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

Citation
Zhang, Ruyang, Yang Zhao, Minjie Chu, Amar Mehta, Yongyue Wei, Yao Liu, Pengcheng Xun, et al. 2013. A large scale gene-centric association study of lung function in newly-hired female cotton textile workers with endotoxin exposure. PLoS ONE 8(3): e59035.

Published Version
doi:10.1371/journal.pone.0059035

Permanent 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

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

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

Ruyang Zhang1,2,3, Yang Zhao1,3,9, Minjie Chu1,2, Amar Mehta3,4,5, Yongyue Wei1,3, Yao Liu1,2, Pengcheng Xun6, Jianlai Li1,3, Li Su1,2, Hongxi Zhang7, Zhibin Hu1,2, Hongbing Shen1,2, Feng Chen1,2, David C. Christiani3,5,8

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 Putuo District 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<6.29×10^{-5}), including rs1910047 (P = 3.07×10^{-5}, FDR = 0.0778) and rs9469089 (P = 6.19×10^{-5}, FDR = 0.0967), as well as other eight suggestive (P<5×10^{-5}) 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_{trend} = 3.01×10^{-18}). However, the association was different among age subgroups (P = 7.11×10^{-6}) and endotoxin subgroups (P = 1.08×10^{-2}). 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.

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 Tfr113His and Hie439Arg 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 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 Metalworks 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 spiromgrams 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 (Genra 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 genie 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 27,811 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 th individual, \( FEV_{1i} \) is the observed FEV₁ value at different months (\( month_i \)), \( u_i \) is baseline FEV₁ and \( e_i \) is the random error at th month. 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.

\[
FEV_{1ij} = \mu_0 + ES \times month_{ij} + e_{ij}
\]

(1)

Secondly, we assessed the associations between ES and each SNP by fitting multiple-variable linear regression models (Equa-
adjusted for height, age, FEV\(_1\) at baseline and the average level of log-transformed endotoxin exposure. In Equation 2, \(\epsilon\) is the random error for \(i^{th}\) individual. Two genetic models (dominant and additive) of inheritance for each SNP were considered respectively. In additive model, SNPs with low frequency (\(\leq 5\%\)) of rare homozygous were excluded for lack of statistical robustness [23].

\[
ES_i = \beta_0 + \beta_1 \times SNP_i + \beta_2 \times \text{height}_i + \beta_3 \times \text{age}_i + \beta_4 \times \text{FEV}_1 + \beta_5 \times \text{endotoxin}_i + \epsilon_i
\]  

Equation 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.

\[
ES_i = \beta_0 + \beta_A \times A_i + \beta_B \times B_i + \beta_{AB} \times A_i \times B_i + \sum \beta_{\text{CoVariate}_p} + \epsilon_i
\]

Equation 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 FEV\(_1\) 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 \(1.81 \times 10^{-6}\), which is very conservative. And, the Human CVD array has a dense gene-centric design, similar studies have used a less stringent level around \(1 \times 10^{-5}\) 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 \(6.29 \times 10^{-5}\)). We imputed un-genotyped SNPs using Minimac software [28] based on LD information from hg/19 1000 Genomes database (with ASN as the reference set, released November 2010). Gene functional network was built using MetaCore\textsuperscript{TM} 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 Haplovew (version 4.2) [30].

**Results**

The mean decline of FEV\(_1\) for the overall 301 subjects was \(6.79 \pm 25.09\) ml per-month. Individuals were classified to different subgroups by age, height, FEV\(_1\) at baseline and endotoxin level respectively and the rate of FEV\(_1\) 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 deducted 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 \(P<6.29 \times 10^{-5}\). Individuals carrying the TT genotype of rs1910047 had average 4.32 ml decline of FEV\(_1\) per-month, while those carrying the TA or AA genotypes had 18.32 ml decline of FEV\(_1\) per-month. After adjustment of covariates, individuals carrying the AA genotype had more rapid FEV\(_1\) decline (15.17 ml/month) compared with those carrying the TA or AA genotype (\(P=3.70 \times 10^{-5}\), FDR = 0.0778). For rs9469089, individuals carrying the GG or GA genotype also had more rapid FEV\(_1\) decline (12.50 ml/month) than those carrying the GA genotype (\(P=6.19 \times 10^{-5}\), FDR = 0.0967).

