Interstitial lung diseases (ILD) comprise a heterogeneous group of lung disease characterized by common clinical syndromes and patterns of lung injury which pose growing burden on health and social economic consequences. Its etiology remains elusive. By integrating transcriptome-wide association studies analysis of ILD and chemical-gene interaction networks implemented by CGSEA software, we systematically evaluated the association between ILD and 11,190 chemicals in this study. We detected several chemicals significantly associated with ILD (permutated empirical P values < 0.05). Briefly, a total of 56 chemicals were detected for ILD in lung tissue, and 121 in whole blood respectively. Among the chemicals identified for ILD in lung tissue and whole blood, we found 7 common chemicals, including St. Thomas' Hospital cardioplegic solution, cytarabine, ginsenoside Rg3, cholecalciferol, fluoxetine, oxidized-L-alpha-1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphorylcholine and excitatory amino acid agonists. Our findings shed light on the underlying impact of chemical exposure on the development and progression of ILD, which will pave the way for more effective prevention and treatment strategies, ultimately improving the health outcomes and quality of life of those affected by ILD.

# Contributed equally to this work

**ARTICLE INFO**

**Article history:**

Received: June 12, 2023  
Accepted: September 12, 2023  
Published: November 15, 2023

**Keywords:**

Interstitial lung diseases, chemicals, transcriptome-wide association studies, gene-set enrichment analysis

**Doi:** http://dx.doi.org/10.14715/cmb/2023.69.11.25

Copyright: © 2023 by the C.M.B. Association. All rights reserved.

**Introduction**

Interstitial lung diseases (ILD), also termed diffuse parenchymal lung diseases, represent a diverse group of lung diseases that share common patterns of lung injury and clinical syndromes (1, 2). The majority of ILD is characterized by inflammation or fibrosis of the interstitial space, which can result in impaired gas exchange and subsequent symptoms such as dyspnea, decreased exercise tolerance, and diminished life quality (3). Interstitial fibrosis is the predominant presentation in most cases (4). According to a recent review, the estimated incidence of progressive fibrosing ILD varies from 2.1 (Europe) to 32.6 (US)/100,000 per person-years, with an estimated prevalence ranging from 6.9 (Europe) to 70.3 (US)/100,000 persons (5, 6). Idiopathic pulmonary fibrosis (IPF), which is the most prevalent form of fibrosing ILD, is associated with poor prognosis as evidenced by rising rates of hospital admission and mortality cases (7, 8), which poses a growing burden on the health and social economic consequences.

The current hypothesis for the pathogenesis of ILD posits that the disease may result from inflammation, fibrosis, or a combination of both, with the initial inflammatory response potentially progressing to fibrosis. Inflammation in ILD may be various underlying causes, among which autoimmune diseases such as systemic sclerosis and rheumatoid arthritis are the most prevalent (9, 10). In individuals with IPF, the onset of fibrosis appears to be influenced by three distinct factors: cumulative exposure to harmful agents resulting in excessive epithelial damage over the lifespan, aging and genetic susceptibility (11). The combined impact of these events is believed to culminate in the premature senescence of alveolar epithelial stem cells, ultimately leading to an aberrant wound-healing response following any subsequent epithelial injury (12).

The prevailing consensus until now suggests that both environmental and genetic factors all contribute to the onset and progression of ILD. Genetic analyses conducted on individuals with IPF and replicated on those with other fibrotic ILD have uncovered numerous single nucleotide polymorphisms (SNPs) linked to progressive fibrosis (13, 14). Additionally, a mounting body of evidence suggests that various ILDs are influenced by both common and rare genetic variants, affecting disease development and clinical presentation. One example is the association between mutations in TERC, TERT, RETL1 and PARN, which participate in telomere length maintenance, and an elevated risk of IPF (15). Several non-genetic factors have also been reported for IPF, including tobacco smoking, male gender and older age (16). Observational data have implicated gastroesophageal reflux, obstructive sleep apnea, air pollution, herpesvirus infection, and certain occupational exposures in ILD (17-20). However, the etiology of ILD remains elusive.

