Preplanned Studies

A Multi-omics Framework Combining Genomics and Proteomics for Silicosis Prediction in Chinese Workers — Jiangsu Province, China, 2023–2024

Furu Wang[1,2]; Qianqian Gao[2]; Chenjie Li[1]; Lei Han[2]; Chuanfeng Zhang[2]; Zhengdong Zhang[1,#]; Baoli Zhu[2,#]

Summary

What is already known about this topic?

Currently, the detection of silicosis relies on imaging and pulmonary function tests, which are effective only at identifying the advanced stages. Additionally, no effective protein biomarkers or genetic risk models exist for the early detection or targeted intervention of silicosis.

What is added by this report?

This study integrates genomics and proteomics to identify new genetic loci associated with susceptibility to silicosis. Using Mendelian randomization and protein quantitative trait loci (pQTL) analysis, 2 functionally significant genetic variants [rs6677666 (WLS) and rs2272528 (COL4A4)] and 5 protein biomarkers (MMP12, EGF, Gal_9, GZMA, and ICOSLG) mechanistically linked to silicosis pathogenesis were identified. A diagnostic causal protein risk score (CPRS) model was then constructed to provide a robust tool for early detection in high-risk populations.

What are the implications for public health practice?

These findings provide new insights into the early diagnosis of silicosis, and support the development of preventive and screening strategies for populations at risk, enhancing public health policies for the control and management of silicosis.

ABSTRACT

Objective: In this study, we aimed to identify novel genetic loci and protein biomarkers associated with silicosis susceptibility in Chinese workers through integrated proteomic and genomic analyses and to develop an early diagnostic prediction model.

Methods: A genome-wide association study (GWAS) was conducted on 163 patients with silicosis and 183 controls, followed by Olink proteomic profiling of 92 plasma proteins. Protein quantitative trait loci (pQTL) mapping, Mendelian randomization (MR), and Bayesian co-localization were used to infer causal relationships. A causal protein risk score (CPRS) model integrating genetic and proteomic data was developed and validated using 10-fold cross-validation.

Results: GWAS identified 16 novel risk loci ($P<1\times10^{-5}$), including rs6677666 (WLS) and rs2272528 (COL4A4). MR analysis revealed eight plasma proteins associated with silicosis risk, with MMP12, EGF, Gal_9, GZMA, and ICOSLG showing significant differential expression ($P<0.05$). The CPRS model combining these proteins demonstrated a high diagnostic accuracy (AUC=0.915), outperforming traditional clinical variables.

Conclusion: This multi-omics study uncovered genetic and proteomic markers linked to silicosis susceptibility and established a robust predictive model. The integration of GWAS and proteomics offers novel insights into the pathogenesis of silicosis, and supports development of early detection and prevention policies for high-risk populations.

Silicosis is an occupational lung disease caused by the inhalation of respirable crystalline silica (RCS) dust and is characterized by chronic inflammation, collagen deposition, and fibrotic lesions that drive a continuous progression from subclinical pathology to severe lung tissue damage (*1*). Currently, no effective cure exists (*2*). Early detection and timely intervention are critical strategies for extending the life expectancy of individuals with silicosis. Recent advances in high-throughput technologies and bioinformatics have enabled comprehensive multi-omics analyses. Therefore, in this study, we aimed to identify novel protein biomarkers and potential therapeutic targets for silicosis to inform public health strategies for its prevention and treatment. First, we conducted a

genome-wide association study (GWAS) involving 163 patients with silicosis and 183 controls, followed by Olink proteomic profiling of 92 plasma proteins. Protein quantitative trait loci (pQTL) analysis, Mendelian randomization (MR), and Bayesian co-localization were used to infer causal relationships. Finally, a causal protein–based risk score, incorporating proteins showing consistent causal evidence, was developed and evaluated for its association with silicosis risk and predictive performance using appropriate statistical models and validation techniques.

