A Large Scale Gene-Centric Association Study of Lung Function in Newly-Hired Female Cotton Textile Workers with Endotoxin Exposure

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version
doi:10.1371/journal.pone.0059035

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:10611678

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
A Large Scale Gene-Centric Association Study of Lung Function in Newly-Hired Female Cotton Textile Workers with Endotoxin Exposure

Ruyang Zhang1,2,9, Yang Zhao1,3,9, Minjie Chu1,2, Amar Mehta3,4,5, Yong Yue Wei1,3, Yao Liu1,2, Pengcheng Xun6, Jianbing Bai1, Hao Yu1, Li Su3, Hongxi Zhang7, Zhibin Hu1,2, Hongbing Shen1,2, Feng Chen1,9, David C. Christiani3,8

1 Department of Epidemiology and Biostatistics and Ministry of Education (MOE) Key Lab for Modern Toxicology, School of Public Health, Nanjing Medical University, Nanjing, China, 2 Section of Clinical Epidemiology, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Cancer Center, Nanjing Medical University, Nanjing, China, 3 Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, United States of America, 4 Swiss Tropical and Public Health Institute, Basel, Switzerland, 5 University of Basel, Basel, Switzerland, 6 Department of Epidemiology and Biostatistics, School of Public Health, Indiana University Bloomington, Bloomington, Indiana, United States of America, 7 Pu t u o Di strict PeOPles Hospital, Shanghai Second Medical University, Shanghai, China, 8 Pulmonary and Critical Care Unit, Department of Medicine, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, United States of America.

Abstract

Background: Occupational exposure to endotoxin is associated with decrements in pulmonary function, but how much variation in this association is explained by genetic variants is not well understood.

Objective: We aimed to identify single nucleotide polymorphisms (SNPs) that are associated with the rate of forced expiratory volume in one second (FEV1) decline by a large scale genetic association study in newly-hired healthy young female cotton textile workers.

Methods: DNA samples were genotyped using the Illumina Human CVD BeadChip. Change rate in FEV1 was modeled as a function of each SNP genotype in linear regression model with covariate adjustment. We controlled the type 1 error in study-wide level by permutation method. The false discovery rate (FDR) and the family-wise error rate (FWER) were set to be 0.10 and 0.15 respectively.

Results: Two SNPs were found to be significant (P<6.29x10^-5), including rs1910047 (P = 3.07x10^-5, FDR = 0.0778) and rs9469089 (P = 6.19x10^-5, FDR = 0.0967), as well as other eight suggestive (P<5x10^-6) associated SNPs. Gene-gene and gene-environment interactions were also observed, such as rs1910047 and rs1049970 (P = 0.0418, FDR = 0.0895); rs9469089 and age (P = 0.0161, FDR = 0.0264). Genetic risk score analysis showed that the more risk loci the subjects carried, the larger the rate of FEV1 decline occurred (P = 3.01x10^-18). However, the association was different among age subgroups (P = 7.11x10^-6) and endotoxin subgroups (P = 1.08x10^-5). Functional network analysis illustrates potential biological connections of all interacted genes.

Conclusions: Genetic variants together with environmental factors interact to affect the rate of FEV1 decline in cotton textile workers.


Editor: Dan E. Arking, Johns Hopkins University, United States of America

Received November 29, 2012; Accepted February 8, 2013; Published March 19, 2013

Copyright: © 2013 Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by grants from USA: The National Institute for Occupational Safety and Health (RO104241), National Institute of Environmental Health Sciences (ES00002), National Natural Science Foundation of China (81072389, 30901232), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (10KJA33034), the Research Fund for the Doctoral Program of Higher Education of China (20113234110002), the Research and Innovation Project for College Graduates of Jiangsu Province (CXXZ11_0733), and the grant from the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: fengchen@njmu.edu.cn (FC); dchris@hsph.harvard.edu (DCC)
† These authors contributed equally to this work.

Introduction

According to an official statement from the American Thoracic Society (ATS) [1], cigarette smoking is not the sole meaningful cause of chronic obstructive pulmonary disease (COPD). There is also strong evidence suggesting a causal relationship between COPD and occupational exposures (e.g. endotoxin [2]) or genetic syndromes (e.g. α1-antitrypsin deficiency [3]). Generally released from bacterial lysis, endotoxin is ubiquitous in the airborne environment [4]. High endotoxin levels are observed in cotton
textile mills processing cotton. Exposure to cotton dust leads to chronic respiratory disease and excessive loss of pulmonary function [5–7]. Recently, genome-wide association studies (GWAS) [8–10] and large-scale meta-analysis of GWAS [11–13] of COPD also offered mechanistic insight into pulmonary function regulation due to individual genetic heterogeneity (e.g. 6p21.32 region) in non-occupationally exposed populations.