Recently, the increasing understanding of how environmental exposures impact human health, as uncovered by the Comparative Toxicogenomics Database (CTD), along with the development of the post-GWAS methodology,
such as transcriptome-wide association studies (TWAS) has advanced the field. A growing number of environmental risk factors have been identified with human complex diseases. For example, Cheng et al. have developed a flexible tool, CGSEA, for detecting the relationships between chemicals and complex diseases (21). By using the CGSEA, Lu et al. have identified a series of celiac disease-related chemicals (22). Although several risk environment factors of ILD have been identified by previous studies, many unknown environmental factors that induce inflammation or interstitial fibrosis involving ILD need to be further elucidated.

In the current study, by integrating TWAS analysis of ILD and chemical-gene interaction networks implemented by CGSEA software, we aimed to systematically evaluate the association between ILD and 11,190 chemicals.

**Materials and Methods**

**GWAS datasets of ILD**

The GWAS summary statistics of ILD (UK Biobank field: 41202, 41204) was downloaded from the UK Biobank (http://geneatlas.roslin.ed.ac.uk/downloads). Briefly, the dataset includes 1,136 ILD cases diagnosed by ICD10 (13 with alveolar and parietoalveolar conditions (field code: J840), 891 with other interstitial pulmonary diseases with fibrosis (field code: J841), 50 with other specified interstitial pulmonary diseases (field code: J848) and 424 with interstitial pulmonary disease, unspecified (field code: J849)) and 451,128 controls. The dataset included a total of 9,113,133 imputed variants that were subjected to quality control. The methods, data processing, and quality control approaches were described in detail in previous publications.

**TWAS analysis**

Functional Summary-based Imputation (FUSION) is a commonly employed tool for TWAS analysis to construct predictive models and assess transcriptome-wide associations (23). In this study, FUSION software was utilized to conduct TWAS for ILD (23). First, gene expression weights were computed for different tissues by various prediction models implemented in FUSION. Subsequently, these pre-computed gene expression weights were combined with GWAS summary statistics to calculate the association statistics between specific gene expression and ILD. Briefly, for each gene, SNP-expression weights within 1-Mb cis loci of the gene were first calculated with different statistics methods, including least absolute shrinkage and selection operator, best linear unbiased predictor, elastic net, top SNPs and Bayesian sparse linear mixed model (24). The imputed gene expression data can be viewed as a linear model of genotypes with weights based on the correlation between SNPs and gene expression in the training data while accounting for linkage disequilibrium (LD) among SNPs (23). Specific to this study, the expression weights of lung tissue and whole blood RNA array were downloaded (http://gusevlab.org/projects/fusion/), and used as reference data in the TWAS of ILD. The gene expression weight reference data of lung tissue were obtained from the GTEx v8 multi-tissue expression dataset which has been built in FUSION software (23). The expression weight reference data of whole blood was collected from the 1,264 subjects of the Young Finns Study (YFS) (25, 26).

**Chemical-gene expression annotation dataset**

We used the chemical-gene interaction dataset implemented in CGSEA software (https://github.com/ChengS-QXJTU/CGSEA) (21) in this study. Briefly, the associations between chemical exposure and gene expression alterations were retrieved from the CTD (http://ctdbase.org/downloads/), including polycyclic compounds, organic chemicals, enzymes and coenzymes biological factors. The CTD was established to curate toxicologically important genes based on the published high-throughput experimental data (27). The chemical-gene interaction networks in CTD used in our study contains 1,788,149 annotation terms of chemical-gene pairs from human and mice. Then the authors were able to generate 11,190 chemicals-related gene sets (21). An overview of the information curation process of CTD can be found in the previous study (28).