Study participants included patients with silicosis undergoing medical follow-up and retired workers with a history of silica dust exposure but without a diagnosis of pneumoconiosis (control group). The control group was matched to the silicosis group based on the duration of silica dust exposure. Data collection included demographics (sex and age), clinical parameters (pneumoconiosis type), occupational factors (occupation type and cumulative dust exposure), lifestyle factors (smoking and drinking habits), and workplace safety information (personal protective equipment usage). The silicosis classification followed the International Labour Organization (ILO) guidelines (*3*).

GWAS was conducted using 163 patients and 183 controls selected from the aforementioned participants. Genotyping was performed using Illumina Infinium Asian Screening Array-24 v1.0 BeadChip platform (Illumina Inc., San Diego, CA, USA), which was specifically designed for East Asian populations. This chip contains approximately 670,000 markers and provides robust coverage of low-frequency variants [minor allele frequency (MAF)=1%–5%] common in the East Asian region. The baseline characteristics of the participants are provided in the Supplementary Table S1 (available at https://weekly.chinacdc.cn/).

Following quality control, 157 patients with silicosis and 182 controls were genotyped. A total of 7,897,033 autosomal single-nucleotide polymorphisms (SNPs) passed the quality control standards and were included for further analysis. Manhattan and quantile–quantile plots for the GWAS results are presented in Supplementary Figure S1 (available at https://weekly.chinacdc.cn/). After adjusting for potential confounding factors, including age, sex, dust concentration, and pack-years of smoking, along with the first 10 principal components, we identified 42,210 risk loci with $P<0.05$. Among these, 16 independent risk loci achieved genome-wide significance ($P<1\times10^{-5}$): (Table 1 and Supplementary Figure S2, available at https://weekly.chinacdc.cn/). However, these loci have not been previously reported to be associated with silicosis.

Among these, rs3738756, rs28536654, rs4776983,

TABLE 1. Sixteen newly identified risk loci for silicosis in the Chinese population.

UniqID	rsID	chr	pos	Allele A	Allele B	Annotation	*OR*	*P* (1×10^{-5})	nSNPs	Nearest gene
1:68640101:C:T	rs6677666	1	68640101	C	T	Intronic	0.373	0.505	25	WLS
1:110658260:C:T	rs3738756	1	110658260	C	T	Intergenic	2.555	0.339	32	UBL4B
2:227976749:C:G	rs2272528	2	227976749	C	G	Intronic	0.378	0.316	22	COL4A4
4:57963611:A:G	rs4865182	4	57963611	A	G	Intronic	0.417	0.695	4	IGFBP7
5:26013029:A:G	rs6885607	5	26013029	A	G	Exonic	0.109	0.709	30	RNU4-43P
5:26324525:A:G	rs112721803	5	26324525	A	G	Intronic	0.136	0.365	104	RP11-351N6.1
5:164319653:G:T	rs17070163	5	164319653	G	T	Intronic	0.418	0.464	11	LINC03000
6:112154244:A:T	rs1391373	6	112154244	A	T	Intergenic	0.314	0.872	93	HLA-DQA1
7:34543087:C:T	rs7802435	7	34543087	C	T	Intronic	0.427	0.877	4	NPSR1-AS1
10:47737036:C:T	rs28536654	10	47737036	C	T	Upstream	2.250	0.848	1	Lnc-ANXA8L1-1
12:113444117:C:G	rs2240185	12	113444117	C	G	Intronic	0.427	0.666	10	OAS2
15:68115177:A:G	rs4776983	15	68115177	A	G	Intronic	5.882	0.072	1	SKOR1
16:83103233:C:T	rs9923246	16	83103233	C	T	Intronic	2.653	0.101	5	CDH13
16:86383556:A:G	rs7191547	16	86383556	A	G	Intronic	0.368	0.769	3	LINC00917
18:35461186:A:G	rs28814664	18	35461186	A	G	Intergenic	0.307	0.965	42	MIR4318
21:31135702:C:T	rs72551940	21	31135702	C	T	Exonic	0.339	0.774	2	GRIK1

Abbreviation: SNPs=single-nucleotide polymorphisms; *OR*=odds ratio.

and rs9923246 were positively correlated ($P<1\times10^{-5}$) with the risk of developing silicosis. Gene expression regulation analysis revealed that the G mutation of rs4865182 downregulated *IGFBP7* gene expression in the blood. *IGFBP7* is a detrimental factor that exacerbates acute inflammatory responses such as acute lung injury (*4*).

