Toxicology Letters xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Toxicology Letters

journal homepage: www.elsevier.com/locate/toxlet

Pathway analysis for a genome-wide association study of pneumoconiosis

Ting Wang [a,1], Jingjin Yang [a,1], Xiaoming Ji [a], Minjie Chu [b,c], Ruyang Zhang [b], Juncheng Dai [b], Guangfu Jin [b], Zhibin Hu [b], Hongbing Shen [b], Chunhui Ni [a,*]

[a] Department of Occupational Medicine and Environmental Health, School of Public Health, Nanjing Medical University, Nanjing, China
[b] Department of Epidemiology and Biostatistics and Ministry of Education (MOE) Key Lab for Modern Toxicology, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, China
[c] Department of Epidemiology and Biostatistics, School of Public Health, Nantong University, Nantong, Jiangsu, China

HIGHLIGHTS

- The identify candidate causal SNPs and pathways (ICSNPathway) method was used to perform pathway analysis of a GWAS dataset related to pneumoconiosis.
- After quality control filtering, the GWAS dataset harbored genotypes of 710,999 SNPs in 202 pneumoconiosis cases and 198 exposed controls.
- Identified were 18 candidate SNPs, 13 genes, 30 pathways, and 13 biological mechanisms that might increase the risk of developing pneumoconiosis.

ARTICLE INFO

Article history:
Received 6 August 2014
Received in revised form 24 October 2014
Accepted 26 October 2014
Available online xxx

Keywords:
Pneumoconiosis
Genome-wide association study
Pathway-based analysis

ABSTRACT

Objective: The aim of this investigation was to identify pathways involved in pneumoconiosis susceptibility, clarify their potential mechanisms, and generate SNP-to-gene to pathway hypotheses using an analytical pathway-based approach.

Methods: The Identify candidate causal SNPs and pathways (ICSNPathway) was used to perform pathway analysis of a GWAS dataset for pneumoconiosis, which, after quality control filtering, harbored genotypes of 710,999 SNPs in 202 pneumoconiosis cases and 198 exposed controls. The first stage involved the pre-selection of candidate SNPs by linkage disequilibrium analysis and functional annotation of the most significant SNPs; the second stage involved annotation of biological mechanisms for the selected candidate SNPs using improved-gene set enrichment analysis.

Results: ICSNPathway analysis identified 18 candidate SNPs, involving 13 genes and 30 candidate pathways and revealed 13 hypothetical biological mechanisms. The strongest hypothetical biological mechanism was that rs8120 and rs2292151 alters the role of *TICAM1*, a gene involved in various pathways and processes, including positive regulation of tumor necrosis factor (TNF) production, innate immune response-activating signal transduction, positive regulation of the innate immune response, and the biosynthesis of type I interferon ($0.001 < p < 0.008$; 0.001< false discovery rate (FDR) <0.035). The second strongest mechanism was that rs2230656 modulates *HIST3H3* to affect its role in chromatin assembly processes ($p < 0.001$; FDR <0.001). The third mechanism was that rs11592462 modulates *CDH23*, which regulates organization of the inner ear stereocilia, auditory receptor cell morphogenesis, ear morphogenesis, and cellular homeostasis ($0.001 < p < 0.006$; $0.001 < \text{FDR} < 0.044$). Of 13 candidate genes, *TICAM1*, *HIST3H3*, *CA1*, *CA3*, *PTPRZ1*, and *IL27RA* are associated with fibrosis. Some of the 30 candidate pathways, which include positive regulation of TNF production, innate immune response-activating signal transduction, and regulation of innate immune response, may be associated with susceptibility to pneumoconiosis. Other candidate genes and pathways were novel or lacking fibrosis-related research.

Conclusion: By applying ICSNPathway analysis to the pneumoconiosis GWAS data, we identified candidate SNPs, genes such as *TICAM1* and HIST3H3, and pathways involved in the positive regulation of

* Corresponding author at: Department of Occupational Medicine and Environmental Health, School of Public Health, Nanjing Medical University, 818 East Tianyuan Road, Nanjing 211166, China. Tel.: +86 25 86868419; fax: +86 25 86868499.
E-mail address: chni@njmu.edu.cn (C. Ni).
[1] These authors contributed equally to this work

http://dx.doi.org/10.1016/j.toxlet.2014.10.028
0378-4274/

TNF production that may contribute to pneumoconiosis susceptibility. Further analyses are needed to validate the results.

1. Introduction

Pneumoconiosis is a systemic disease caused by long-term inhalation of dust and characterized by irreversible lung tissue fibrosis. Pathological varieties of silicosis, the most common type of pneumoconiosis in China, include simple (nodular) silicosis, progressive massive fibrosis, silicoproteinosis, and diffuse interstitial fibrosis (Mossman and Churg, 1998). Gross pathological examination of affected lungs reveals discrete hard nodules, usually with upper-lobe predominance. Hilar and peribronchial lymph nodes are frequently enlarged (Leung et al., 2012). Patients with pneumoconiosis often have abnormal lung function, as reflected by FEV1/FVC ratios.