**Chemical-related gene set enrichment analysis**

The CTD chemical-gene interaction networks implemented in CGSEA software and TWAS expression association testing statistics (TWAS Z-score) of ILD were used in this study to conduct extended gene set enrichment analysis (GSEA) for exploring the associations between chemicals and ILD. A weighted Kolmogorov-Smirnov-like running-sum statistic was employed for this purpose (29). To determine statistical significance, 5,000 permutations were conducted to obtain the empirical distributions of GSEA statistics of each chemical. The P value of each chemical was then calculated from the permuted empirical distribution of GSEA statistics. Chemical gene sets containing fewer than 10 genes or more than 200 genes were excluded in this study to control the impact of different gene set sizes on the results, based on previous research (30). An empirical P value of less than 0.05 was considered significant.

**Statistical analysis**

The bioinformatics analyses conducted in this study utilized R software (v.4.0.2). The Student’s t-test was employed for continuous variables, and subsequent adjustment of P values was performed based on Benjamini and Hochberg's test. Statistical significance was determined if the p values < 0.05.

**Results**

**TWAS analysis of ILD in lung tissue**

A total of 239 significant genes in lung tissue and 135 in whole blood with empirical permutation p-value < 0.05 were detected for ILD, such as RNASEK (P = 0.0002), WNT3 (P = 0.0004), BIN1 (P = 0.0006) and VPS53 (P = 0.0008). The full significant results of TWAS were presented in Figure 1.

**ILD-associated chemicals identified by CGSEA in lung tissue**

We detected several chemicals significantly associated with ILD (permuted empirical P values < 0.05). Briefly, a total of 56 chemicals were detected for ILD in lung tissue, such as boron compounds (P = 0.0014), troxerutin (P = 0.0014), aripiprazole (P = 0.0018) and chloroacetic acid (P = 0.0018) (Figure 2). Table 1 summarizes the top 10
ILD in whole blood.

ILD-associated chemicals identified by CGSEA both in lung tissue and whole blood

Among the chemicals identified for ILD in lung tissue and whole blood, we found 7 common chemicals, including St. Thomas' Hospital cardioplegic solution, cytarabine, ginsenoside Rg3, cholecalciferol, fluoxetine, oxidized-L-alpha-1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphorylcholine and excitatory amino acid agonists (Table 2).

Discussion

Fibrotic ILD commonly presents as IPF and is associated with a dismal prognosis, with a median untreated sur-
During cellular stress and damage (43). Studies in adenostin, a purine-signaling nucleoside that will be produced in excess as a prototype therapeutic molecule for ILD (42). Adenosine is a potential cause of ILD, a possible signal for ILD and that it could exert its effects through bronchoconstriction, recruiting inflammatory cells, and elevating levels of inflammatory cytokines and mechanical insults (40, 41). We identified a series of chemicals showing correlation evidence with ILD with tissue-specific, such as fluoxetine, vitamin D3, Cholecalciferol, and NAcetyl-L-cysteine, which have been reported by previous study. Among those identified chemicals, some may serve as adverse drug reactions with the risk of ILD while some may have the potential for treatment of ILD. Fluoxetine, a selective serotonin reuptake inhibitor, possesses anti-obssessive-compulsive, anti-bulimic and anti-depressant properties. Previous studies have reported its associations with various forms of ILD such as eosinophilic pneumonia, ILD with a granulomatosis component and pneumonitis (34-37). A recent systematic review showed that fluoxetine might be a possible cause of ILD and that it could exert its effects by inducing bronchoconstriction, recruiting inflammatory cells in the pulmonary parenchyma, and elevating levels of the inflammatory marker exhaled nitric oxide (38). These effects could contribute to the development of chronic interstitial pneumonia in humans, highlighting the importance of further research on the pathogenetic and risk factors of ILD. The vinyl chloride detected in the current study may be a risk chemical for ILD by affecting proteins and the immunologic mechanisms triggered by the altered protein for the development in some cases of IPF (39). The oxidized 1-palmitoyl-2-arachidonyl-sn-glycero-3-phosphorylcholine (OxPAPC) is a potentially protective chemical that can enhance the basal barrier properties of lung endothelial cells and provide protection against vascular permeability induced by various agonists, including cell wall components, bacterial pathogens, thrombin, endotoxins, inflammatory cytokines and mechanical insults (40, 41). Considering the pathogenesis of ILD, OxPAPC may be a prototype therapeutic molecule for ILD (42). Adenosine is a purine-signaling nucleoside that will be produced in excess during cellular stress and damage (43). Studies in adenosine deaminase deficient mice have indicated that chronic adenosine elevations are linked to pulmonary fibrosis, suggesting that adenosine serves as a profibrotic signal in the lung (44).