Even for occupationally exposed populations, there is evidence that single nucleotide polymorphisms (SNPs) may play an important role in pulmonary function while taking endotoxin effect into account. Our previous study in 20-year longitudinal cohort suggested that Tyr113His and His139Arg polymorphisms in microsomal epoxide hydrolase (mEH) gene [14], rs1800629 in tumor necrosis factor (TNF) gene and rs909253 in lymphotxin alpha (LTA) gene [15], may modify the association between endotoxin exposure and annual decline of FEV₁ in cotton textile workers, though no statistically significant marginal effects of those polymorphisms were observed.

In this study, we aimed to explore the association between the rate of FEV₁ decline and genetic heterogeneity in previously unexposed newly-hired young healthy female cotton textile workers entering endotoxin-exposed work areas in yarn preparation based on large scale genomic data. To the best of our knowledge, there is no large scale association study of lung function in occupational settings.

Materials and Methods

Ethics Statement

The institutional review boards of the Harvard School of Public Health, the Putuo District People’s Hospital, and the Human Resources Administration of China approved the study. All participants gave written informed consent before the study.

Study Population

The study population was derived from the Newly-Hired Chinese Textile Workers Study, an intensive, repeated measures prospective cohort study in the cotton textile industry consisting of 384 adults in Shanghai, China. All subjects were self-reported Han Chinese ethnicity, females, and lifelong non-smokers, asymptomatic of cardiopulmonary disease and naive to occupational endotoxin exposure before entering cotton textile mill. Detailed information on subject selection, methods for testing pulmonary function, and exposure assessment has been described previously [16–18]. A baseline examination took place in March of 1997, and follow-up examinations occurred at 3 months, 12 months, and 18 months later. Only subjects with DNA samples that met quality assurance requirements for genotyping and analysis (n = 301; refer to quality control section for details) were included in this study. Among the 301 subjects, one hundred and sixty-three (54.2%) subjects with average age of 18.4 years (range 16.0–28.8 years) were newly employed, while 138 (45.8%) subjects with average age of 33.1 years (range 16.6–46.7 years) were currently working in converted silk mills. Sixty-two (20.6%) subjects were followed up to 3 months, 42 (14.0%) subjects were followed up to 12 months, and 197 (65.4%) subjects were followed up to 18 months.

Exposure Assessment and Pulmonary Function Tests

Briefly, airborne cotton dust was sampled with a settled vertical elutriator (General Metaworks Corp., Mequon, WI) fixed in work area using 37 mm PVC filters, and endotoxin concentration was analyzed using chromogenic Limulus amoebocyte lysate assay, as previously described [19]. The workers remained in the same area throughout the observation. We recorded the endotoxin concentration for all work areas at each follow-up. Spirometric measurements were conducted by a trained technician at cotton mills using an 8-liter water-sealed field spirometer (W.E. Collins Co., Braintree, MA) calibrated twice each day with a 3-liter syringe. Forced expiratory spirometers were carried out before and after work shifts on the first day back to work after a two day rest, and each worker performed up to seven trials to produce three acceptable curves [16]. The present analysis focused on change rate in pre-shift FEV₁ over the 18 month period of follow-up.

DNA Isolation and SNP Genotyping

DNA samples were extracted from whole blood using a DNA purification kit (Gentra Systems, Inc., Research Triangle, NC). Of the 384 DNA samples collected from subjects who participated in the baseline examination, 53 samples were excluded due to inadequate DNA content or low DNA quality. Genotyping was performed using the Illumina HumanCVD Beadchip (IBC array) by laboratory personnel without the knowledge of phenotype information. The IBC array incorporates about 50,000 SNPs to efficiently capture genetic diversity across >2,000 genetic regions related to cardiovascular, inflammatory and metabolic phenotypes. Genetic variation within the majority of these regions is captured at density equal to or greater than that afforded by genome-wide genotyping products [20]. Of the 329 samples that were sent for genotyping, 312 were genotyped successfully.

Quality Control

We performed systematic quality control procedures to filter both unqualified samples and SNPs before the association analysis. Briefly, SNPs were removed for a total of 49,094 genotyped ones if they were located in non-autosomal chromosomes (1,126), had a call rate ≤95% (954), minor allele frequency (MAF) <0.05 (19,357), P<0.001 for Hardy-Weinberg equilibrium (HWE) test (46). For these 312 samples, one subject was removed with a genotyping call rate <95%. None was ambiguous for genetic sex. Seven subjects failed the cryptic relatedness check. We also detected outliers (three individuals were excluded) and population stratification using EIGENSTRAT 4.2 [21], a method based on principal component analysis. As a result, a total of 301 subjects and 27,611 SNPs were left for subsequent analysis. Quality control procedures were implemented in PLINK (version 1.07) [22].