Genome-wide pQTL analysis of 182 healthy controls identified 689 significant cis-pQTL variant–protein interactions ($P<0.05$). In addition to transcriptional regulatory regions, such as promoters and enhancers, pQTL variants were also found in post-transcriptional regulatory regions, such as untranslated regions (UTRs). The cis-pQTL variants were predominantly located in the intronic regions of adjacent genes (51%), suggesting that introns play a significant role in the regulation of protein expression.

Proteomic MR analysis identified eight proteins that were significantly correlated with silicosis ($P<0.05$). Four proteins showed positive associations: MMP12 [odds ratio (*OR*)=4.717, $P=0.027$], CD244 (*OR*=3.251, $P=0.039$), GZMA (*OR*=7.520, $P=0.002$), and ICOSLG (*OR*=12.351, $P=0.016$); four proteins demonstrated negative effects: EGF (*OR*=0.129, $P=0.014$), MUC_16 (*OR*=0.442, $P=0.005$), Gal_9 (*OR*=0.022, $P=0.003$), and CD28 (*OR*=0.367, $P=0.012$) (Figure 1).

To assess the causal relation between proteins and different stages of silicosis, we performed a GWAS

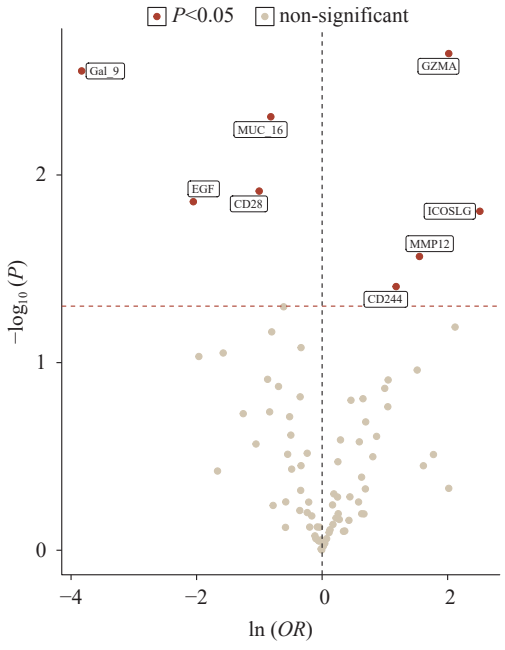


FIGURE 1. Protein-wide MR of cis-pQTLs and silicosis.
Abbreviation: MR=mendelian randomization.

stratified by disease stage. A total of 134 patients with stage I disease vs. healthy controls and 6,172,798 variants were included in the subsequent analysis. MR and pQTL analyses identified CD28, GZMA, and MMP12 as causally associated with Stage I disease. In an analysis of 22 patients with stage II disease *vs.* healthy controls, only CD244 exhibited a causal association.

Among the eight proteins with causal associations, we selected five (MMP12, EGF, Gal_9, GZMA, and ICOSLG) that demonstrated significant differential expression between patients and healthy controls ($P<0.05$) to develop a Combined Protein Risk Score (CPRS). The CPRS exhibited excellent diagnostic capability for detecting silicosis, with an area under the receiver operating characteristic (ROC) curve (AUC) of 0.808 [95% confidence interval (*CI*): 0.763, 0.853; Model 1]. The traditional model including age and years of exposure, also demonstrated good diagnostic performance (ROC=0.844, 95% *CI*: 0.801, 0.888; Model 0). When CPRS and clinical variables were incorporated into Model 2, the diagnostic performance significantly improved ($P=1.546\times10^{-5}$), reaching an AUC of 0.915 (95% *CI*: 0.873, 0.937) (Figure 2A). To validate the robustness of our findings, we performed a 10-fold cross-validation with 500 repetitions. The average AUC was 0.846 (95% *CI*: 0.802, 0.890) for Model 0, 0.809 (95% *CI*: 0.764, 0.854) for Model 1, and 0.908 (95% *CI*: 0.877, 0.939) for Model 2 (Figure 2B).

