

RESEARCH

Open Access



Integrative eQTL and Mendelian randomization analysis reveals key genetic markers in mesothelioma

Jinsong Li^{1,2}, Xingmeng Wang^{1,2}, Yaru Lin^{1,2}, Zhengliang Li^{4*} and Wei Xiong^{1,2,3*}

Abstract

Background Mesothelioma is a rare cancer that originates from the pleura and peritoneum, with its incidence increasing due to asbestos exposure. Patients are frequently diagnosed at advanced stages, resulting in poor survival rates. Therefore, the identification of molecular markers for early detection and diagnosis is essential.

Methods Three mesothelioma datasets were downloaded from the GEO database for differential gene expression analysis. Instrumental variables (IVs) were identified based on expression quantitative trait locus (eQTL) data for Mendelian randomization (MR) analysis using mesothelioma Genome-Wide Association Study (GWAS) data from the FINNGEN database. The intersecting genes from MR-identified risk genes and differentially expressed genes were identified as key co-expressed genes for mesothelioma. Functional enrichment analyses, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA), as well as immune cell correlation analysis, were performed to elucidate the roles of key genes in mesothelioma. Additionally, the differential expression of key genes in mesothelioma was validated in independent GEO datasets and TCGA datasets. This integrative research combining multiple databases and analytical methods established a robust model for identifying mesothelioma risk genes.

Results The research conducted in our study identified 1608 genes that were expressed differentially in mesothelioma GEO datasets. By combining these genes with 192 genes from MR analysis, we identified 14 key genes. Notably, MPZL1, SOAT1, TACC3, and CYBRD1 are linked to a high risk of mesothelioma, while TGFBR3, NDRG2, EPAS1, CPA3, MNDA, PRKCD, MTUS1, ALOX15, LRRN3, and ITGAM are associated with a lower risk. These genes were found to be enriched in pathways associated with superoxide metabolism, cell cycle regulation, and proteasome function, all of which are linked to the development of mesothelioma. Noteworthy observations included a significant infiltration of M1 macrophages and CD4+T cells in mesothelioma, with genes SOAT1, MNDA, and ITGAM showing a positive correlation with the level of M1 macrophage infiltration. Furthermore, the differential expression analyses conducted on the GEO validation set and TCGA data confirmed the significance of the identified key genes.

*Correspondence:

Zhengliang Li
407874365@qq.com
Wei Xiong
skyeljah201@gmail.com

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



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Conclusion This integrative eQTL and Mendelian randomization analysis provides evidence of a positive causal association between 14 key co-expressed genes and mesothelioma genetically. These disease critical genes are implicated in correlations with biological processes and infiltrated immune cells related to mesothelioma. Moreover, our study lays a theoretical foundation for further research into the mechanisms of mesothelioma and potential clinical applications.

Keywords eQTL localization, Mesothelioma, Mendelian randomization, Biomarkers, Tumor immunity

Introduction

Mesothelioma is an uncommon and treatment-resistant form of cancer that develops from serosal mesothelial cells. Its onset is frequently linked to exposure to asbestos. With the implementation of bans on asbestos, it is anticipated that the rates of both incidence and mortality related to mesothelioma will reach their highest levels in the middle of the 21st century [1]. Additionally, mesothelioma is characterized by the frequent loss or mutation of tumor suppressor genes such as cyclin dependent kinase inhibitor 2 A (CDKN2A), BRCA1 associated protein 1 (BAP1), neurofibromin 2 (NF2), and cullin 1 (CUL1) [2]. Due to the difficulty of early diagnosis and effective treatment with traditional methods, mesothelioma is often detected at an advanced stage, with a median survival time of less than one year and a poor prognosis [3]. Hence, the identification of a set of reliable biomarkers for mesothelioma holds considerable importance in facilitating early detection and the implementation of clinical immunotherapy [4].

Expression quantitative trait loci (eQTL) are genomic regions statistically associated with variations in gene expression levels. By identifying how specific genetic variants influence gene activity, eQTLs link genetic variation to complex traits and diseases [5]. Mendelian randomization (MR) is a method that uses genetic variation to determine the causal relationship between observed risk factors and disease outcomes. By leveraging genetic variation as a “natural experiment,” MR offers a more efficient way to test causality, effectively minimizing the influence of confounding factors [6, 7].

In this research, MR was employed to investigate the correlation between human whole-genome eQTL data and mesothelioma Genome-Wide Association Study (GWAS) data [8]. Additionally, mesothelioma transcriptome data sourced from the Gene Expression Omnibus (GEO) database was integrated to pinpoint significant genes. Analyses conducted using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA) methods provided insights into the pathophysiological implications of key genes in the development and progression of mesothelioma. Using the cell immune infiltration analysis algorithm “CIBERSORT,” we constructed an immune infiltration model for mesothelioma and analyzed the correlation between the expression levels of key genes

and immune infiltrating cells in mesothelioma. Ultimately, we validated the differential expression of key genes using external data from the Cancer Genome Atlas (TCGA) and GEO database, enhancing the reliability of our findings. This research seeks to investigate potential molecular indicators of mesothelioma and assess their biological implications on the disease. The findings are intended to guide forthcoming studies on the mechanisms of mesothelioma and the development of immunotherapeutic approaches (Fig. 1).

Materials and methods

Data collection and processing

Four human mesothelioma transcriptome datasets were obtained from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), including GSE2549 (45 mesothelioma samples and 9 normal samples) [9], GSE42977 (39 mesothelioma samples and 9 normal samples) [10], GSE29370 (21 mesothelioma samples and 1 normal sample), and GSE51024 (55 mesothelioma samples and 41 normal samples) [11]. The GSE2549, GSE42977, and GSE29370 datasets were integrated into a training set using the “limma” and “sva” packages in R software (version 4.3.1). Batch correction was conducted through principal component analysis (PCA). The training dataset comprised 105 samples of mesothelioma and 19 samples of normal mesothelial tissue. The GSE51024 dataset was utilized as a validation set for further confirmation (Table 1). Furthermore, transcriptome expression data from 86 mesothelioma samples (TCGA-MESO.htseq_fpk.m.tsv) and 110 normal lung tissue samples (TCGA-LUAD.htseq_fpk.m.tsv and TCGA-LUSC.htseq_fpk.m.tsv) were obtained from TCGA database (<https://portal.gdc.cancer.gov/>) for external validation.

Differential gene expression analyses (DEGs)

In the R programming environment, the “limma” and “dplyr” packages were utilized to detect differentially expressed genes (DEGs) in mesothelioma samples compared to normal controls, with statistical significance determined at a threshold of $P < 0.05$ and $|\text{Fold Change}| < 0.585$. Following this analysis, visual representations of the DEGs were generated using the “ggplot2” package and the “pheatmap” package to draw volcano plot and heatmap.