Eight SNPs (rs32588, rs4855881, rs10515978, rs601675, rs11761231, rs10129426, rs10201627 and rs1049970) with \(P<5 \times 10^{-4}\) were considered suggestive for accelerated changes of FEV\(_1\). Two SNPs (rs32586 and rs32589) are in perfect LD with rs32588 (\(R^2 = 1.00\), 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 (\(\lambda\)) of 1.026 indicated little inflation of the
large scale genotyping study results due to population stratification (Figure S3 in File S1). We again did a sensitivity analysis with further adjustment of first eight PCs generated by EIGENSTRAT. However, none PCs was significant in the model (FDR = 0.0308 respectively); rs4855881 had interaction with age (trend = 3.01 with FDR = 0.0264; $P = 0.0206$; FDR = 0.0264 and $P = 0.0356$; FDR = 0.0300 respectively; rs4855881 had interaction with endotoxin level ($P = 0.0454$, FDR = 0.4086).

### 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 ($P_{\text{trend}} = 3.01 \times 10^{-10}$). Bootstrap re-sampling analysis showed that bootstrap $P_{\text{trend}}$ was less than $1 \times 10^{-10}$ 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 ($P = 7.11 \times 10^{-7}$) or GRS and endotoxin ($P = 1.08 \times 10^{-5}$) adjusted for other covariates (Figure 3).

### Table 2. Summary of results of the association study for top-10 SNPs using dominant model.

| SNP        | Genotype | N   | Mean± SD * | $\beta$(SE) b | $P^b$ | Proportion c |
|------------|----------|-----|------------|---------------|-------|-------------|
| rs1910047  | TT       | 248 | -4.32±21.86| -15.17(3.62)  | 3.70 \times 10^{-5} | 89.25% |
|            | TA+AA    | 53  | -18.32±34.61|               |       |             |
| rs9469089  | GG       | 217 | -3.07±23.35| -12.50(3.08)  | 6.19 \times 10^{-3} | 96.35% |
|            | GC+CC    | 83  | -16.69±26.98|               |       |             |
| rs32588    | TT       | 268 | -8.71±23.60| 17.21(4.39)   | 1.11 \times 10^{-6} | 85.50% |
|            | TC+CC    | 33  | 8.84±31.18  |               |       |             |
| rs4855881  | AA       | 257 | -9.08±22.28| 14.83(3.92)   | 1.84 \times 10^{-6} | 83.35% |
|            | AG+GG    | 44  | 6.60±34.96  |               |       |             |
| rs10515978 | AA       | 138 | -12.59±23.71| 10.09(2.77)   | 3.16 \times 10^{-6} | 90.75% |
|            | AG+GG    | 163 | -1.88±25.25 |               |       |             |
| rs601675   | AA       | 160 | -2.19±18.74| -10.14(2.79)  | 3.32 \times 10^{-6} | 85.80% |
|            | AG+GG    | 141 | -12.00±29.98|               |       |             |
| rs11761231 | AA       | 167 | -11.54±21.07| 10.05(2.77)   | 3.35 \times 10^{-6} | 89.60% |
|            | AG+GG    | 134 | -0.86±28.32 |               |       |             |
| rs10129426 | CC       | 100 | -13.44±27.26| 10.52(2.93)   | 3.88 \times 10^{-6} | 82.85% |
|            | CT+TT    | 201 | -3.38±23.31 |               |       |             |
| rs10201627 | GG       | 241 | -4.23±23.05| -12.35(3.47)  | 4.41 \times 10^{-6} | 85.45% |
|            | GT+TT    | 60  | -17.05±30.10|               |       |             |
| rs1049970  | CC       | 194 | -2.99±23.16| -10.30(2.91)  | 4.63 \times 10^{-6} | 86.60% |
|            | CT+TT    | 106 | -14.12±26.77|               |       |             |

*Mean and standard deviation of the rate of FEV₁ decline (ml/month);
$^b$Coefficient, standard error and $P$ value to SNP in linear regression model adjusted for height, age and FEV₁ at baseline, and average log transformed endotoxin level;
$^c$Proportions of $P$ value $\leq 0.005$ in 2,000 bootstrap samples.

doi:10.1371/journal.pone.0059035.t002

Association Study in Cotton Textile Workers
Figure 1. The functional network based on four pairs of gene-gene interactions identified in this study. Four gene-gene interaction functional paths were highlighted, including \( \text{TBX3} - \text{VE}_{-}\text{cadherin} (\text{CDH5}), \text{PERC} (\text{PPARGC1B}) - \text{LMAN1}, \text{PERC} (\text{PPARGC1B}) - \text{Podocalyxin}_{-}\text{like 1} (\text{PODXL}) \) and \( \text{Podocalyxin}_{-}\text{like 1} (\text{PODXL}) - \text{APEH} \).