DISCUSSION

Our study identified 16 novel loci and eight plasma proteins associated with susceptibility to silicosis, among which two independent SNPs (rs6677666 and rs2272528) demonstrated particularly significant biological relevance. The rs6677666 variant, located within the intron of the WLS gene, may influence pulmonary fibrosis through Wnt/β-catenin signaling activation and epithelial–mesenchymal transition (EMT). Similarly, rs2272528 in COL4A4 promotes fibrosis through collagen IV overexpression and extracellular matrix (ECM) accumulation (*5*). The ECM provides structural support as an extracellular scaffold, whereas EMT involves transformation of epithelial cells into mesenchymal cells. Both of these processes contribute to tissue fibrosis when dysregulated. Although the specific functions of RNU4-43P and LINC03000 in silicosis are uncharacterized, emerging evidence indicates that

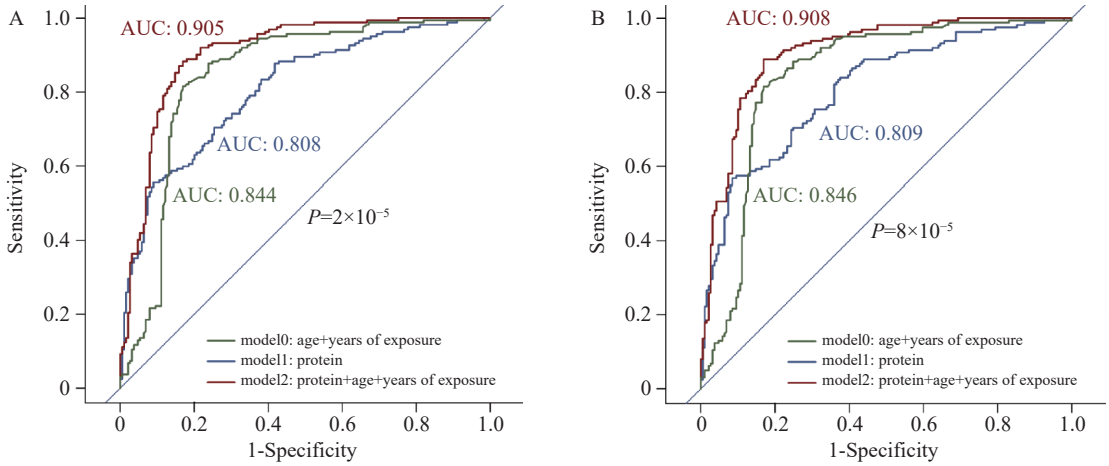


FIGURE 2. ROC curves of the protein diagnostic model. (A) ROC curves of the protein diagnostic model; (B) ROC curves after the robustness test.
Abbreviation: ROC=receiver operating characteristic; AUC=area under the curve.

pseudogenes can act as competitive endogenous RNAs to regulate parental genes (*6*), whereas long non-coding RNAs (lncRNAs) can directly modulate profibrotic pathways such as TGF-β (*7*). Thus, we speculated that RNU4-43P and LINC03000 may influence silicosis susceptibility by similarly affecting mRNA splicing stability or pro-fibrotic gene networks. Using MR analysis with pQTLs as instrumental variables, we established causal relationships between eight proteins and silicosis. Subsequent validation identified five significantly differentially expressed proteins (MMP12, EGF, Gal_9, GZMA, and ICOSLG) between patients and controls. These biomarkers are biologically relevant: MMP12 contributes to elastin degradation and fibrosis, with elevated levels confirmed in silica-exposed workers (*8*); EGF promotes pulmonary epithelial/fibroblast proliferation (*9*); Gal_9 mediates immune regulation through Tim-3 interactions; GZMA reflects cytotoxic immune activity (*10*); and ICOSLG modulates $CD4^+$ T cell activation (*11*), with emerging roles in fibrotic disorders. Their functions align with the pathogenic mechanisms of silicosis, including macrophage activation, immune cell recruitment, and chronic inflammation.