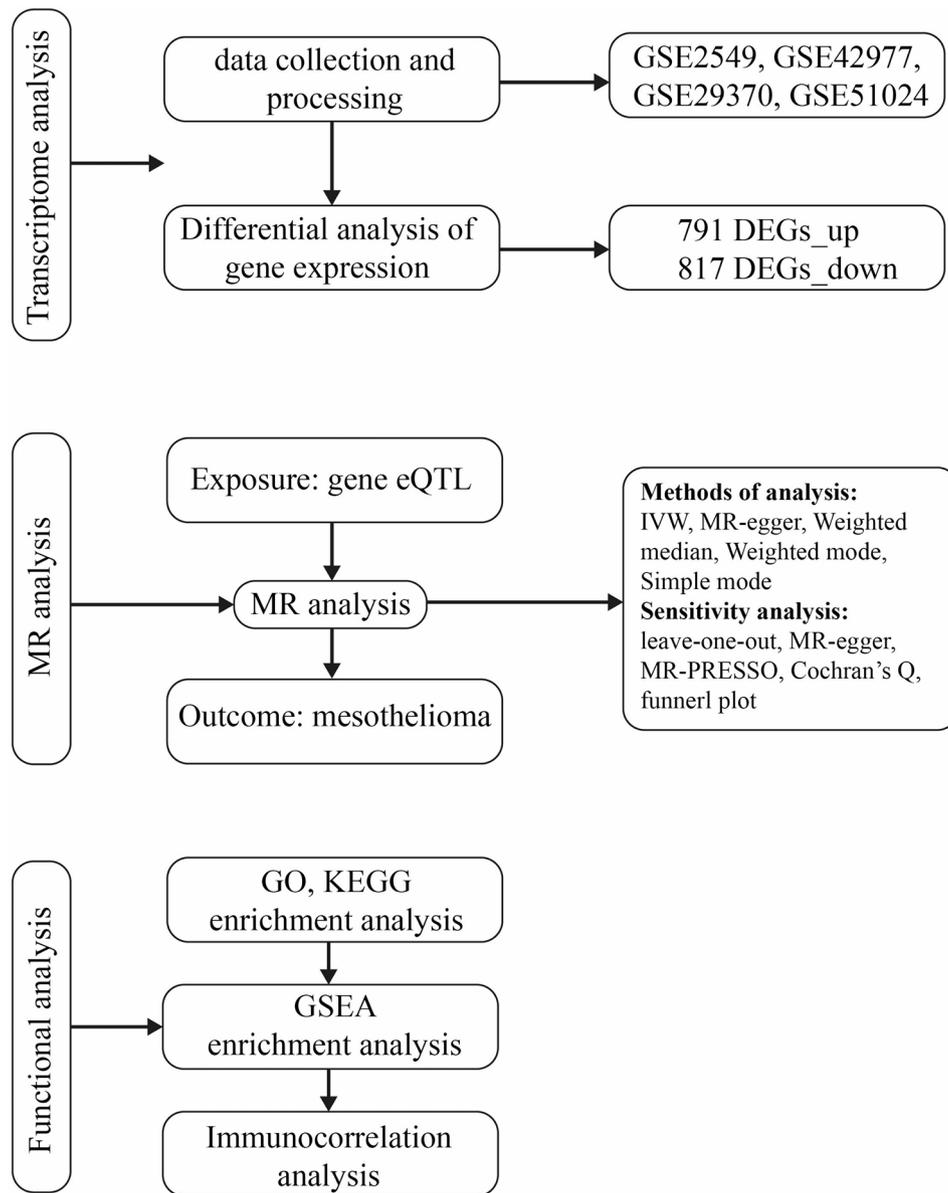


Fig. 1 Workflow diagram of this study

Table 1 Microarray data details in this study

GEO accession	Platform	Mesothelioma	Normal	Data type
GSE2549	GPL96-57554	45	9	Training
GSE42977	GPL570-55999	39	9	Training
GSE29370	GPL6790-11603	21	1	Training
GSE51024	GPL570-55999	55	41	Testing

Mendelian randomization

Mendelian randomization data sources

In the R programming software, we employed Mendelian randomization utilizing GWAS data to investigate the causal relationship between genes in the human genome and mesothelioma. We downloaded the eQTL dataset from the OpenGWAS database (<https://gwas.mrcieu.ac.uk/>)

to serve as the exposure GWAS data, encompassing 19,942 genes. We acquired the finn-b-C3_MESO_THELIOMA dataset from the FINNGEN database as the outcome GWAS data for mesothelioma, comprising 218,792 samples and 16,380,466 single nucleotide polymorphisms (SNPs). Both the exposure and outcome GWAS datasets for mesothelioma are sourced from European populations, ensuring uniformity and coherence of the data. The MR analysis conducted in this research utilized publicly GWAS summary data, thus no additional ethical approval or informed consent was required.

Selection of instrumental variables (IVs)

The selection of instrumental variables (IVs) was conducted by applying specific criteria. This involved utilizing the “TwoSampleMR” package to assess the eQTL dataset of each gene and identifying SNPs that exhibited a strong association with gene expression levels ($P < 5e-08$) [12]. Set the linkage disequilibrium condition to $r^2 < 0.001$, and a physical distance threshold of 10,000 base pairs. To determine the presence of weak variable bias, it is recommended to compute the F-statistic of exposure, with instrumental variables having an F-value below 10 indicating the existence of such bias, which should be excluded [13].

Mendelian randomization analysis

Following the identification of IVs with effective causal detection capability, a two-sample MR analysis was conducted on the exposure and outcome using the “TwoSampleMR” and “VariantAnnotation” packages [14, 15]. Five analytical methods were employed for the MR analysis: Inverse-variance weighted (IVW), MR-Egger, weighted median, weighted mode, and simple mode, with IVW being the primary method [16]. An assumption in MR analysis is that the selected IVs can only affect the outcome through exposure. Therefore, to make the results between exposure and outcome more stable and reliable, testing for horizontal pleiotropy is necessary. In this research investigation, the MR-Egger intercept test was employed to evaluate horizontal pleiotropy, with a significance threshold set at $P < 0.05$ [17]. Deviation of the MR-Egger intercept from zero indicates the existence of horizontal pleiotropy, which raises concerns about the reliability of the results. Such deviation suggests that the SNPs under consideration are linked solely to the exposure variable and are not influenced by other potential confounding factors. Additionally, due to variations in the selected population and experimental conditions, there may be heterogeneity among samples, potentially biasing the results. Therefore, Cochran’s Q test was used to detect heterogeneity ($P > 0.05$), and a leave-one-out analysis was performed to assess the consistency of the pleiotropy test results.

Screening and validation of key genes for mesothelioma

Based on the analysis results in MR, genes exhibiting heterogeneity and pleiotropy were filtered out. The genes associated with a significance level of $P < 0.05$ were divided into high-risk genes ($OR > 1.0$) and low-risk genes ($OR < 1.0$) based on the odds ratio (OR) value. The “VennDiagram” package was utilized to intersect the key associated genes with the upregulated and downregulated genes of DEGs, and a Venn diagram was drawn. Using the “grid,” “readr,” and “forestploter” packages, we

screened and plotted a forest plot of strongly associated genes based on the results of IVW and weighted median analyses. The “circlize” package was used to create a circular genomic map to display the chromosomal locations of key genes [18]. In the GSE51024 and TCGA datasets, differential expression validation of key genes was performed using the “limma,” “reshape2,” and “ggpubr” packages.

Functional enrichment analysis of key genes

Using the “clusterProfiler,” “org.Hs.eg.db,” and “enrichplot” packages, we conducted GO functional and KEGG pathway enrichment analyses on key genes ($P < 0.05$, adjusted $P < 0.05$). This analysis encompassed the examination of cellular components (CC), molecular functions (MF), biological processes (BP), and signaling pathways associated with the identified key genes, followed by the creation of graphical representations in the form of bubble charts. The GSEA analysis for each key gene’s high and low expression groups was conducted on the GEO training set expression matrix data by utilizing the “c2.cp.kegg.Hs.symbols” dataset sourced from the MSigDB database.

Immune cell infiltration and immune correlation analysis

Immune cell infiltration levels in the control and experimental groups of the GEO mesothelioma dataset were examined using the “LM22” file and the “CIBERSORT” package. Statistical significance was determined at a threshold of $P < 0.05$ with 1000 permutations. Visualization of the results was achieved through the generation of boxplots and bar plots utilizing the “corrplot” and “ggplot2” packages. Furthermore, the relationship between mesothelioma immune infiltrating cells and key genes was investigated using the “linkET” package.

Statistical analysis

Quantitative data is represented using the mean \pm standard deviation (SD). Kaplan-Meier method was used to plotted survival curve and log-rank test was performed. In consideration of other confounding factors and the impact of a suppressor effect, variables with a significant value of $P < 0.1$ were subjected to a multivariate Cox proportional hazard analysis and further screened by forward selection method to evaluate their independent effect. The Cox proportional hazard regression models were used for multivariate analysis, and the relative risks of dying were expressed as adjusted OR and corresponding 95% confidence intervals (CIs). Pearson’s analysis was used to analyze the correlation between disease critical genes and infiltrated immune cells. Statistical significance was determined by the independent paired *t*-test or one-way analysis of variance with Dunnett’s post hoc analysis

using SPSS software (version 24.0; IBM Corp.). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Results

DEGs identification

Batch correction was performed on the mesothelioma training set (Fig. 2A, B), with the training set details provided in Table 1. Differential gene expression analysis of the mesothelioma training set identified a total of 1608 DEGs, including 791 upregulated and 817 downregulated genes (Supplementary Table S1), as displayed in the volcano plot and heatmap (Fig. 2C, D).

Instrumental variables selection (IVs) and MR analysis

Following the elimination of linkage disequilibrium and weak IVs, a total of 26,152 SNPs were pinpointed as robust IVs strongly linked to genes (Supplementary Table S2). Using two-sample MR analysis to evaluate the effect of each SNP on mesothelioma, sensitivity analyses were conducted on SNPs with IVW $P < 0.05$ and without

pleiotropy. Consequently, 182 genes associated with mesothelioma were identified (Supplementary Table S3), including 90 high-risk genes and 92 low-risk genes. By intersecting high-risk and low-risk genes identified from MR analysis with upregulated and downregulated DEGs, respectively, we identified 14 key co-expressed genes. The human chromosomal locations of these 14 genes are shown in Fig. 3. Among them, the genes myelin protein zero like 1 (MPZL1), sterol O-acyltransferase 1 (SOAT1), transforming acidic coiled-coil containing protein 3 (TACC3), cytochrome b reductase 1 (CYBRD1) are associated with high risk of mesothelioma, while the genes transforming growth factor beta receptor 3 (TGFB3), NDRG family member 2 (NDRG2), endothelial PAS domain protein 1 (EPAS1), carboxypeptidase A3 (CPA3), myeloid cell nuclear differentiation antigen (MNDA), protein kinase C delta (PRKCD), microtubule associated scaffold protein 1 (MTUS1), arachidonate 15-lipoxygenase (ALOX15), leucine rich repeat neuronal 3 (LRRN3), integrin subunit alpha M (ITGAM) are