doi:10.1371/journal.pone.0059035.g001

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

| Gene1\(^a\) | Gene2\(^a\) | $N$ | Mean± SD\(^b\) | $\beta$(SE)\(^c\) | $P$ | $P_{\text{adj}}$ \(^d\) | FDR\(^e\) |
|------------|------------|-----|----------------|----------------|-----|----------------|-----------|
| rs1910047  | rs1049970  |     | −1.39±23.13   | Ref.           |     | 4.18×10\(^{-2}\) | 0.0895    |
| TT         | CC         | 159 | −10.05±17.89  | −7.69(3.09)    | 1.35×10\(^{-2}\) |         |
| TA+AA      | CC         | 35  | −10.25±22.17  | −9.79(4.36)    | 2.56×10\(^{-2}\) |         |
| TA+AA      | CT+TT      | 18  | −34.00±47.81  | −32.59(5.73)   | 3.19×10\(^{-8}\) |         |
| rs10515978 | rs32588    |     | −8.20±7.51    | 4.11(6.29)     | 5.14×10\(^{-1}\) |         |
| AA         | TT         | 123 | −13.12±24.95  | Ref.           |     | 5.09×10\(^{-3}\) | 0.0313    |
| AA         | TC+CC      | 15  | −4.97±21.79   | 7.44(2.83)     | 8.96×10\(^{-3}\) |         |
| AG+GG      | TT         | 145 | 23.04±36.25   | 35.68(5.83)    | 2.96×10\(^{-9}\) |         |
| AG+GG      | TC+CC      | 18  | 23.04±36.25   | 35.68(5.83)    | 2.96×10\(^{-9}\) |         |
| rs11761231 | rs32588    |     | −12.38±21.74  | Ref.           |     | 5.16×10\(^{-3}\) | 0.0313    |
| AA         | TT         | 150 | −12.14±21.75  | Ref.           |     | 9.51×10\(^{-3}\) | 0.0390    |
| AA         | TC+CC      | 17  | −4.13±11.92   | 5.01(5.96)     | 4.02×10\(^{-1}\) |         |
| AG+GG      | TT         | 118 | −4.04±25.10   | 7.01(2.85)     | 1.43×10\(^{-2}\) |         |
| AG+GG      | TC+CC      | 16  | 22.62±39.11   | 36.34(6.06)    | 6.01×10\(^{-9}\) |         |
| rs11761231 | rs4855881  |     | −7.80±16.14   | 5.15(5.22)     | 3.25×10\(^{-1}\) |         |
| AA         | AA         | 144 | −12.14±21.75  | Ref.           |     | 3.25×10\(^{-1}\) |         |
| AA         | AG+GG      | 23  | −7.80±16.14   | 5.15(5.22)     | 3.25×10\(^{-1}\) |         |
| AG+GG      | AA         | 113 | −5.18±22.45   | 6.90(2.90)     | 1.81×10\(^{-2}\) |         |
| AG+GG      | AG+GG      | 21  | 22.37±42.90   | 31.87(5.44)    | 1.27×10\(^{-8}\) |         |

\(^a\)rs1910047 (TBX3), rs1049970 (CDH5), rs10515978 (LMAN1), rs32588 (PPARGC1B) and rs4855881 (APEH);
\(^b\)Mean and stand deviation of the rate of FEV1 decline (ml/month);
\(^c\)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.
\(^d\)P value for gene-gene interaction.
\(^e\)FDR q-value calculated based on all pair-wise (45 pairs) $P$ values of top-10 SNPs.

doi:10.1371/journal.pone.0059035.g001
In addition, the T-box genes may play an important role in the process of lung growth [31,32]. The results of gene-environment interaction analysis for top-10 SNPs (P<0.05). Table 4.