Statistical Methods

As presented in Figure S1 in File S1, the FEV₁ declines linearly with increased exposure time, but the rate of FEV₁ decline varies among individuals. We aimed to evaluate the association between SNP and the rate of FEV₁ decline. Firstly, we estimated the rate of monthly FEV₁ decline for each individual in a linear regression model through Equation 1. For the ᵗᵗʰ individual, \( \text{FEV}_{1ij} \) is the observed FEV₁ value at different months \( \text{month}_{ij} \), \( u_i \) is baseline FEV₁ and \( e_i \) is the random error at \( \text{month}_{ij} \). The estimated slope (ES) reflected the rate of FEV₁ decline as a function of month for each individual. The mean and median \( R^2 \) of the linear model for all 301 subjects are 0.63 and 0.73 respectively indicating the present model is not bad to describe the monthly decline rate of FEV₁. ES was then used as the outcome in the following association analysis.

\[
\text{FEV}_{1ij} = u_i + \text{ES} \times \text{month}_{ij} + e_i
\]

Secondly, we assessed the associations between ES and each SNP by fitting multiple-variable linear regression models (Equa-
tion 2) adjusted for height, age, FEV1 at baseline and the average level of log-transformed endotoxin exposure. In Equation 2, $\epsilon_i$ is the random error for the $i$th individual. Two genetic models (dominant and additive) of inheritance for each SNP were considered respectively. In additive model, SNPs with low frequency (≤5%) of rare homozygous were excluded for lack of statistical robustness [23].

$$ES_i = \beta_0 + \beta_1 \times \text{SNPi} + \beta_2 \times \text{heighti} + \beta_3 \times \text{agei} + \beta_4 \times \text{FEV1i} + \beta_5 \times \text{endoi} + \epsilon_i$$ (2)

The bootstrap re-sampling analysis was then performed to control the false positive discovery for significant SNPs [24]. We generated 2,000 bootstrap samples, and SNPs were tested 2,000 times respectively. The results are not likely to be falsely positive if the SNP has more than 1,600 times (80%) bootstrap $P$ value less than the predefined threshold (e.g. 0.005 for 10 SNPs).

Gene-gene and gene-environment interactions were tested using Equation 3. Here, $\beta_A$ and $\beta_B$ are the main effects of factor $A$ and $B$, respectively. $\beta_{AB}$ is the interaction.

$$ES_i = \beta_0 + \beta_A \times \text{Ai} + \beta_B \times \text{Bi} + \beta_{AB} \times \text{Ai} \times \text{Bi} + \sum \beta_m \times \text{CoVariate} \mu_i + \epsilon_i$$ (3)

We also constructed a genetic risk score (GRS) by counting the number of risk genotypes of top 10 SNPs that the individuals carried [25]. The association between the GRS and the rate of FEV1 decline was also tested in multiple-variable linear regression models to evaluate the joint effects of multiple genetic factors. Additionally, stratification analysis of GRS by age and endotoxin level was performed.

Due to the linkage disequilibrium (LD) structure among SNPs, a standard Bonferroni correction would yield a significance level of approximately $1 \times 10^{-5}$, which is very conservative. And, the Human CVD array has a dense gene-centric design, similar studies have used a less stringent level around 1 previously [26,27]. In this exploratory study, we controlled the false discovery rate (FDR) and the family-wise error rate (FWER) in study-wide level by permutation method (see details in Appendix S1). The FDR was set to be 0.10, meanwhile the FWER were set to be 0.15 (corresponding threshold for each SNP was 6.29×10$^{-5}$). We imputed un-genotyped SNPs using Minimac software [28] based on LD information from hg19 1K Genomes database (with ASN as the reference set, released November 2010). Gene functional network was built using MetaCoreTM online software [29] (GeneGo, Inc., Carlsbad, CA) to illustrate potential biological connections of interacted genes identified in this study. Statistical analysis and data management were carried out through: PLINK (Version 1.07) [22] and R software (version 2.14.0; The R Foundation for Statistical Computing). Manhattan plots were generated using Haploview (version 4.2) [30].