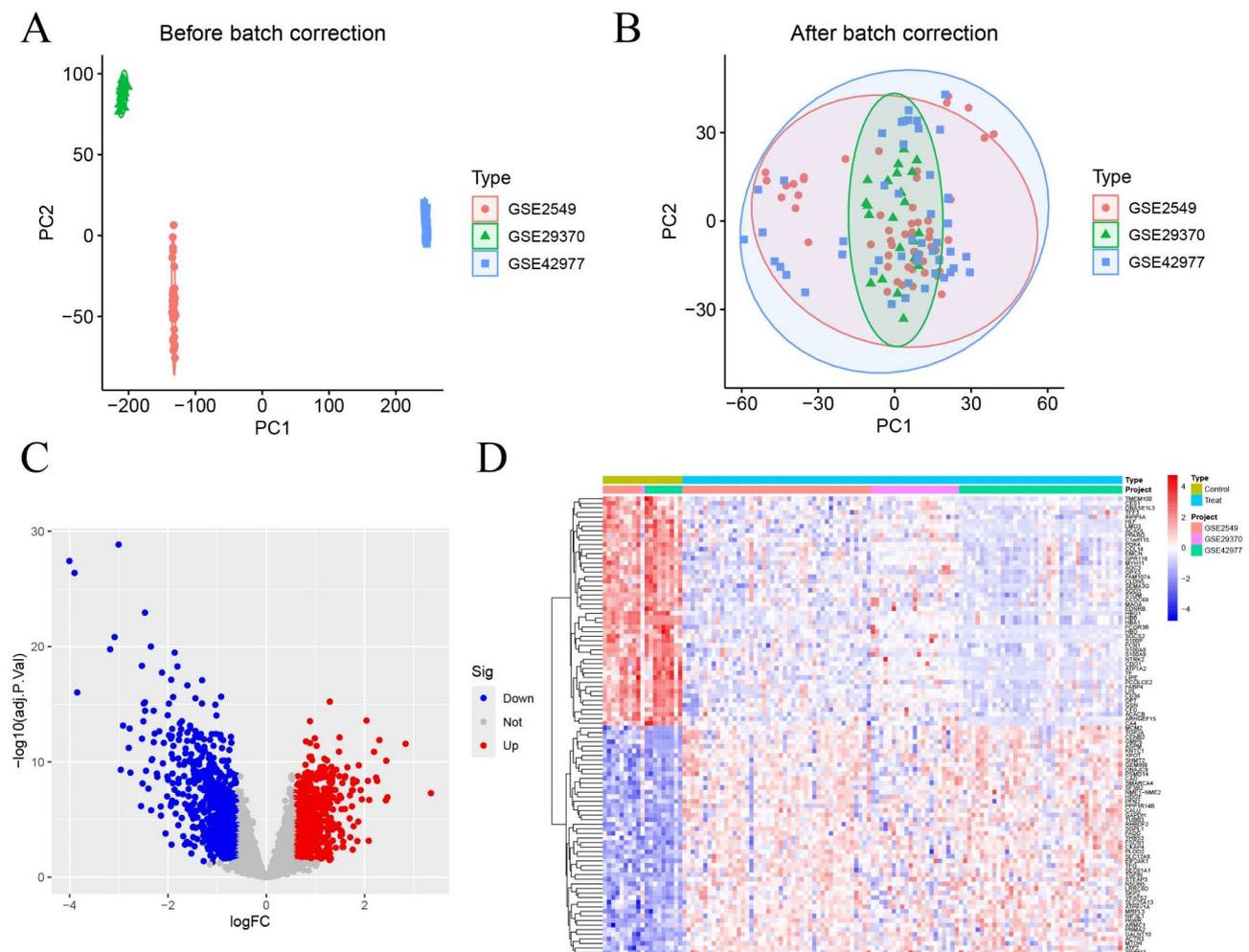


Fig. 2 Batch correction and variance analysis. **(A)** Before the batch correction. **(B)** After the batch correction. **(C)** Volcano plot of differential expression genes. **(D)** Heatmap of differential expression genes

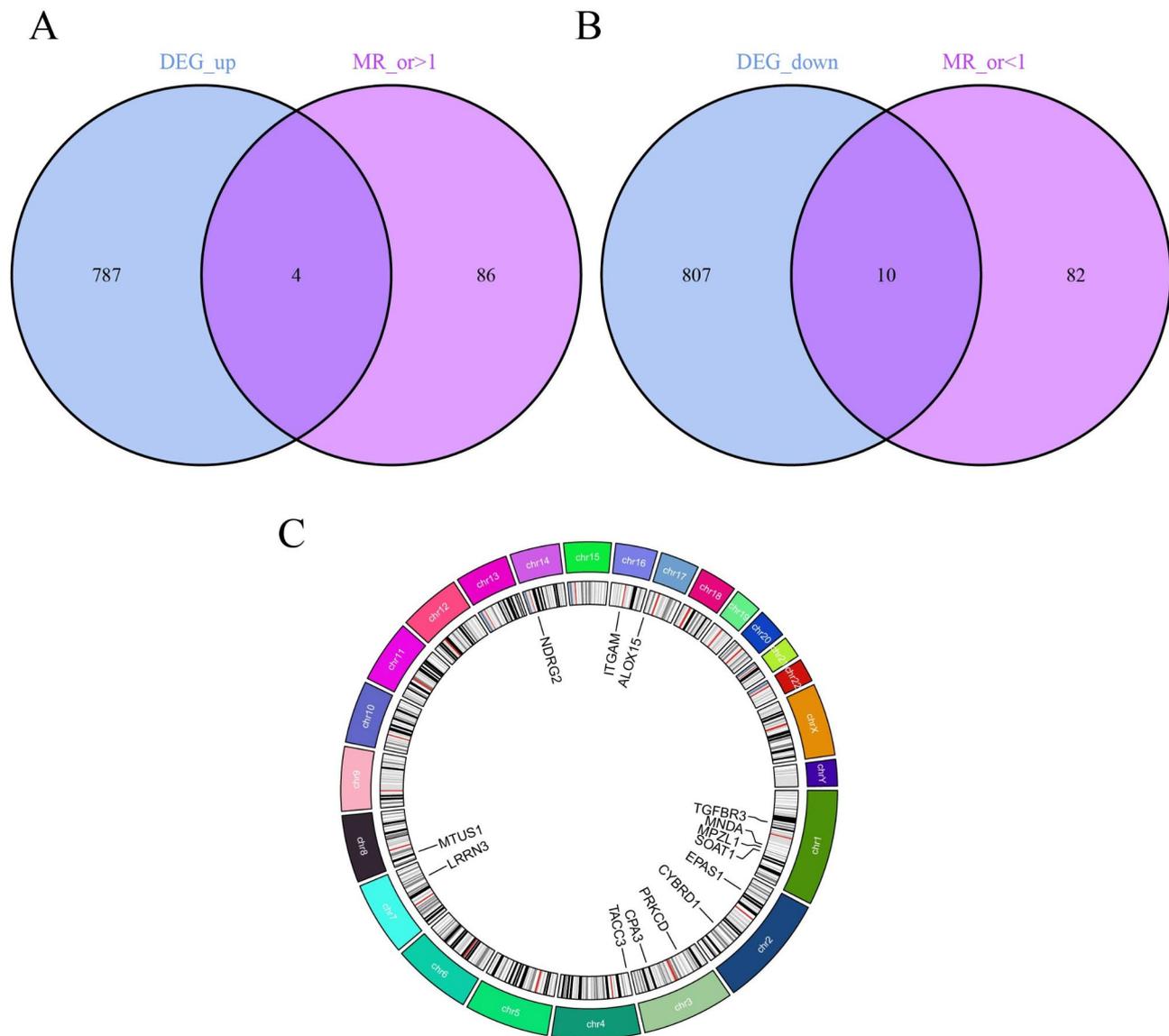


Fig. 3 Screening and localization of critical genes. **(A)** Disease upregulated DEGs are intersected with genes with OR values greater than one in the MR results. **(B)** Disease downregulated DEGs are intersected with genes with OR values less than one in the MR results. **(C)** Position of disease-critical genes on human chromosomes

low-risk genes for mesothelioma (Fig. 4). The MR analysis of the causal relationship between the 14 key genes and mesothelioma revealed that the expression levels of MPZL1 (OR=1.728; 95% CI: [1.143–2.612]; $P=0.009$), SOAT1 (OR=1.579; 95% CI: [1.050–2.374]; $P=0.028$), TACC3 (OR=1.964; 95% CI: [1.057–3.649]; $P=0.033$), and CYBRD1 (OR=1.771; 95% CI: [1.019–3.078]; $P=0.043$) were positively correlated with mesothelioma. Conversely, the expression levels of TGFBR3 (OR=0.252; 95% CI: [0.070–0.912]; $P=0.036$), NDRG2 (OR=0.442; 95% CI: [0.202–0.969]; $P=0.041$), EPAS1 (OR=0.207; 95% CI: [0.058–0.741]; $P=0.016$), CPA3 (OR=0.431; 95% CI: [0.200–0.930]; $P=0.032$), MND4 (OR=0.477; 95% CI: [0.261–0.874]; $P=0.017$), PRKCD (OR=0.425; 95%

CI: [0.205–0.880]; $P=0.021$), MTUS1 (OR=0.597; 95% CI: [0.360–0.992]; $P=0.046$), ALOX15 (OR=0.569; 95% CI: [0.351–0.924]; $P=0.022$), LRRN3 (OR=0.197; 95% CI: [0.058–0.670]; $P=0.009$), and ITGAM (OR=0.179; 95% CI: [0.043–0.751]; $P=0.019$) were negatively correlated with mesothelioma.