Current silicosis surveillance methods (chest radiography, high-resolution CT (HRCT), and pulmonary function tests) primarily detect advanced disease, highlighting the critical need for early biomarkers. Our multi-omics approach integrated GWAS using the ASA chip with Olink-targeted proteomics, enabling pQTL analysis to identify genetic regulators of protein expression. This strategy identified five protein bio-CPRS markers incorporated into a composite risk prediction score for early diagnosis before radiographic changes occurred. This model offers substantial clinical utility for early intervention and personalized management, particularly in patients with comorbidities such as tuberculosis or lung cancer (*12*).

Notable strengths of this study include our novel multi-omics framework combining GWAS and CPRS proteomics, a systematic analytical pipeline (pQTL, MR), and population-specific genetic profiling. However, certain limitations exist: although patients and controls were matched for silica exposure duration, residual confounding by factors such as age or smoking cannot be ruled out. Additionally, although MR suggested causal relationships between proteins and silicosis, sample size constraints necessitated validation in larger cohorts.

In conclusion, we identified two functionally significant genetic variants and five protein biomarkers that are mechanistically linked to silicosis pathogenesis. The CPRS is a robust tool for early detection in high-risk populations. By integrating genetic and proteomic markers, this study deepens our understanding of silicosis mechanisms and offers a foundation for improved public health strategies, targeted therapies, and personalized prevention approaches for this irreversible occupational disease.

Conflicts of interest: No conflicts of interest.

Acknowledgments: The medical institutions and staff who participated in this survey.

Ethical statement: Approved by the Ethics Committee of Jiangsu Provincial Center for Disease Prevention and Control (Approval No. JSJK2022-

B002-01).

Funding: Supported by the Jiangsu Provincial Social Development Program of the Key R&D Project (BE2022803), Natural Science Foundation of Jiangsu (BK20201485), Jiangsu Provincial Key Medical Discipline (ZDXK202249), and Scientific Research Project of Jiangsu Health Commission (M2022085).

doi: 10.46234/ccdcw2025.270

[#] Corresponding authors: Zhengdong Zhang, zdzhang@njmu.edu.cn; Baoli Zhu, zhubl@jscdc.cn.

[1] Department of Genetic Toxicology, The Key Laboratory of Modern Toxicology of Ministry of Education, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing City, Jiangsu Province, China; [2] Jiangsu Provincial Center for Disease Control and Prevention, Nanjing City, Jiangsu Province, China.