Results

The mean decline of FEV1 for the overall 301 subjects was 6.79±25.09 ml per-month. Individuals were classified to different subgroups by age, height, FEV1 at baseline and endotoxin level respectively and the rate of FEV1 decline (ES) in each subgroup was presented in Table 1. The large standard deviation (SD) of ES reflects a large individual variation of ES among subjects, which probably be caused by different environmental exposure, demographic characteristics or genetic background. We fitted a multi-variable linear regression model (Equation 2) and predicted the ES for each subject. The SD of the predictive ES was presented in Table S1 in File S1. When deduced the effects of non-genetic factors, the SD of predictive ES ranges from ~4 to ~6. We assumed that the genetic factors might contribute to the remained variation of ES.

Results of Individual SNP Analysis

In additive model, the frequency of rare homozygosity of the top-10 SNPs was less than 2.5% (data not shown) and none was significant. Thus, only SNPs identified by the dominant model are presented (Table 2, Tables S2 and S3 in File S1). Two risk SNPs (rs1910047 and rs9469089) were significant with $P$<6.29×10$^{-5}$. Individuals carrying the TT genotype of rs1910047 had average 4.32 ml decline of FEV1 per-month, while those carrying the TA or AA genotypes had 18.32 ml decline of FEV1 per-month. After adjustment of covariates, individuals carrying the AA genotype had more rapid FEV1 decline (15.17 ml/moth) compared with those carrying the TA or AA genotype ($P$=3.70×10$^{-5}$, FDR = 0.0778). For rs9469089, individuals carrying the GC or CC genotype also had more rapid FEV1 decline (12.50 ml/moth) than those carrying the GG genotype ($P$=6.19×10$^{-5}$, FDR = 0.0967).

Eight SNPs (rs32586, rs4855881, rs10515978, rs601675, rs11761231, rs10129426, rs10201627 and rs1049970) with $P$<5×10$^{-4}$ were considered suggestive for accelerated changes of FEV1. Two SNPs (rs32586 and rs32589) are in perfect LD with $R^2$ = 1.00, rs3749073 is also in high LD with rs10201627 ($R^2$ = 0.98) (Table S4 in File S1). The bootstrap re-sampling analysis was then performed for the top-10 SNPs to internally validate the results. All top-10 SNPs had a bootstrap $P$ value<0.005 more than 1,600 times out of 2,000 bootstrap samples (over 80% significant), indicating that those SNPs were not likely to be false-positive results.

Manhattan plot using $P$ values derived from the dominant model are presented in Figure S2 in File S1. A small genomic control inflation factor (k) of 1.026 indicated little inflation of the

### Table 1. The mean and standard deviation of the rate of FEV1 decline (ES) in different subgroups.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Mean ± SD (ml/month) of the rate of FEV1 decline (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ≤18 years</td>
<td>18≤ Age ≤25 years</td>
</tr>
<tr>
<td>Heightb</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.65±24.53 30</td>
</tr>
<tr>
<td>High</td>
<td>-1.73±23.47 43</td>
</tr>
<tr>
<td>FEV1b</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>-0.60±13.16 31</td>
</tr>
<tr>
<td>High</td>
<td>-0.86±29.43 42</td>
</tr>
<tr>
<td>Endotoxinb</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>-8.86±29.19 15</td>
</tr>
<tr>
<td>High</td>
<td>1.29±22.00 58</td>
</tr>
</tbody>
</table>

bSubjects were divided into two groups according to median (160 cm, 2630 ml and 163 EU/m$^3$ for height, FEV1, and endotoxin respectively); doi:10.1371/journal.pone.0059035.t001

aSubjects with age ≥25 years were divided into two additional groups according median (25 years); doi:10.1371/journal.pone.0059035.s001
Results of Gene-gene and Gene-environment Interaction Analysis

We tested all pairwise gene-gene interactions for these ten SNPs, and interactions with P value < 0.05 and FDR < 0.10 were presented in Table 3. rs1910047 which had significant marginal effect had synergistic interaction with rs4855881 (P = 0.039, FDR = 0.016). The other three interactions which composed of four SNPs, also showed synergy between SNPs. The interaction P values (FDRs) were 0.0051 (0.0313), 0.0052 (0.0313) and 0.0095 (0.0390) for rs10129426_ruru 1049970 and rs1910047 respectively. We mapped these interacted SNPs to their nearest genes (Table S3 in File S1), and built network (Figure 1) using the six genes (TBX3, CDH5, PPARC1B, LMAN1, PODXL and APEH). There are four highlighted functional paths corresponding to four pairs of gene-gene interactions: TBX3 - KL4 - VE_cadherin (CDH5); PERC (PPARGC1B) - ESR1(nuclear) - ATF6 alpha-LMAN1; PERC (PPARGC1B) - ESR1(nuclear) - Provastone receptor - p300 - NF_kb - Podocalyxin_like 1 (PODXL); Podocalyxin_like 1 (PODXL) - NF_kb - c_Myc - APEH. The information of all 63 nodes in the network is given in Appendix S3. Fourteen nodes (Ubiquitin, AP_1 and other twelve blue-circle marked ones) directly connect with the four functional paths and form the gene-gene interaction functional network.