Gene sensitivity analysis and differential expression verification

We conducted sensitivity analyses on 14 key mesothelioma genes using MR-Egger regression and Cochran's test. The results indicated no heterogeneity or pleiotropy, confirming the reliability of our findings (Table 2). The funnel plot analysis indicated that no individual SNP

exposure	nsnp	method	pval	OR(95% CI)
ALOX15	3	Weighted median	0.028	0.571 (0.346 to 0.941)
	3	Inverse variance weighted	0.022	0.569 (0.351 to 0.924)
CPA3	8	Weighted median	0.017	0.319 (0.125 to 0.815)
	8	Inverse variance weighted	0.032	0.431 (0.200 to 0.930)
CYBRD1	9	Weighted median	0.042	1.965 (1.023 to 3.775)
	9	Inverse variance weighted	0.043	1.771 (1.019 to 3.078)
EPAS1	4	Weighted median	0.039	0.194 (0.041 to 0.918)
	4	Inverse variance weighted	0.016	0.207 (0.058 to 0.741)
ITGAM	4	Weighted median	0.092	0.221 (0.038 to 1.281)
	4	Inverse variance weighted	0.019	0.179 (0.043 to 0.751)
LRRN3	5	Weighted median	0.159	0.321 (0.066 to 1.559)
	5	Inverse variance weighted	0.009	0.197 (0.058 to 0.670)
MNDA	4	Weighted median	0.002	0.373 (0.200 to 0.694)
	4	Inverse variance weighted	0.017	0.477 (0.261 to 0.874)
MPZL1	11	Weighted median	0.086	1.622 (0.934 to 2.816)
	11	Inverse variance weighted	0.009	1.728 (1.143 to 2.612)
MTUS1	3	Weighted median	0.050	0.588 (0.346 to 1.000)
	3	Inverse variance weighted	0.046	0.597 (0.360 to 0.992)
NDRG2	7	Weighted median	0.042	0.387 (0.155 to 0.965)
	7	Inverse variance weighted	0.041	0.442 (0.202 to 0.969)
PRKCD	5	Weighted median	0.027	0.398 (0.176 to 0.900)
	5	Inverse variance weighted	0.021	0.425 (0.205 to 0.880)
SOAT1	5	Weighted median	0.055	1.547 (0.992 to 2.413)
	5	Inverse variance weighted	0.028	1.579 (1.050 to 2.374)
TACC3	4	Weighted median	0.034	1.997 (1.052 to 3.791)
	4	Inverse variance weighted	0.033	1.964 (1.057 to 3.649)
TGFB3	4	Weighted median	0.121	0.293 (0.062 to 1.380)
	4	Inverse variance weighted	0.036	0.252 (0.070 to 0.912)

Fig. 4 Disease critical genes causally associated with mesothelioma

affected the outcome, implying the absence of directional pleiotropy for individual SNP non-violation and bias estimation. The leave-one-out analysis further confirmed the absence of horizontal pleiotropy, demonstrating the robustness and reliability of our analytical methods and results (Fig. 5, Supplementary Fig. 1,2). Furthermore, this research examined the variations in expression of 14 pivotal genes in mesothelioma by utilizing the validation set GSE51024 dataset and the TCGA database. The findings indicated notable distinctions in the levels of expression of these critical genes in mesothelioma, thereby validating their differential expression (Fig. 6).

Functional enrichment analysis of key genes

The analysis of GO for the identified key genes demonstrated notable enrichment in various biological processes, including phagocytosis, superoxide anion generation, superoxide metabolism, and neutrophil activation. Furthermore, these genes exhibited enrichment in cellular components such as primary lysosomes, azurophilic granules lumen, and cytoplasmic vesicles. Additionally, they were found to be enriched in molecular functions such as sulfur compound binding, amide binding, and serine/threonine kinase activity (Fig. 7A, supplementary table S4). The KEGG analysis of the key genes revealed their enrichment in pathways such as ferroptosis, steroid biosynthesis, cholesterol metabolism, and the

Table 2 Sensitivity analysis of mesothelioma critical genes

Gene	$P_{MR-Egger}$	$P_{MR-Egger, Q}$	$P_{IVW, Q}$
ALOX15	0.969	0.477	0.776
CPA3	0.536	0.261	0.311
CYBRD1	0.347	0.911	0.882
EPAS1	0.825	0.598	0.779
ITGAM	0.477	0.707	0.694
LRRN3	0.412	0.659	0.644
MNDA	0.729	0.213	0.342
MPZL1	0.588	0.668	0.724
MTUS1	0.663	0.930	0.839
NDRG2	0.703	0.686	0.776
PRKCD	0.753	0.457	0.606
SOAT1	0.549	0.937	0.929
TACC3	0.536	0.348	0.442
TGFBR3	0.357	0.852	0.631

renin-angiotensin system (Fig. 7B, supplementary table S5).

The GSEA enrichment analysis of key genes

To investigate the enrichment differences of key genes in high and low expression groups in mesothelioma, gene set enrichment analysis was conducted on the key genes. The results of GSEA showed that the antigen processing and presentation signaling pathway had the highest enrichment level in the high expression group, followed by significant enrichment in the cytokine-cytokine receptor interaction and proteasome signaling pathways (Fig. 8). The key genes were most enriched in the cell cycle signaling pathway in the low expression group, followed by the proteasome, spliceosome and DNA replication signaling pathways (Fig. 9). These findings suggest that alterations in the expression levels of key genes may have an impact on the occurrence and development of mesothelioma at the levels of protein synthesis and metabolism, protein interactions, and even the cell cycle.

Analysis of immune cell infiltration levels and their association with key genes in mesothelioma

Using the CIBERSORT algorithm to measure differences in immune cell infiltration between mesothelioma and normal mesothelial tissues, we obtained the immune cell infiltration profile of mesothelioma. Compared to the control group, the proportions of CD4⁺ naive T cells and M1 macrophages were significantly increased in mesothelioma, suggesting their potential impact on the disease (Fig. 10A.B). Additionally, to investigate the immunological characteristics of the key genes, we conducted a correlation analysis between the key genes and immune infiltrating cells (Fig. 10C). The results indicate that key genes are correlated with various immune infiltrating cells. M1 macrophages show significant positive correlations with the expression levels of SOAT1, MNDA,

and ITGAM, and negative correlations with ALOX15 and LRRN3. Neutrophils exhibit positive correlations with the expression levels of MPZL1, CYBRD1, NDRG2, and MNDA, and negative correlations with TACC3, TGFBR3, and LRRN3. Additionally, the infiltration levels of T cells gamma delta significantly positively correlate with M1 macrophage infiltration levels. In contrast, the infiltration level of T cell follicular helper is significantly negatively correlated with T cell CD4 memory resting. The above research discovered that certain key genes interact with immune cells in mesothelioma. Specifically, the genes MNDA, TACC3, and CYBRD1 are significantly associated with infiltration of various immune cells.

Discussion

Mesothelioma is a rare, aggressive, and difficult-to-treat malignant tumor. Most patients are diagnosed at an advanced stage, with a median survival time of only 9-12 months and a 5-year survival rate of less than 10%. Consequently, mesothelioma ranks among the cancers with the poorest survival rates [19]. Asbestos is a significant causative factor in mesothelioma, inducing chromosomal and genetic damage in mammalian cells [20]. Mesothelioma is characterized by extensive chromosomal rearrangements, gene mutations, and deletions, primarily involving human chromosomes 9 (9p21), 21q, and 3 (3p21) [21]. Based on transcriptomic pathway analysis, mesothelioma cells exhibit alterations in cellular proliferation, apoptosis, differentiation, and migration. These alterations in pathways can disturb the homeostasis of normal mesothelial cells [22].

In this study, we employed MR to analyze eQTL data of 19,942 genes to investigate their causal relationship with mesothelioma. The MR analysis revealed 182 genes associated with the condition, which were classified into high-risk and low-risk groups according to their OR values. By intersecting high-risk genes with upregulated DEGs, we identified four key high-risk genes associated with mesothelioma: MPZL1, SOAT1, TACC3, and CYBRD1. Conversely, intersecting low-risk genes with downregulated DEGs revealed ten key low-risk genes: LRRN3, TGFBR3, NDRG2, EPAS1, CPA3, MNDA, PRKCD, MTUS1, ALOX15, and ITGAM. This study represents the first application of MR analysis using eQTL data integrated with transcriptome data to identify pivotal risk genes in mesothelioma. Additionally, we conducted functional enrichment analysis of these key genes and explored their immune correlations. This finding has the potential to offer new insights into investigating the mechanisms underlying mesothelioma and advancing the development of immunosuppressive therapy.