Submitted: August 06, 2025
Accepted: November 24, 2025
Issued: December 19, 2025

REFERENCES

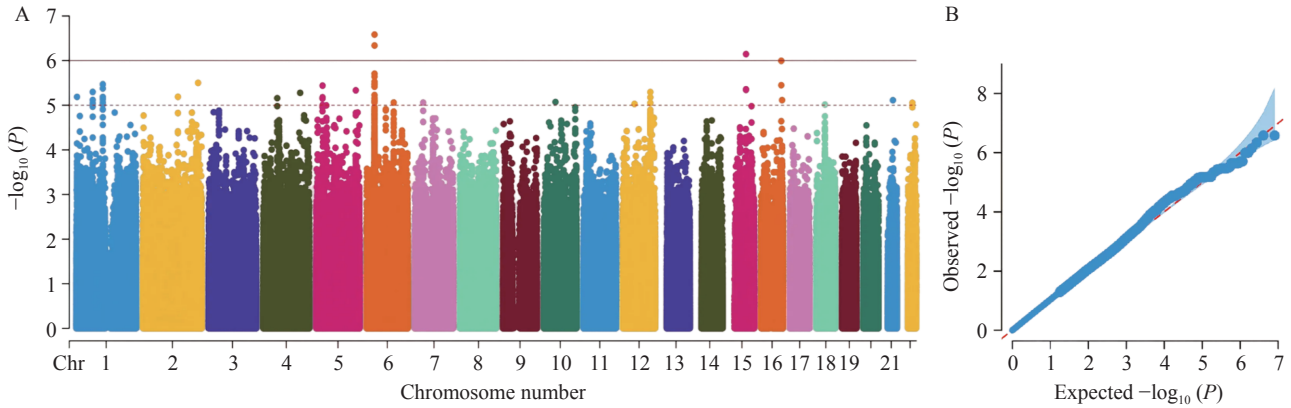
1. Li T, Yang XY, Xu H, Liu HL. Early identification, accurate diagnosis, and treatment of silicosis. Can Respir J 2022;2022(1):3769134. https://doi.org/10.1155/2022/3769134.
2. Chu MJ, Ji XM, Chen WH, Zhang RY, Sun CQ, Wang T, et al. A genome-wide association study identifies susceptibility loci of silica-related pneumoconiosis in Han Chinese. Hum Mol Genet 2014;23(23):6385 – 94. https://doi.org/10.1093/hmg/ddu333.
3. Cao ZJ, Song MY, Liu Y, Pang JL, Li ZG, Qi XM, et al. A novel pathophysiological classification of silicosis models provides some new insights into the progression of the disease. Ecotoxicol Environ Saf 2020;202:110834. https://doi.org/10.1016/j.ecoenv.2020.110834.
4. Wu YT, Xu HW, Mao LP, Zhao RR, Chu JF, Huang LL, et al. Identification of novel pQTL-SNPs associated with lung adenocarcinoma risk: a multi-stage study. Cancer Med 2024;13(17):e70247. https://doi.org/10.1002/cam4.70247.
5. Cao ZL, Xiao QL, Dai XN, Zhou ZW, Jiang R, Cheng YS, et al. circHIPK2-mediated σ-1R promotes endoplasmic reticulum stress in human pulmonary fibroblasts exposed to silica. Cell Death Dis 2017;8(12):3212. https://doi.org/10.1038/s41419-017-0017-4.
6. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A *ceRNA* hypothesis: the Rosetta Stone of a hidden RNA language? Cell 2011;146(3):353-8. http://dx.doi.org/10.1016/j.cell.2011.07.014.
7. Zhang XL, Hong RY, Chen WQ, Xu MW, Wang LF. The role of long noncoding RNA in major human disease. Bioorg Chem 2019;92:103214. https://doi.org/10.1016/j.bioorg.2019.103214.
8. Chaudhuri R, McSharry C, Brady J, Donnelly I, Grierson C, McGuinness S, et al. Sputum matrix metalloproteinase-12 in patients with chronic obstructive pulmonary disease and asthma: relationship to disease severity. J Allergy Clin Immunol 2012;129(3):655 – 63.e8. https://doi.org/10.1016/j.jaci.2011.12.996.
9. Miao RM, Ding BM, Zhang YY, Xia Q, Li Y, Zhu BL. Proteomic profiling change during the early development of silicosis disease. J Thorac Dis 2016;8(3):329 – 41. https://doi.org/10.21037/jtd.2016.02.46.
10. Ahamed MT, Forshed J, Levitsky A, Lehtiö J, Bajalan A, Pernemalm M, et al. Multiplex plasma protein assays as a diagnostic tool for lung cancer. Cancer Sci 2024;115(10):3439 – 54. https://doi.org/10.1111/cas.16300.
11. Wang SD, Zhu GF, Chapoval AI, Dong HD, Tamada K, Ni J, Chen LP. Costimulation of T cells by B7-H2, a B7-like molecule that binds ICOS. Blood 2000;96(8):2808 – 13. https://doi.org/10.1182/blood.V96.8.2808.
12. Zhang TT, Wang YY, Sun YL, Song MY, Pang JL, Wang MY, et al. Proteome, lysine acetylome, and succinylome identify posttranslational modification of STAT1 as a novel drug target in silicosis. Mol Cell Proteomics 2024;23(6):100770. https://doi.org/10.1016/j.mcpro.2024.100770.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE S1. Baseline characteristics of participants in the case–control study of plasma proteins for silicosis.