Homocysteine, cholecalciferol and niclosamide are interesting chemicals identified in whole blood for ILD in this study. Homocysteine, a sulfhydryl-containing amino acid obtained from dietary methionine, has been found to be associated with autoimmune diseases including systemic lupus erythematosus (45) and systemic sclerosis (46, 47). Sekiguchi et al. proposed a potential association between elevated plasma homocysteine levels and ILD in dermatomyositis patients (48). Cholecalciferol, also known as Vitamin D3, is a steroid hormone that exhibits anti-inflammatory and cytoprotective properties. According to a great number of previous studies, Vitamin D3 may have prognostic and therapeutic potential in IPF patients by regulating the MAPK pathway through its impact on PSAT1 expression, both in vivo and in vitro (49-52). Niclosamide is an FDA-approved anti-helminthic drug that has exhibited pleiotropic pharmacological activities recently (53). It has been reported that Niclosamide could alleviate pulmonary fibrosis by attenuating matrix proteins, epithelial-to-mesenchymal transition, PI3K-mTORC1 pathway and Wnt/β-catenin signaling both in vitro and in vivo (54, 55).

The extended classical algorithm of GSEA, the CGSEA tool, was applied to evaluate the association between chemicals and ILD in this study with some advantages. Firstly, this approach, which examines the functional relationships between chemicals and ILD from a genetic standpoint, is expected to yield more reliable results compared to traditional methods that face challenges in accurately measuring in vivo exposure levels. Secondly, in addition to the previously reported chemicals identified in the current study, we also detected a series of novel chemicals that showed significant association signals with ILD. Further experiments are warranted to explore the potential biological mechanism of those chemicals in the pathogenesis of ILD. However, some limitations of this study should be noted. Firstly, the chemical-gene interactions were curated from a relevant literature set pertaining to the effects of chemicals on gene expression. Independent clinical samples or cohorts are required to validate our findings. Furthermore, it would be advantageous to elucidate the directionality of the relationship between chemicals and ILD.

In summary, we conducted a systematic analysis of the potential relationships between environmental chemicals and ILD. Our results revealed a set of candidate chemicals that exhibited significant association signals with ILD. It is our expectation that these findings will contribute new insights into the investigation of the etiology of ILD.

<table>
<thead>
<tr>
<th>Chemicals</th>
<th>Empirical P value</th>
<th>Lung tissue</th>
<th>Whole blood</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. Thomas’ Hospital cardioplegic solution</td>
<td>0.0024</td>
<td>0.0160</td>
<td></td>
</tr>
<tr>
<td>Cytarabine</td>
<td>0.0074</td>
<td>0.0062</td>
<td></td>
</tr>
<tr>
<td>ginsenoside Rg3</td>
<td>0.0112</td>
<td>0.0066</td>
<td></td>
</tr>
<tr>
<td>Cholecalciferol</td>
<td>0.0362</td>
<td>0.0020</td>
<td></td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>0.0366</td>
<td>0.0030</td>
<td></td>
</tr>
<tr>
<td>oxidized-L-alpha-1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphorylcholine</td>
<td>0.0392</td>
<td>0.0018</td>
<td></td>
</tr>
<tr>
<td>Excitatory Amino Acid Agonists</td>
<td>0.0486</td>
<td>0.0120</td>
<td></td>
</tr>
</tbody>
</table>

Note: ILD, interstitial lung disease.
serve as a fundamental resource for comprehending the impact of chemicals on the development of this disease.

Conflict of interests
The authors declared no conflict of interest.

References