Although thoracoscopic biopsy can be used for early pathological diagnosis of mesothelioma, its application is limited due to significant harm to the patients [23]. To

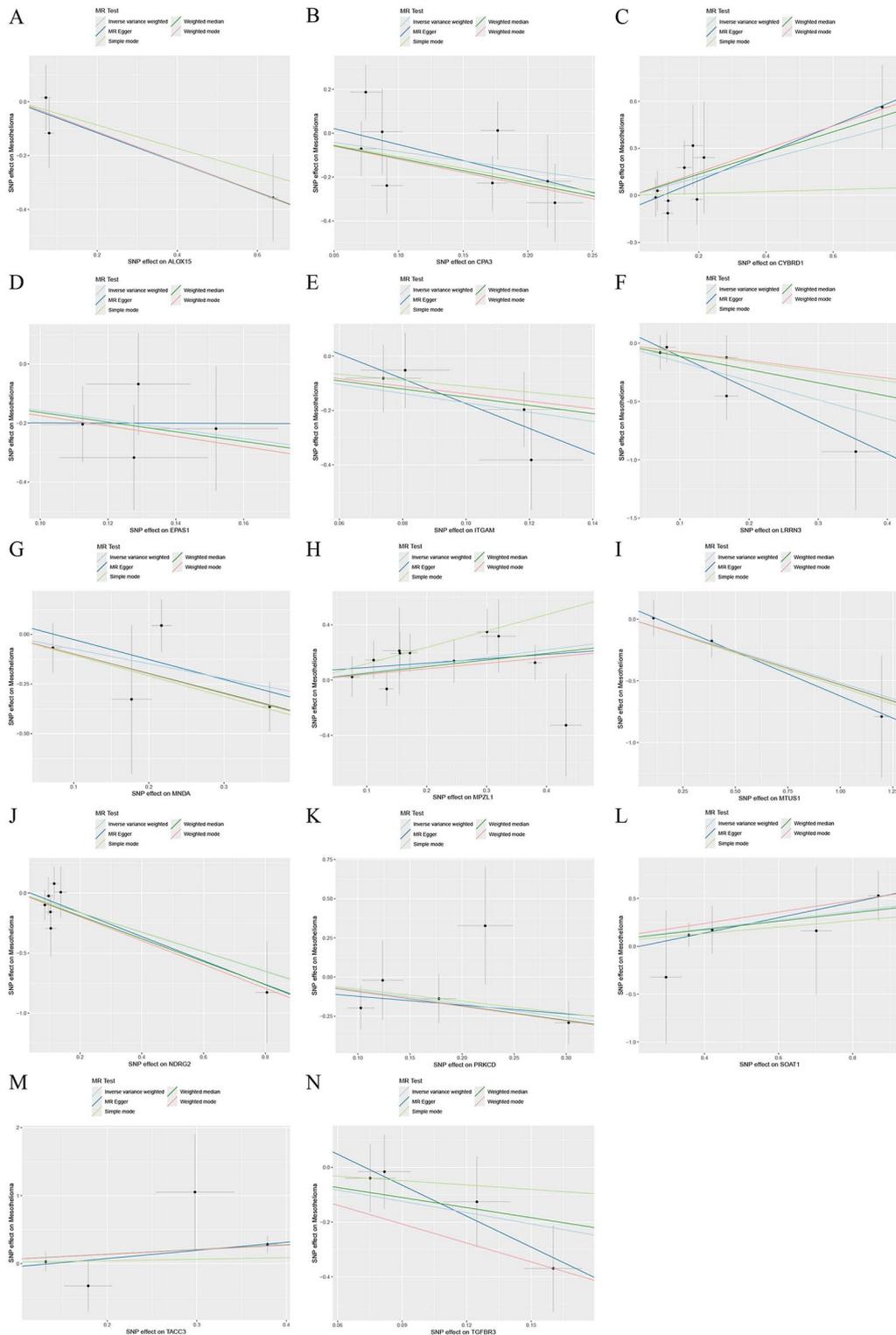


Fig. 5 Scatterplot of MR analysis of the association between mesothelioma critical genes and mesothelioma. **(A)** Scatterplot of MR analysis of ALOX15. **(B)** Scatterplot of MR analysis of CPA3. **(C)** Scatterplot of MR analysis of CYBRD1. **(D)** Scatterplot of MR analysis of EPAS1. **(E)** Scatterplot of MR analysis of ITGAM. **(F)** Scatterplot of MR analysis of LRRN3. **(G)** Scatterplot of MR analysis of MNDA. **(H)** Scatterplot of MR analysis of MPZL1. **(I)** Scatterplot of MR analysis of MTUS1. **(J)** Scatterplot of MR analysis of NDRG2. **(K)** Scatterplot of MR analysis of PRKCD. **(L)** Scatterplot of MR analysis of SOAT1. **(M)** Scatterplot of MR analysis of TACC3. **(N)** Scatterplot of MR analysis of TGFBR3

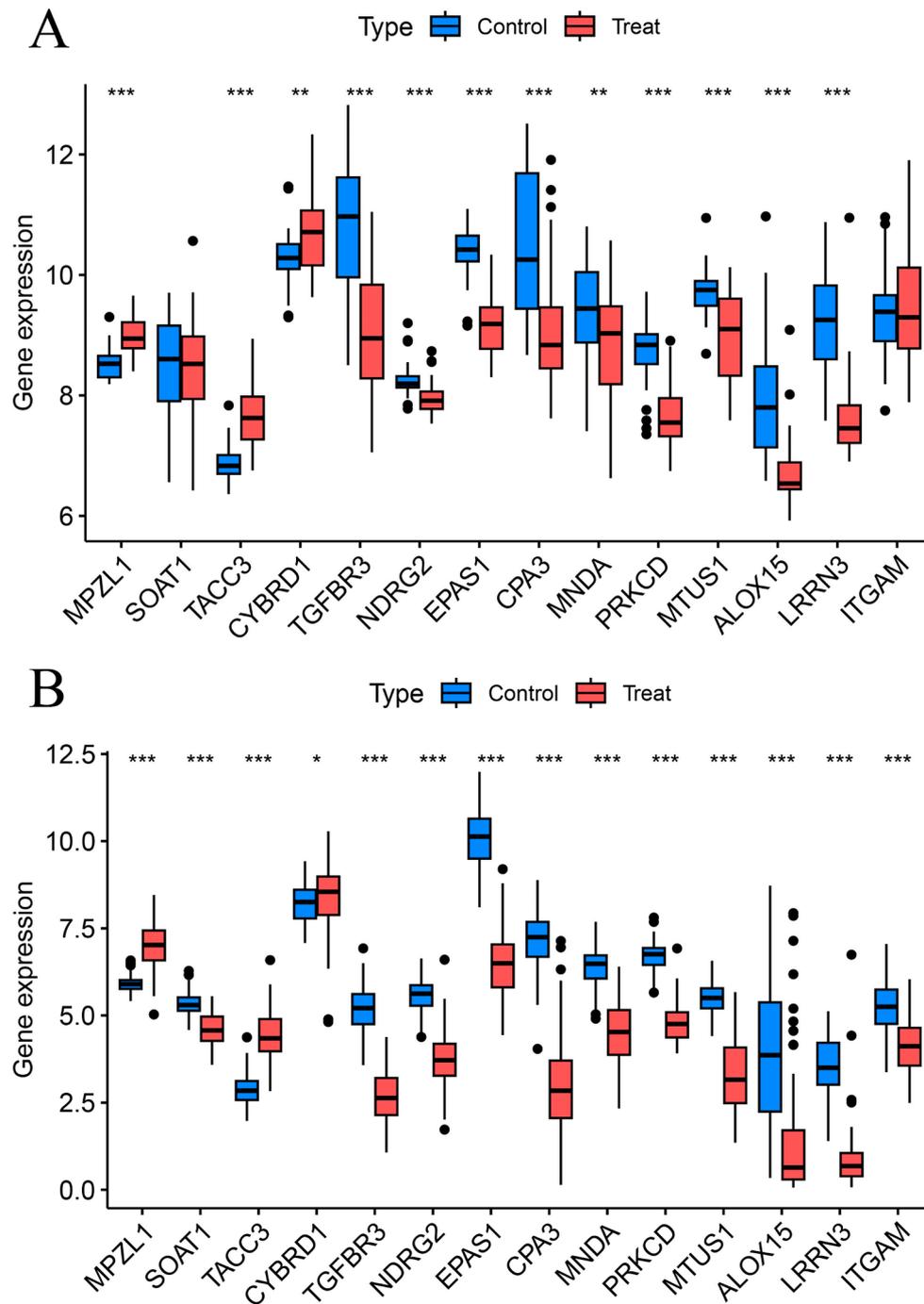


Fig. 6 Expression validation of disease critical genes in GEO testing group and TCGA database. **(A)** GEO testing group, **(B)** TCGA database. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. TCGA: The Cancer Genome Atlas. GEO: Gene Expression Omnibus.

date, as an environmental tumor associated with asbestos exposure, mesothelioma currently lacks clinically validated serum biomarkers for early diagnosis [24]. In this study, we identified the MPZL1 gene as one of the core genes with a strong causal association with mesothelioma. The MPZL1 gene is located on human chromosome 1q24.2 and contains 7 exons (Fig. 3C). MPZL1 is a member of the immunoglobulin superfamily and is

also known as the concanavalin receptor. As a single transmembrane protein, MPZL1's intracellular domain contains two immunoreceptor tyrosine-based inhibitory motifs (ITIMs) that specifically bind to the SH2 domains of the tyrosine phosphatase SHP-2/PTPN11 [25, 26]. In hepatocellular carcinoma, MPZL1 recruits SHP-2 to phosphorylate Src kinase at Tyr426, subsequently leading to cortactin phosphorylation, which promotes the