From the results of gene-environment interaction analysis (Table 4), three SNPs (rs9469089, rs1049970 and rs10201627) with FDR < 0.10 exhibited interaction with age (P = 0.016, FDR = 0.0264; P = 0.0206, FDR = 0.0356 and P = 0.0350, FDR = 0.0300 respectively; rs4855881 had interaction with endotoxin level (P = 0.0454, FDR = 0.4006).

Results of Genetic Risk Score Analysis

As presented in Figure 2 and Table S9 in File S1, the more risk loci the subjects carried, the larger the rate of FEV1 decline occurred (P trend = 3.01 × 10^-10). Bootstrap re-sampling analysis showed that bootstrap P trend was less than 1 × 10^-10 with 92.75% of 2,000 bootstrap samples. Additionally, we performed stratification analysis of GRS by age and endotoxin level (Table S10 in File S1). Though, the rate of FEV1 decline became larger with the increase of GRS, the association differed among age subgroups or endotoxin subgroups, which indicated that there were interactions between GRS and age (P = 7.11 × 10^-7) or GRS and endotoxin (P = 1.08 × 10^-5) adjusted for other covariates (Figure 3).
Four gene-gene interaction functional paths were highlighted, including TBX3 - VE_cadherin (CDH5), PERC (PPARGC1B) - LMAN1, PERC (PPARGC1B) - Podocalyxin_like 1 (PODXL) and Podocalyxin_like 1 (PODXL) - APEH.

doi:10.1371/journal.pone.0059035.g001

Table 3. The results of gene-gene interaction analysis for top-10 SNPs (P<0.05).

<table>
<thead>
<tr>
<th>Gene1a</th>
<th>Gene2a</th>
<th>N</th>
<th>Mean ± SDb</th>
<th>/ (SE)c</th>
<th>Pd</th>
<th>Padjd</th>
<th>FDR*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1910047</td>
<td>rs1049970</td>
<td>159</td>
<td>−1.39±23.13</td>
<td>Ref.</td>
<td>4.18×10−2</td>
<td>0.0895</td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>CC</td>
<td>123</td>
<td>−13.12±24.95</td>
<td>Ref.</td>
<td>5.09×10−3</td>
<td>0.0313</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>TT</td>
<td>150</td>
<td>−12.38±21.74</td>
<td>Ref.</td>
<td>9.51×10−3</td>
<td>0.0390</td>
<td></td>
</tr>
</tbody>
</table>

*rs1910047 (TBX3), rs1049970 (CDH5), rs10515978 (LMAN1), rs32588 (PPARGC1B) and rs4855881 (APEH);  
*Mean and stand deviation of the rate of FEV1 decline (ml/month).  
*Coefficient, stand error, P value were acquired from linear regression model adjusted for height, age and FEV1 at baseline, and average log transformed endotoxin level.  
*P value for gene-gene interaction.  
*FDR q-value calculated based on all pair-wise (45 pairs) P values of top-10 SNPs.
Discussion

Host genetic factors probably influence susceptibility to the rate of FEV\textsubscript{1} decline after occupational exposure to endotoxin. This exploratory study is the first to provide evidence that genetic factors may play an important role in determining the rate of FEV\textsubscript{1} decline, as well as eight suggestive associated SNPs. Further analysis indicated that potential gene-gene interaction and gene-environment interaction were ubiquitous in the genetic architecture of complex traits.

rs9469089 is located at 6p21.32, within the first intron of the TBX5 gene, which may affect lung function, accelerating FEV\textsubscript{1} decline. TBX3 is a downstream target of Wnt/\beta-catenin pathway, which is a key regulator in cell proliferation and differentiation [34]. Recent evidence has suggested that TBX3 is over-expressed in a number of cancers including lung cancer [34], as well as in transformed lung fibroblast cells [35].

rs9469089 is located at 6p21.32, within the first intron of the RNF5 gene, which may negatively regulate the virus-triggered immune response [37,38]. Besides, rs9469089 is about 2 kb downstream of advanced glycosylation end product-specific receptor (AGER), which is consistently reported to be associated with lung function [11-13].