Characteristic	Case (*n*=163)	Control (*n*=189)	*P*
Sex (*n*, %)			0.312
Female	2 (1.2)	0 (0)	
Male	161 (98.8)	189 (100.0)	
Age [mean (SD)]	72.0 [69.0, 75.0]	67.0 [64.0, 71.0]	<0.001
Smoking status (*n*, %)			0.707
Never	43 (26.4)	54 (28.6)	
Former	63 (38.7)	65 (34.4)	
Current	57 (35.0)	70 (37.0)	
Pack-year [median (IQR)]	15.0 (0.0, 25.5)	10.0 (0.0, 25.0)	0.126
Drinking status (*n*, %)			0.651
Never	40 (24.5)	39 (20.6)	
Former	47 (28.8)	60 (31.7)	
Current	76 (46.6)	90 (47.6)	
Cumulative dust exposure (mg/m^3·y)	1,688.0 (1,506.1, 1,857.2)	1,648.1 (1,253.4, 1,788.1)	0.014
Use of personal protective equipment (PPE)			0.597
No	15 (9.2)	23 (12.2)	
Yes	147 (90.2)	164 (86.8)	

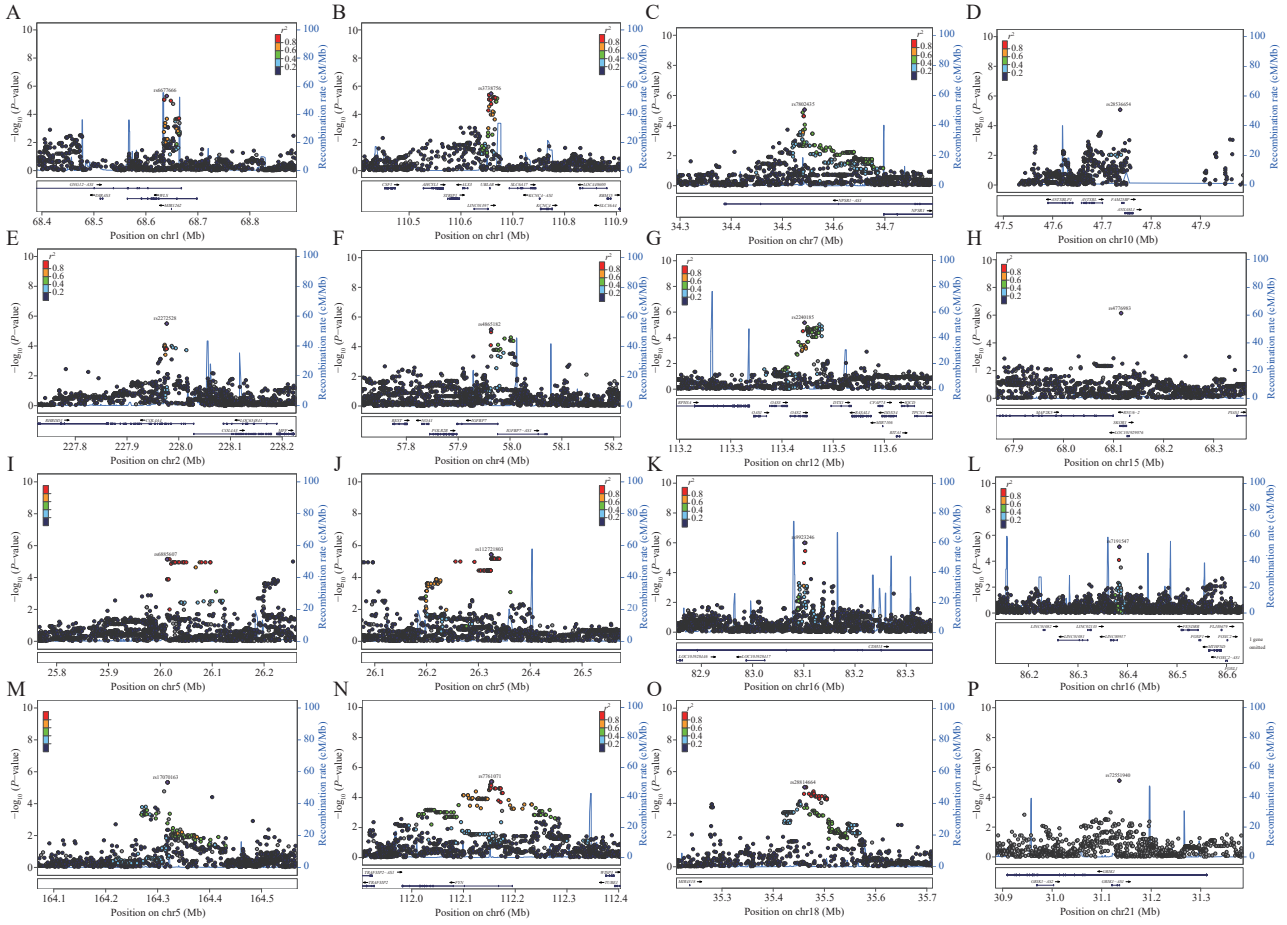
Abbreviation: SD=standard deviation; IQR=interquartile range; PPE=personal protective equipment.