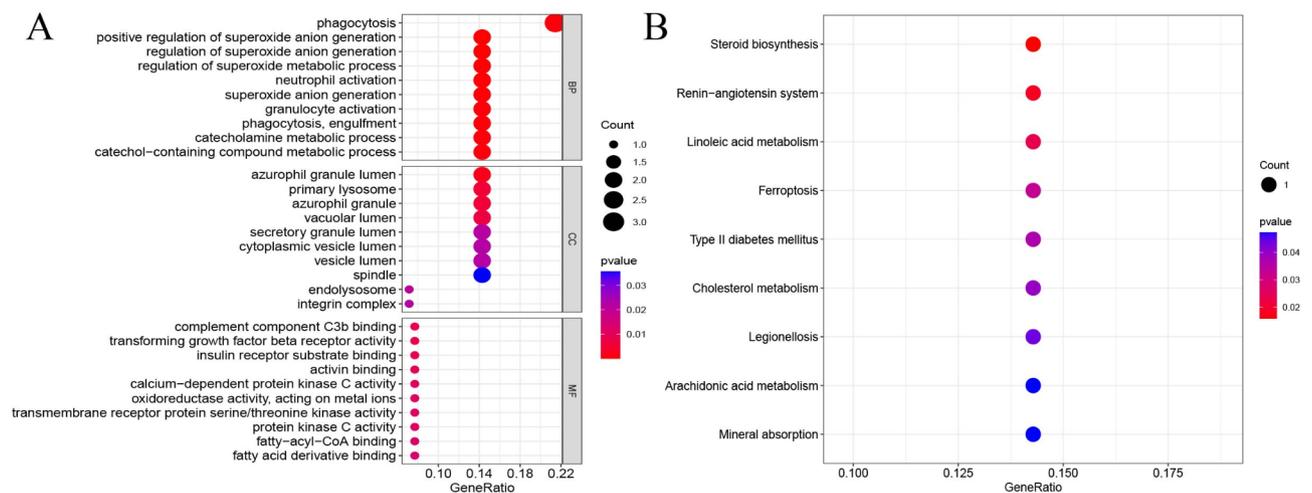


Fig. 7 Functional enrichment analysis of critical genes. **(A)** GO enrichment analysis of mesothelioma critical genes. **(B)** KEGG enrichment analysis of mesothelioma critical genes

migration of hepatocellular carcinoma cells [27]. Additionally, MPZL1 is highly expressed in various cancers, such as lung cancer, glioma, ovarian cancer, and gallbladder cancer [28–32]. MPZL1 promotes metastasis of non-small cell lung cancer (NSCLC) by upregulating COL11A1 and is associated with a poor prognosis in NSCLC [29]. MPZL1 can bind with β 1 transforming growth factor to promote the development of lung adenocarcinoma. The upregulation of MPZL1 in lung adenocarcinoma negatively regulates the infiltration of CD8 + T cells, which is associated with resistance to immunotherapy [32]. However, in another study, MPZL1 was found to inhibit the proliferation, metastasis, invasion, and cancer stem cell characteristics of lung adenocarcinoma by activating the β -catenin/TCF-4 signaling pathway [33], highlighting the complex role of MPZL1 in tumor development.

Additionally, we discovered that the genes SOAT1, TACC3, and NDRG2 are linked to the prognosis of mesothelioma patients. Elevated levels of SOAT1 and TACC3 were connected to shorter survival times ($p < 0.05$), while increased expression of NDRG2 was related to a more favorable prognosis (Supplementary Fig. 3). The SOAT1 gene is located on human chromosome 1q25.2 (Fig. 3C) and encodes a protein that resides in the endoplasmic reticulum. As a member of the acyltransferase family, it is primarily expressed in adrenal tissues [34]. The SOAT1 protein catalyzes the synthesis of fatty acid-cholesterol esters and promotes epithelial-mesenchymal transition (EMT) in hepatocellular carcinoma by regulating cholesterol metabolism [35]. Furthermore, the SOAT1 protein can activate the PI3K/AKT signaling pathway to promote lung cancer invasiveness by downregulating intracellular free cholesterol levels [36]. According to Huang's pan-cancer analysis, the SOAT1 gene exhibits

increased expression in various cancers, such as lung adenocarcinoma, lung squamous cell carcinoma, breast cancer, hepatocellular carcinoma, and gastric cancer. It is associated with infiltration of different immune cells, including T cells, neutrophils, and macrophages, as well as co-expression with numerous immune-related genes. These findings align with our analysis of SOAT1 gene immunoreactivity [37]. TACC3 is a transforming acidic coiled-coil protein expressed in the spindle, centrosome, and nucleus, it is involved in maintaining the stability of cell mitosis [38, 39]. In breast cancer, the loss or mutation of the p53 gene can upregulate TACC3 expression via FOXM1, leading to increased centrosome amplification (CA) and enhanced invasiveness [40]. Similarly, p53 is a commonly mutated tumor suppressor gene in mesothelioma, and TACC3 could be a downstream gene involved in its pathogenesis [41]. Additionally, TACC3 can promote the EMT of gastric cancer cells through the ERK/Akt/cyclin D1 signaling pathway, which is a critical process in the metastasis of mesothelioma [42, 43]. Similarly, literature reports that high TACC3 expression is associated with tumor immune cell infiltration and poor prognosis in lung adenocarcinoma patients, serving as an independent prognostic indicator [44, 45]. The NDRG2 gene, also known as SYLD, belongs to the hydrolase superfamily and is a cytoplasmic protein that promotes neurite growth. It is downregulated in various cancers, including breast and colorectal cancer, and is considered a negative indicator of tumor metastasis [46, 47]. The expression of the NDRG2 gene is downregulated in human lung cancer tissues and is considered a favorable prognostic indicator for lung cancer [48]. Salvianolic acid B (Sal B) can induce oxidative stress and inhibit the growth and metastasis of lung cancer A549 cells through the NDRG2/PTEN signaling pathway [49]. In a