rs1049970 which interacted with rs1910047 is a missense variant of the cadherin 5, type 2 (CDH5) gene that may play an important role in endothelial cell biology through control of the cohesion and organization of the intercellular junctions. Deficiency or truncation of Cdh5 induced endothelial apoptosis and abolished transmission of the endothelial survival signal in a mouse

### Table 4. The results of gene-environment interaction analysis for top-10 SNPs (P<0.05).

<table>
<thead>
<tr>
<th>Gene\textsuperscript{a}</th>
<th>Environment</th>
<th>N</th>
<th>Mean ± SD\textsuperscript{b}</th>
<th>(/ \text{SE})\textsuperscript{c}</th>
<th>(P)</th>
<th>(P_{\text{ca}})\textsuperscript{d}</th>
<th>FDR\textsuperscript{e}</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs9469089</td>
<td>Age (years)</td>
<td>53</td>
<td>3.16 ± 25.29</td>
<td>Ref.</td>
<td>0.0161</td>
<td>0.0264</td>
<td></td>
</tr>
<tr>
<td>rs9469089</td>
<td>&lt;18</td>
<td>53</td>
<td>3.16 ± 25.29</td>
<td>Ref.</td>
<td>0.0161</td>
<td>0.0264</td>
<td></td>
</tr>
<tr>
<td>rs9469089</td>
<td>18 ~ 25</td>
<td>80</td>
<td>0.81 ± 28.47</td>
<td>-2.55(4.38)</td>
<td>0.5607</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9469089</td>
<td>&gt;25</td>
<td>84</td>
<td>-10.71 ± 12.20</td>
<td>-23.80(10.91)</td>
<td>0.0300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9469089</td>
<td>&lt;18</td>
<td>20</td>
<td>-11.10 ± 15.35</td>
<td>-13.23(6.17)</td>
<td>0.0329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9469089</td>
<td>18 ~ 25</td>
<td>26</td>
<td>-22.69 ± 40.84</td>
<td>-25.84(5.85)</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9469089</td>
<td>&gt;25</td>
<td>37</td>
<td>-15.49 ± 18.17</td>
<td>-28.43(11.53)</td>
<td>0.0143</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1001627</td>
<td>Age (years)</td>
<td>47</td>
<td>2.03 ± 23.22</td>
<td>Ref.</td>
<td>0.0206</td>
<td>0.0264</td>
<td></td>
</tr>
<tr>
<td>rs1001627</td>
<td>&lt;18</td>
<td>47</td>
<td>2.03 ± 23.22</td>
<td>Ref.</td>
<td>0.0206</td>
<td>0.0264</td>
<td></td>
</tr>
<tr>
<td>rs1001627</td>
<td>18 ~ 25</td>
<td>78</td>
<td>0.78 ± 28.14</td>
<td>-2.63(3.55)</td>
<td>0.5644</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1001627</td>
<td>&gt;25</td>
<td>69</td>
<td>-10.67 ± 12.94</td>
<td>-22.43(11.30)</td>
<td>0.0481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10201627</td>
<td>Age (years)</td>
<td>52</td>
<td>-14.17 ± 16.02</td>
<td>-25.78(11.20)</td>
<td>0.0220</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10201627</td>
<td>&lt;18</td>
<td>52</td>
<td>-14.17 ± 16.02</td>
<td>-25.78(11.20)</td>
<td>0.0220</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10201627</td>
<td>18 ~ 25</td>
<td>37</td>
<td>-15.49 ± 18.17</td>
<td>-28.43(11.53)</td>
<td>0.0143</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10201627</td>
<td>&gt;25</td>
<td>29</td>
<td>-19.86 ± 40.90</td>
<td>-21.77(5.83)</td>
<td>0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4855881</td>
<td>Endotoxin (EU/m\textsuperscript{3})</td>
<td></td>
<td>0.0454</td>
<td></td>
<td>0.4086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4855881</td>
<td>&lt;163</td>
<td>132</td>
<td>-12.90 ± 22.51</td>
<td>Ref.</td>
<td>0.0454</td>
<td>0.4086</td>
<td></td>
</tr>
<tr>
<td>rs4855881</td>
<td>&lt;163</td>
<td>132</td>
<td>-12.90 ± 22.51</td>
<td>Ref.</td>
<td>0.0454</td>
<td>0.4086</td>
<td></td>
</tr>
<tr>
<td>rs4855881</td>
<td>163 ~ 255</td>
<td>125</td>
<td>-5.04 ± 21.40</td>
<td>1.46(5.42)</td>
<td>0.7876</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4855881</td>
<td>&lt;163</td>
<td>15</td>
<td>12.82 ± 45.55</td>
<td>27.98(6.51)</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4855881</td>
<td>&lt;163</td>
<td>15</td>
<td>12.82 ± 45.55</td>
<td>27.98(6.51)</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}rs9469089 (\textit{RNFS}), rs1049970 (\textit{CDH5}), rs1001627 (\textit{GPR55}) and rs4855881 (\textit{APEH});