| Gene | Environment | N  | Mean±SD | β (SE) | P   | Pq   | FDR |
|------|-------------|----|---------|--------|------|------|-----|
| rs9469089 | Age (years) | 53 | 3.16±25.29 | Ref. |      |      |     |
|      | <18         | 53 | 3.16±25.29 | Ref. |      |      |     |
|      | 18 ~ 25     | 80 | 0.81±28.47 | -2.55(4.38) | 0.5607 |      |     |
|      | >25         | 84 | -10.71±12.20 | -23.80(10.91) | 0.0300 |      |     |
|      | GC+CC <18   | 20 | -11.10±15.35 | -13.23(6.17) | 0.0329 |      |     |
|      | GC+CC 18 ~ 25 | 26 | -22.69±40.84 | -25.84(5.85) | <0.0001 |      |     |
|      | GC+CC >25   | 37 | -15.49±18.17 | -28.43(11.53) | 0.0143 |      |     |
| rs1049970 | Age (years) | 47 | 2.03±23.22 | Ref. |      |      |     |
|      | <18         | 47 | 2.03±23.22 | Ref. |      |      |     |
|      | 18 ~ 25     | 78 | 0.78±28.14 | -2.63(4.55) | 0.5644 |      |     |
|      | >25         | 69 | -10.67±12.94 | -22.43(11.30) | 0.0481 |      |     |
|      | CT+TT <18   | 25 | -7.36±23.50 | -11.05(5.89) | 0.0618 |      |     |
|      | CT+TT 18 ~ 25 | 29 | -19.86±40.90 | -21.77(5.83) | 0.0002 |      |     |
|      | CT+TT >25   | 52 | -14.17±16.02 | -25.78(11.20) | 0.0220 |      |     |
| rs10201627 | Age (years) | 60 | 0.62±25.41 | Ref. |      |      |     |
|      | <18         | 60 | 0.62±25.41 | Ref. |      |      |     |
|      | 18 ~ 25     | 87 | 0.44±26.17 | -0.40(4.18) | 0.9247 |      |     |
|      | >25         | 94 | -11.65±15.42 | -19.86(10.84) | 0.0679 |      |     |
|      | GT+TT <18   | 13 | -7.09±12.84 | -9.39(7.23) | 0.1951 |      |     |
|      | GT+TT 18 ~ 25 | 20 | -27.66±48.70 | -28.07(6.25) | <0.0001 |      |     |
|      | GT+TT >25   | 27 | -13.98±10.02 | -22.06(11.57) | 0.0576 |      |     |
| rs4855881 | Endotoxin (EU/m³) | | | | 0.0454 | 0.4086 |     |
|      | AA <163     | 132 | -12.90±22.51 | Ref. |      |      |     |
|      | ≥163        | 125 | -5.04±21.40 | 1.46(5.42) | 0.7876 |      |     |
|      | AG+GG <163 | 15 | 12.82±45.55 | 27.98(6.51) | <0.0001 |      |     |
|      | AG+GG ≥163 | 29 | 3.38±28.41 | 8.97(6.64) | 0.1777 |      |     |

a rs9469089 (RNF5, rs1049970 (CDHS), rs10201627 (GPR55) and rs4855881 (APEH)
b Mean and stand deviation of the rate of FEV₁ decline (ml/month); c 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. d P value for gene-environment interaction. 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.t004

Discussion

Host genetic factors probably influence susceptibility to the rate of FEV₁ 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 FEV₁ decline using a large scale gene-centric approach. We found that two SNPs (rs1910047 and rs9469089) were significantly associated with the rate of FEV₁ 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.

rs1910047 is located at 12p24.21, upstream of two genes: T-box 5 (TBX5, 280 kb upstream) and T-box 3 (TBX3, 7 kb upstream). T-box genes encode transcription factors that can mediate the production of branching signals by the lung mesenchyme in the embryonic mouse lung. TBX4; TBX5 double heterozygous mutants showed decreased lung branching, indicating that T-box genes may play an important role in the process of lung growth [31,32]. In addition, the Tb5 was reported to be associated with asthma susceptibility in the developing lung in rat models [33]. Thus, it is biological plausible that SNP variation may result in aberrant activity of the TBX5 gene, which may affect lung function, accelerating FEV₁ 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 RNF5 (encoding membrane-bound ubiquitin ligase). RNF5 controls the membrane fraction of ATG4B and limits LC3 (ATG8) processing, aberrant of which may limit basal levels of autophagy and influence susceptibility to bacterial infection [36]. Previous studies have collectively suggested that RNF5 negatively regulates 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).
model [39]. 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, mannose-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 on 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]. Beosearcrech 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 ($r^2 = 0.98$) 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

---

**Figure 2. The frequency distribution of genetic risk score (GRS) and the coefficient and 95% CI to genetic risk score in linear regression model adjusted for height, age and FEV1 at baseline, and average log transformed endotoxin level.** The more risk loci the subjects carried, the larger the rate of FEV1 decline occurred ($P_{\text{trend}} = 3.01 \times 10^{-18}$).

doi:10.1371/journal.pone.0059035.g002
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 $4.83 \times 10^{-6}$, $6.07 \times 10^{-3}$ 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–S5, 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