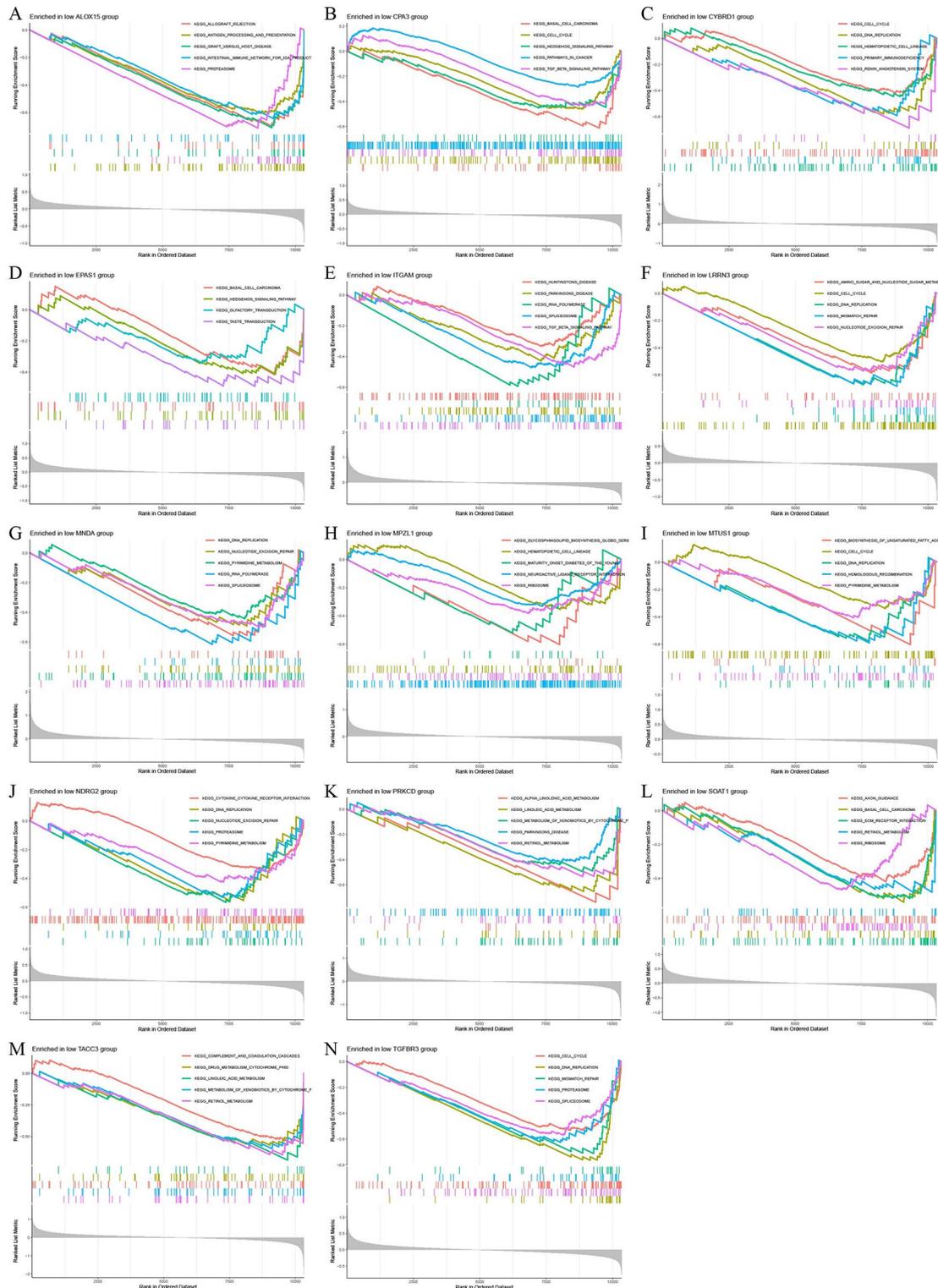


Fig. 9 GSEA enrichment analysis of disease critical genes in mesothelioma. **(A)** GSEA enrichment results of ALOX15 low expression group. **(B)** GSEA enrichment results of CPA3 low expression group. **(C)** GSEA enrichment results of CYBRD1 low expression group. **(D)** GSEA enrichment results of EPAS1 low expression group. **(E)** GSEA enrichment results of ITGAM low expression group. **(F)** GSEA enrichment results of LRRN3 low expression group. **(G)** GSEA enrichment results of MMDA low expression group. **(H)** GSEA enrichment results of MPZL1 low expression group. **(I)** GSEA enrichment results of MTUS1 low expression group. **(J)** GSEA enrichment results of NDRG2 low expression group. **(K)** GSEA enrichment results of PRKCD low expression group. **(L)** GSEA enrichment results of SOAT1 low expression group. **(M)** GSEA enrichment results of TACC3 low expression group. **(N)** GSEA enrichment results of TGFB3 low expression group

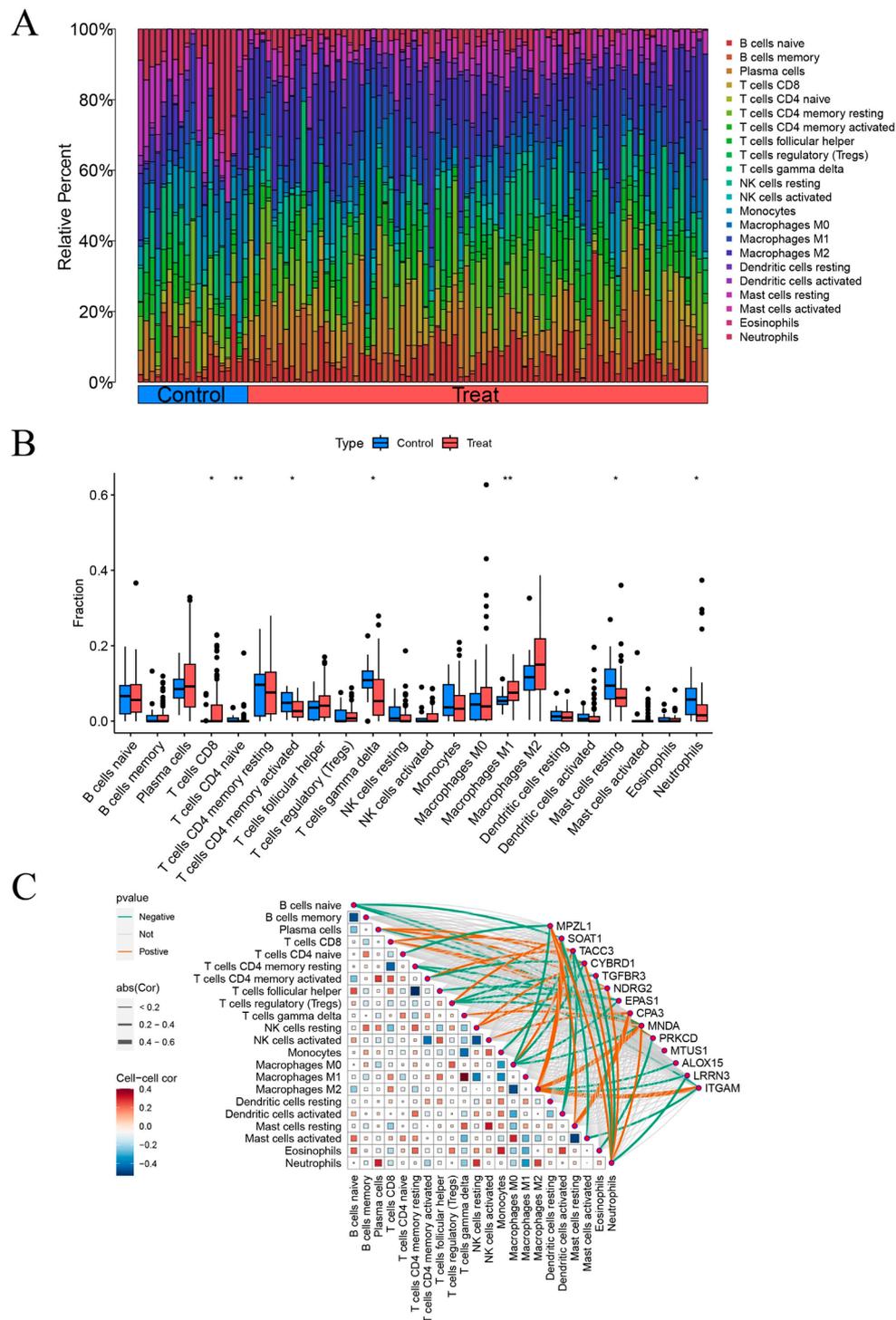


Fig. 10 Immune infiltration analyses, correlations between disease critical genes and infiltrating immune cell types. **(A)** Stacked histogram of the proportions of immune infiltration cells between control and mesothelioma groups. **(B)** Box plot of the infiltration level of immune cells between control and mesothelioma groups. **(C)** Correlation analysis between disease critical genes and infiltrated immune cells. * $P < 0.05$, ** $P < 0.01$

separate study, overexpression NDRG2 protein via a survivin promoter suppressed the viability and invasiveness of lung cancer cells [50]. Additionally, NDRG2 can negatively regulate the progression of small cell lung cancer through the PTEN-AKT-mTOR signaling pathway [51].

Currently, there are no research reports on the roles of SOAT1, TACC3, and NDRG2 in mesothelioma. Further investigation is needed to understand their functions in the progression of mesothelioma.

However, our study has some unavoidable limitations. Firstly, mesothelioma is a rare malignant tumor, leading to a limited sample size across multiple databases, such as GEO and TCGA database. Moreover, due to the absence of normal mesothelial samples in TCGA database, we utilized gene expression matrices from normal lung tissue for comparison and validation. Secondly, the GWAS data used for MR analysis were all derived from European populations, which may limit the generalizability of the findings to other populations. Lastly, our study developed a predictive model consisting of 14 genes. To improve prediction accuracy, further analysis and screening are required to identify the most pertinent key genes, which should then be validated through experimental studies.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12931-025-03219-4>.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We express our sincerest gratitude to all the study participants of the GWAS data, as well as to all the organizations that provided fund support.

Author contributions

J.S.L: Data curation and analysis, Visualization, Writing—original draft, Writing—review & editing. X.M.W: visualization, sensitivity analysis. Y.R.L: Data acquisition. Z.L.L and W.X: Fund support, conceptualization, data management, methodology, supervision.

Funding

This research was funded by the National Natural Science Foundation of China (Nos. 32160167 and 82160516). Yunnan Provincial Applied Basic Research Program General Project (Nos. 202201AT070004 and 202301AT070023). Key Program of Yunnan Provincial Applied Basic Research Program (No. 202001BB0500080). Yunnan Province Ten Thousand Talent Program (2019). Yunnan Provincial Local University Joint Project—Youth Project (No. 202101BA070001–282). Scientific Research Fund of Yunnan Provincial Department of Education (Nos. 2022J0688, 2022J0716 and 2022Y808). The Dali City Science and Technology Planning Project Support (No. 2021KBG032).

Data availability

All original data for this study are in the article or supplementary material, and all GWAS data as well as transcriptomic data are publicly available.