\textsuperscript{b}Mean and stand deviation of the rate of FEV\textsubscript{1}, decline (ml/month);

\textsuperscript{c}Coefficient, stand error, \(P\) value were acquired from linear regression model adjusted for height, age and FEV\textsubscript{1} at baseline, and average log transformed endotoxin level.

\textsuperscript{d}P value for gene-environment interaction.

\textsuperscript{e}FDR q-value calculated based on top-10 SNPs (10 pairs for Age-SNP interactions and 10 pairs for SNP-Endotoxin interactions).

doi:10.1371/journal.pone.0059035.0004
Both two SNPs are involved in cell stability, including proliferation, differentiation and apoptosis. Thus, it is biologically conceivable that these two SNPs, or those in LD with them, may directly regulate targeted genes, together resulting in more rapid FEV1 decline.

It is noteworthy that four SNPs interacted with each other adjacently, producing three pairs of end-to-end interactions in this study (rs10515978_rs32588, rs32588_rs11761231 and rs11761231_rs4855881). rs10515978 is within the intron of lectin, mannos-binding, 1 (LMAN1), mutational inactivation of which was a frequent and early event potentially contributing to colorectal tumorigenesis [40]. rs32588 is within the exon of peroxisome proliferator-activated receptor gamma, coactivator 1 beta (PPARGC1B), which was associated with airway hyperreactivity in asthmatic patients [41]. rs11761231 was reported to be associated with rheumatoid arthritis in Caucasians [42–44]. rs4855881 is located at 3p21.31, within the intron of N-acylaminoacyl-peptide hydrolase (APEH). Deletions of which are commonly at the 3p region, and have been found in various types of carcinomas, including lung cancer [45].

Functional network analysis reveals potential biological connections among these four pairs interacted genes. Fourteen nodes as well as other ones directly or indirectly connect to the four functional paths and form the network. Among the fourteen key nodes, genetic variation in the gene encoding p53 can impair the response to cell damage and increase the loss of alveolar epithelial cells [46]. Also, it’s reported that p53 may modify the effect of diminishing PM10 exposure on lung function decline [47]. Vascular endothelial growth factor receptor (VEGFR) was reported to mediate the anti-inflammatory, anti-protease, and anti-apoptosis effects of the lung, hence contribute to an attenuation of emphysema and destructive pulmonary function in COPD [48]. Signal transducer and activator of transcription 3 (Stat3) plays an essential role in the pathogenesis of IgG immune complex (IC)-induced acute lung injury [49]. Beosearcw the morphogenetic protein 4 (BMP4) plays a potential regulatory role of lung fibroblast function [50]. Combined with our gene-gene interaction results, it is possible that the interaction among genes is associated with the rate of FEV1 decline through some possible functional paths.

rs10129426 is about 4 kb downstream of BCL2-associated athanogene 5 (BAG5), which encodes anti-apoptotic protein that functions through interactions with a variety of cell apoptosis and growth related proteins [51]. According to the UCSC and TFSearch [52] databases, polymorphisms of this gene may influence transcription factor binding (GATA-binding factor 2) [53].

rs10201627 which is located in the micro-RNA binding sites of the 3’-UTR of G protein-coupled receptor 55 (GPR55), was predicted to affect miR-26b binding. And it was found that expression of miR-26b was down-regulated in lung cancer tissues [54]. Moreover, rs3749073, as the sole non-synonymous polymorphism (Gly195Val) in the GPR55, is in high LD (r2 = 0.98) with rs10201627. It was predicted that GPR55 with Val195 appeared to induce less phosphorylation of extracellular signal-regulated kinase (ERK) than Gly195 [55]. It was also reported that GPR55 promoted cancer cell proliferation by ERK [56].

Actually, among these top ten SNPs, five SNPs (rs1910047, rs1049970, rs4855881, rs10129426 and rs10201627) are related to cell proliferation or apoptosis and three SNPs (rs9469089, rs32588 and rs11761231) are related to the immune inflammatory response.