SUPPLEMENTARY FIGURE S1. Results of GWAS for the silicosis. (A) Manhattan plot for the silicosis GWAS; (B) Quantile–quantile plot for the silicosis GWAS.

Note: In the Manhattan plot, different colors represent different chromosomes for easier distinction. Each point represents an SNP, i.e., a genetic variant site. A higher point position indicates a stronger statistical association between the locus and the risk of developing silicosis. (A) used to visually display which chromosomal regions contain genetic loci significantly associated with silicosis. A smaller P corresponds to a larger -$\log_{10}$(P), indicating a stronger association between the locus and silicosis. (B) used to assess the quality of the GWAS analysis and to check for systematic bias or true genetic effects. The red dashed line represents the expected distribution line. If no true associations exist, the observed values should largely coincide with the expected values. The blue line represents the observed distribution line. If the blue line deviates significantly from the red dashed line in the region of low P (upper right corner), it indicates the presence of true genetic signals (i.e., SNPs associated with the disease). If the blue line largely overlaps with the red dashed line overall, it suggests that the analysis may not have found significant associations. If the blue line shows excessive deviation in the low P region, it might suggest the presence of confounding factors such as population stratification.

Abbreviation: GWAS=genome-wide association study; SNP=single-nucleotide polymorphism.



SUPPLEMENTARY FIGURE S2. Regional plots showing the chromosomal positions of 16 independent risk loci identified in the silicosis GWAS. (A) Position on chr1 (Mb); (B) Position on chr1 (Mb); (C) Position on chr7 (Mb); (D) Position on chr10 (Mb); (E) Position on chr2 (Mb); (F) Position on chr4 (Mb); (G) Position on chr12 (Mb); (H) Position on chr15 (Mb); (I) Position on chr5 (Mb); (J) Position on chr5 (Mb); (K) Position on chr16 (Mb); (L) Position on chr16 (Mb); (M) Position on chr5 (Mb); (N) Position on chr6 (Mb); (O) Position on chr18 (Mb); (P) Position on chr21 (Mb).
Abbreviation: GWAS=genome-wide association study.