Declarations

Ethical approval

All GWAS pooled data were ethically approved by the respective institutional review boards. The studies were conducted in accordance with local legislative and institutional requirements. Written informed consent was not required to be obtained from the participants or legal guardians/next of kin of the participants as per the national legislative and institutional requirements.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Biochemistry and Molecular Biology, College of Basic Medical Sciences, Dali University, Dali, Yunnan, China

²Key Laboratory of Clinical Biochemistry Testing in Universities of Yunnan Province, College of Basic Medical Sciences, Dali University, Dali, Yunnan, China

³Yunnan Provincial Key Laboratory of Entomological Biopharmaceutical R&D, College of Pharmacy, Dali University, Dali, Yunnan, China

⁴Department of Radiology, The First Affiliated Hospital of Dali University, Dali University, Dali, Yunnan, China

Received: 6 August 2024 / Accepted: 2 April 2025

Published online: 13 April 2025

References

- Alpert N, et al. Epidemiology of mesothelioma in the 21(st) century in Europe and the united States, 40 years after restricted/banned asbestos use. *Transl Lung Cancer Res.* 2020;9(Suppl 1):S28–38.
- Nash A, et al. New markers for management of mesothelioma. *Semin Respir Crit Care Med.* 2023;44(4):491–501.
- Opitz I, et al. ERS/ESTS/EACTS/ESTRO guidelines for the management of malignant pleural mesothelioma. *Eur J Cardiothorac Surg.* 2020;58(1):1–24.
- Yap TA, et al. Efficacy and safety of pembrolizumab in patients with advanced mesothelioma in the open-label, single-arm, phase 2 keynote-158 study. *Lancet Respir Med.* 2021;9(6):613–21.
- Zhang Y, et al. An overview of detecting gene-trait associations by integrating Gwas summary statistics and Eqtls. *Sci China Life Sci.* 2024;67(6):1133–54.
- Erdin CA, et al. Mendelian randomization. *JAMA.* 2017;318(19):1925–6.
- Thanassoulis G, et al. Mendelian randomization: Nature's randomized trial in the post-genome era. *JAMA.* 2009;301(22):2386–8.
- Zhu Z, et al. Integration of summary data from Gwas and Eqtl studies predicts complex trait gene targets. *Nat Genet.* 2016;48(5):481–7.
- Gordon GJ, et al. Identification of novel candidate oncogenes and tumor suppressors in malignant pleural mesothelioma using large-scale transcriptional profiling. *Am J Pathol.* 2005;166(6):1827–40.
- De Rienzo A, et al. Sequential binary gene ratio tests define a novel molecular diagnostic strategy for malignant pleural mesothelioma. *Clin Cancer Res.* 2013;19(9):2493–502.
- Yoshikawa Y, et al. Frequent deletion of 3p21.1 region carrying semaphorin 3 g and aberrant expression of the genes participating in semaphorin signaling in the epithelioid type of malignant mesothelioma cells. *Int J Oncol.* 2011;39(6):1365–74.
- Hemani G et al. The mr-base platform supports systematic causal inference across the human phenome. *Elife.* 2018; 7.
- Burgess S, et al. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med.* 2011;30(11):1312–23.
- Obchain V, et al. Variantannotation: A bioconductor package for exploration and annotation of genetic variants. *Bioinformatics.* 2014;30(14):2076–8.
- Yavorska OO, et al. Mendelianrandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol.* 2017;46(6):1734–9.
- Lin Z, et al. Combining the strengths of inverse-variance weighting and Egger regression in Mendelian randomization using a mixture of regressions model. *PLoS Genet.* 2021;17(11):e1009922.
- van Kippersluis H, et al. Pleiotropy-robust Mendelian randomization. *Int J Epidemiol.* 2018;47(4):1279–88.
- Gu Z, et al. Circlize implements and enhances circular visualization in R. *Bioinformatics.* 2014;30(19):2811–2.
- Huang J, et al. Global incidence, risk factors, and Temporal trends of mesothelioma: A population-based study. *J Thorac Oncol.* 2023;18(6):792–802.
- Gariazzo C, et al. Asbestos consumption and malignant mesothelioma mortality trends in the major user countries. *Ann Glob Health.* 2023;89(1):11.
- Wadowski B, et al. The molecular basis of malignant pleural mesothelioma. *Thorac Surg Clin.* 2020;30(4):383–93.
- B Chapel D, et al. Molecular pathways and diagnosis in malignant mesothelioma: A review of the 14th international conference of the international mesothelioma interest group. *Lung Cancer.* 2019;127:69–75.
- Husain AN, et al. Guidelines for pathologic diagnosis of malignant mesothelioma 2017 update of the consensus statement from the international mesothelioma interest group. *Arch Pathol Lab Med.* 2018;142(1):89–108.
- Vorster T et al. The association between the histological subtypes of mesothelioma and asbestos exposure characteristics. *Int J Environ Res Public Health.* 2022;19(21).

25. Zhao R, et al. Dissecting the interaction of shp-2 with Pzr, an Immunoglobulin family protein containing immunoreceptor tyrosine-based inhibitory motifs. *J Biol Chem*. 2000;275(8):5453–9.
26. Zhao R, et al. Identification of a variant form of Pzr lacking immunoreceptor tyrosine-based inhibitory motifs. *Biochem Biophys Res Commun*. 2003;303(4):1028–33.
27. Yeh YT, et al. Amplification of mpzl1/pzr gene in hepatocellular carcinoma. *Hepatobiliary Surg Nutr*. 2014;3(2):87–90.
28. Chen D, et al. Mpzl1 promotes tumor cell proliferation and migration via activation of Src kinase in ovarian cancer. *Oncol Rep*. 2019;42(2):679–87.
29. Feng J, et al. Mpzl1 upregulation promotes tumor metastasis and correlates with unfavorable prognosis in non-small cell lung cancer. *Carcinogenesis*. 2022;43(10):919–29.
30. Liu X, et al. Mpzl1 is highly expressed in advanced gallbladder carcinoma and promotes the aggressive behavior of human gallbladder carcinoma gbc–sd cells. *Mol Med Rep*. 2019;20(3):2725–33.
31. Song H, et al. Mpzl1 as an Hgf/met signaling amplifier promotes cell migration and invasion in glioblastoma. *Genes Dis*. 2024;11(5):101085.
32. Wang A, et al. Mpzl1 promotes lung adenocarcinoma progression by enhancing tumor proliferation, invasion, migration, and suppressing immune function via transforming growth factor- β 1. *Hum Gene Ther*. 2023;34(11–12):540–53.
33. Ge Q, et al. Mpzl1 suppresses the cancer stem-like properties of lung cancer through β -catenin/tcf4 signaling. *Funct Integr Genomics*. 2023;23(4):304.
34. Harned TC et al. Acute acat1/soat1 Blockade increases Mam cholesterol and strengthens er-mitochondria connectivity. *Int J Mol Sci*. 2023; 24(6).
35. Fu R, et al. Soat1 regulates cholesterol metabolism to induce Emt in hepatocellular carcinoma. *Cell Death Dis*. 2024;15(5):325.
36. Mo Y, et al. Soat1 enhances lung cancer invasiveness by stimulating akt-mediated mitochondrial fragmentation. *Biochem Cell Biol*. 2022;100(1):68–74.
37. Huang Y, et al. A pan-cancer analysis identifies soat1 as an immunological and prognostic biomarker. *Oncol Res*. 2023;31(2):193–205.
38. Ding ZM, et al. The role of tacc3 in mitotic spindle organization. *Cytoskeleton (Hoboken)*. 2017;74(10):369–78.
39. Saatci O, et al. Tacc3: A multi-functional protein promoting cancer cell survival and aggressiveness. *Cell Cycle*. 2023;22(23–24):2637–55.
40. Saatci O, et al. Targeting tacc3 represents a novel vulnerability in highly aggressive breast cancers with centrosome amplification. *Cell Death Differ*. 2023;30(5):1305–19.
41. Bueno R, et al. Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. *Nat Genet*. 2016;48(4):407–16.
42. Akanda MR, et al. Tacc3 promotes gastric carcinogenesis by promoting epithelial-mesenchymal transition through the Erk/akt/cyclin d1 signaling pathway. *Anticancer Res*. 2021;41(7):3349–61.
43. Ramundo V et al. The epithelial-to-mesenchymal transition (emt) in the development and metastasis of malignant pleural mesothelioma. *Int J Mol Sci*. 2021; 22(22).
44. Chen Y et al. High expression of tacc3 is associated with the poor prognosis and immune infiltration in lung adenocarcinoma patients. *Dis Markers*. 2022; 2022:8789515.
45. Zhang X, et al. Abnormal tacc3 expression is an independent prognostic biomarker in lung carcinoma. *Front Biosci (Landmark Ed)*. 2022;27(8):252.
46. Lee KW et al. The function of n-myc downstream-regulated gene 2 (ndrg2) as a negative regulator in tumor cell metastasis. *Int J Mol Sci*. 2022; 23(16).
47. Morishita K, et al. Pathophysiological significance of n-myc downstream-regulated gene 2 in cancer development through protein phosphatase 2a phosphorylation regulation. *Cancer Sci*. 2021;112(1):22–30.
48. Li SJ, et al. Expression of ndrg2 in human lung cancer and its correlation with prognosis. *Med Oncol*. 2013;30(1):421.
49. Yang Y, et al. Salvianolic acid B inhibits the growth and metastasis of a549 lung cancer cells through the ndrg2/pten pathway by inducing oxidative stress. *Med Oncol*. 2024;41(7):170.
50. Fanian M, et al. Targeted overexpression of ndrg2 using survivin promoter reduces viability and invasiveness of a549 cell line. *Avicenna J Med Biotechnol*. 2023;15(2):84–90.
51. Ma Z, et al. Ndr2 acts as a negative regulator of the progression of small-cell lung cancer through the modulation of the pten-akt-mtor signalling cascade. *Toxicol Appl Pharmacol*. 2024;485:116915.

Publisher's note

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