAD3, two important genes for mammalian hair. Gene. 2019; 713:143975.
36. Holzinger D, Tenbrock K, Roth J. Alarmins of the S100-family in juvenile autoimmune and auto-inflammatory diseases. Front Immunol. 2019; 10:182.
37. van der Feen DE, Bartelds B, De Boer RA, Berger RMF. Pulmonary arterial hypertension in congenital heart disease: translational opportunities to study the reversibility of pulmonary vascular disease. Eur Heart J. 2017; 38:2034-2041.
38. Li Q, Dimopoulos K, Liu T, Xu Z, Liu Q, Li Y, Zhang J, Gu H. Peripartum outcomes in a large population of women with pulmonary arterial hypertension associated with congenital heart disease. Eur J Prev Cardiol. 2019; 26:1067-1076.

Received November 22, 2024; Revised January 27, 2025; Accepted February 13, 2025.

[§]These authors contributed equally to this work.
**Address correspondence to:*
Zhongzhong Chen, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China.
E-mail: zhongzhongchen@gmail.com

Liangping Cheng, Children's Hospital of Chongqing Medical University, No.136, Zhongshan 2nd Road, Chongqing, China.
E-mail: chelsea_lpc@163.com

Released online in J-STAGE as advance publication February 18, 2025.