As determined by morbidity and mortality, pneumoconiosis is the most prevalent occupational disease in China. In 2013, 26,393 cases of occupational diseases were reported. Among these, 23,152 cases, 87.72% of the total, were attributed to pneumoconiosis (Ministry of Health of the People's Republic of China, 2013). The disease is often associated with underground mining, where workers may be exposed to silica and coal dust. As a consequence of long-term dust exposure, there is increased mortality due to respiratory diseases, lung cancer, and cardiovascular disease (Chen et al., 2012). Although environmental factors are prominent in the pathogenesis of pneumoconiosis, a genetic component of susceptibility has been established. Previous studies using candidate gene approaches have discovered, within genes, single-nucleotide polymorphisms (SNPs) that are associated with susceptibility to pneumoconiosis. These genes include *MUC5B* (Ji et al., 2014), *SELE* (Wang et al., 2013), and *TNF-α* and *IL-1RA* (Wang et al., 2012). However, small sample sizes and limited numbers of SNPs reduced the power of such studies.

Genome-wide association studies (GWASs) provide a powerful approach to investigate variants associated with various diseases and traits. GWASs have led to the discovery and validation of relevant genes (Johnson and O'Donnell, 2009). Our previous study of Han Chinese systematically evaluated genetic variants that were associated with pneumoconiosis susceptibility on a genome-wide scale and found an association for rs73329476 ($p < 5.0 \times 10^{-8}$) and replicated associations for rs4320486 ($p < 0.05$) and rs117626015 ($p < 0.05$) that were associated with pneumoconiosis (Chu et al., 2014). However, these loci account for only a small fraction of pneumoconiosis hereditability; other variants with weak effects may be lost due to the stringent significance level and multiple comparison corrections. Further, although genetic variants were examined at the single marker level in our pneumoconiosis GWAS, the biological mechanisms of the identified SNPs remain controversial. A challenge in interpreting GWAS data is identifying causative SNPs and providing evidence for the mechanisms that could be responsible for the observed diseases and traits (Lee and Song, 2012; Wang et al., 2010).

To overcome the limitations of these approaches, pathway-based approaches have been applied to GWAS datasets (Ramanan et al., 2012). Such analyses are based on the hypothesis that SNPs causing alternations in the expression of genes involved in functional pathways lead to adverse outcomes (Ramanan et al., 2012). The identify candidate causal SNPs and pathways (ICSNPathway) method was developed to identify candidate SNPs and their corresponding candidate pathways by use of GWAS data together with integrating linkage disequilibrium (LD) analysis, functional SNP annotation, and pathway-based analysis (PBA) (Zhang et al., 2011).

To date, no PBA of the pneumoconiosis GWAS has been reported. Here, we applied ICSNPathway analysis to our previous pneumoconiosis GWAS dataset to identify candidate causal SNPs and mechanisms for pneumoconiosis susceptibility and thereby to generate SNP-to-gene-to-pathway hypotheses.

2. Materials and methods

2.1. Study population and GWAS data quality control

We used our previous pneumoconiosis GWAS dataset, which included data for 205 pneumoconiosis cases and 203 exposed controls. All subjects provided written informed consent, and the institutional review boards of each participating institution approved this collaborative study (Chu et al., 2014). The dataset was filtered to remove unqualified samples and SNPs. One sample with overall genotype completion rates <95% was excluded. Four probable relatives were excluded based on pairwise identity by state according to their PI_HAT values in PLINK (all PI_HAT >0.125). In addition, three other samples were removed as population outliers. Retained for further analysis were 202 cases and 198 exposed controls.

900,015 SNPs were filtered by use of these criteria: (1) SNPs were not mapped on autosomal chromosomes; (2) SNPs had a call rate <95%; (3) SNPs had a minor allele frequency <0.05 in all samples; (4) the genotype distributions of SNPs deviated from those expected by the Hardy–Weinberg equilibrium in all samples ($p < 1 \times 10^{-5}$). These thresholds resulted in 710,999 SNPs remaining for further pathway analysis.

2.2. Identification of candidate causal SNPs and pathways

The ICSNPathway analysis was accomplished in two steps (Zhang et al., 2011). The first phase included the pre-selection of candidate causal SNPs using linkage disequilibrium (LD) analysis, in which the threshold value of r^2 was 0.8, and the functional SNP annotation was based on the most significant SNPs. The second phase involved annotation of the biological mechanisms to preselect candidate causal SNPs by the PBA algorithm i-GSEA (improved gene-set enrichment analysis). A complete list of SNP *p*-values from the pneumoconiosis GWAS was used in the ICSNPathway analysis.

LD analysis is used to search for SNPs in LD with most significant SNPs of a GWAS dataset and thereby to identify candidate causal SNPs based on an extended data set, including HapMap data (International HapMap 3Consortium et al., 2010). Another approach utilizes functional SNPs, which are defined as those that may alter protein or gene expression or the role of a protein in the context of a pathway. Functional SNPs include deleterious and non-deleterious, non-synonymous SNPs; those that cause the gain or loss of a stop codon those resulting in a frame shift; and those located in essential splice sites or in regulatory regions. Based on functional SNPs, ICSNPathway analysis preselects candidate causal SNPs in order to understand the underlying genetics of human health.

The ICSNPathway server applied the i-GSEA PBA algorithm to the full list of pneumoconiosis GWAS SNP *p*-values to detect the pathways associated with individual characteristics. Briefly, the

Please cite this article in press as: Wang, T., et al., Pathway analysis for a genome-wide association study of pneumoconiosis. Toxicol. Lett. (2014), http://dx.doi.org/10.1016/j.toxlet.2014.10.028