We then again investigated the marginal effects of rs1051740 (Tyr113His), rs1051741, which was in high LD with rs2234922 (His139Arg, not genotyped in our chip), rs1800629 and rs909253 in this study according to our previous studies of long-term pulmonary function decline [14,15]. None of them reached the statistical significance level. However, rs1051740, rs1051741 and...
rs909253 could still be significant modifiers in the association between endotoxin exposure and the rate of FEV₁ decline. The effect of endotoxin exposure was different among subjects carrying different genotypes of SNP. The \(P\) values of heterogeneity for these three SNPs were 4.83 \(\times 10^{-2}\), 6.07 \(\times 10^{-2}\) and 1.04 \(\times 10^{-2}\) respectively. Although, the subjects in this current study were all newly-hired young healthy female cotton workers, and therefore different from those prospectively followed for 20 years in our previous studies, the results still support what we reported previously. Additionally, four gene-gene interactions and four gene-environment interactions in this study again provided evidence that multiple factors might contribute to the rate of FEV₁ decline interactively.

Overall, the more risk loci carried by the cotton worker, the larger the rate of FEV₁ decline over 18 months. However, younger workers (age <25 years) had more accelerated decline per-month than relatively older ones (age ≥25 years) when GRS increased. Moreover, for those carrying same number of risk loci, younger workers had more rapid decline of FEV₁ than relatively older ones. This provided evidence that endotoxin exposure tended to cause more serious damage to lung function of young workers, even with same genetic background.

Because all the workers of the cohort were lifelong young female non-smokers with relatively pure occupational exposure to endotoxin, this population provided a unique opportunity to observe gene effect to the rate of FEV₁ decline when adjusted environmental effect. The results contribute scientific support to do further functional studies of those genes in relation to the pathogenesis of endotoxin-related airways disease.

However, it is important to recognize the limitations of our study. Firstly, there is no similar external population to validate the observed SNP associations. Although we have performed internal statistical validation by bootstrap re-sampling procedures to minimize false positive discovery, we could not exclude the possibility of false discoveries among our findings. Besides, we controlled both FDR and FWER in study-wide level (0.10 and 0.15 respectively), which guaranteed the confidence of what we found. Secondly, the sample size was not large enough to detect some SNPs with modest effects. Thus, study with enlarged sample size were warranted. Finally, airborne endotoxin concentrations

**Figure 3. Stratification analysis of the associations between the rate of FEV₁ decline and the genetic risk score (GRS) of top-10 SNPs.** Each box and horizontal line represents the Coefficient and 95% CI derived from the linear regression model adjusted for height, age and FEV₁ at baseline, and average log transformed endotoxin level. \(P\) values of interaction between GRS and age/endotoxin were acquired from linear regression model with covariates adjustment.

doi:10.1371/journal.pone.0059035.g003
were estimated from sampling airborne cotton dust at fixed positions in work areas, rather than from sampling the air in the personal breathing zone of each participant, since there is no personal sampling technology developed for these kinds of exposure. The lack of personal air sampling data may be a possible source of exposure misclassification for this study. Thus, we used average of repeated measurement of endotoxin exposure level for each subject.

In summary, we performed the first large scale exploratory study of genetic variation in the rate of FEV₁ decline for newly-hired young healthy female cotton textile workers exposed to endotoxin. Genetic variants that play a role in immune inflammatory response, cell proliferation and cell apoptosis, together with environmental factors interact to affect the rate of FEV₁ decline after initiation of exposure and this decrement happens over short time. Additional replication and gene function studies are necessary to confirm what we found.

Supporting Information
File S1 Supplementary tables (Table S1–S3, S7–S10) and figures (Figure S1–S3).

File S2 The association results of imputed SNPs and genotyped SNPs 500 kb around targeted SNPs (Table S6).

References

Appendix S1 Description of the FWER and FDR controlling procedure.

Appendix S2 Regional plot of the top-10 SNP using P values derived from multi-variable linear regression model adjusted for height, age and FEV₁ at baseline, and average log transformed endotoxin level.

Appendix S3 Sixty-three nodes in the gene-gene interaction functional network.

Acknowledgments
The authors thank all cotton textile workers and textile industry staff who participated in this study. We are grateful to Helian Dai and Marcia Chertok for research assistance, Dr. Hakon Hankonarson of Children’s Hospital of Pennsylvania for running the IBC-CVD chips, and Drs Honggang Yi and Zhihang Peng for double-checking results of statistical analysis. We also appreciate two anonymous reviewers for their valuable suggestions.

Author Contributions
Conceived and designed the experiments: DCC FC HS ZH. Performed the experiments: AM LS HZ. Analyzed the data: RZ YZ YW JB. Wrote the paper: RZ YZ MC YL PX HY.