1. Eisner MD, Anthonisen N, Coultas D, Kuenzli N, Perez-Padilla R, et al. (2010) An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 182: 693–718.
2. Harkema JR, Hotchkiss JA (1993) Ozone- and endotoxin-induced mucous cell metaplasias in rat airway epithelium: novel animal models to study toxicant-induced epithelial transformation in airways. Toxicol Lett 68: 251–263.
3. Janus ED, Phillips NT, Carrell RW (1985) Smoking, lung function, and alpha 1-antitrypsin deficiency. Lancet 1: 152–154.
4. Liebers V, Brenning T, Raulf-Heimsoth M (2006) Occupational endotoxin-exposure and possible health effects on humans. Am J Ind Med 49: 474–491.
5. Beck GJ, Schachter EN, Maurer MD, Schilling RS (1982) A prospective study of chronic lung disease in textile workers. Am Intern Med 97: 645–651.
6. Castellan RM, Olenchook SA, Kinsey KB, Hankinson JL (1987) Inhaled endotoxin and decreased spirometric values. An exposure-response relation for cotton dust. N Engl J Med 317: 605–610.
7. Kennedy SM, Christiani DC, Eisen EA, Wegman DH, Greaves IA, et al. (1987) Cotton dust and endotoxin exposure-response relationships in cotton textile workers. Am Rev Respir Dis 135: 194–200.
8. Wilk JB, Walter RE, Laramie JM, Gotthelf DJ, O'Connor GT (2007) Framingham Heart Study genome-wide association results: for pulmonary function measures. BMC Med Genet 8 Suppl 1: S1.
9. Wilk JB, Chen TH, Gotthelf DJ, Walter RE, Nagle MW, et al. (2009) A genome-wide association study of pulmonary function measures in the Framingham Heart Study. PLoS Genet 5: e1000429.
10. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, et al. (2011) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 5: e1000421.
11. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, et al. (2010) Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat Genet 42: 45–52.
12. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, et al. (2010) Genome-wide association study identifies five loci associated with lung function. Nat Genet 42: 36–44.
13. Arigas MS, Loth DW, Wain LV, Gharib SA, Obeidat M, et al. (2011) Genotype-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet 43: 1082–1090.
14. Han J, Zhou W, Wang X, Zhang H, Sun B, et al. (2005) Microsilicic acid epoxide hydrolase, endotoxin, and lung function decline in cotton textile workers. Am J Respir Crit Care Med 171: 165–170.
15. Zhang H, Han J, Wang X, Zhou W, Sun B, et al. (2007) TNF polymorphisms modify endotoxin exposure-associated longitudinal lung function decline. Occup Environ Med 64: 409–413.

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.
31. Cebra-Thomas JA, Bronner J, Gardner R, Lam GK, Sheipe H, et al. (2005) T-box gene products are required for mesenchymal induction of epithelial branching in the embryonic mouse lung. Dev Dyn 226: 82–90.
32. Arora R, Metzger RJ, Papaioannou VE (2012) Multiple roles and interactions of Tbx4 and Tbx5 in development of the respiratory system. PLoS Genet 8: e1002866.
33. Carpe N, Mandeville I, Ribeiro L, Ponton A, Martin JG, et al. (2010) Genetic influences on asthma susceptibility in the developing lung. Am J Respir Cell Mol Biol 43: 720–730.
34. Lu J, Li XP, Dong Q, Kung HF, He ML (2010) TBX2 and TBX3: the special value for anticancer drug targets. Biochem Biophys Acta 1806: 268–274.
35. Yin XJ, Xu JN, Zou CQ, He FS, Fang FD (2004) Genes differentially expressed in human lung fibroblast cells transformed by glycidyl methacrylate. Biomed Environ Sci 17: 397–407.
36. Kuang E, Okumura CY, Sheffy-Levin S, Varsano T, Shu VC, et al. (2012) Regulation of ATG4B Stability by RNF5 limits basal levels of autophagy and influences susceptibility to bacterial infection. PLoS Genet 8: e1003007.
37. Zhong B, Zhang Y, Tan B, Liu TT, Wang YY, et al. (2011) The E3 ubiquitin ligase RN5 targets virus-induced signaling adapter for ubiquitination and degradation. J Immunol 184: 6249–6255.
38. Zhong B, Zhang L, Lei C, Li Y, Mao AP, et al. (2009) The ubiquitin ligase RNF5 regulates antiviral responses by mediating degradation of the adaptor protein MITA. J Cell Biochem 114: 325–335.
39. Carmeliet P, Lampugnani MG, Moons L, Breviario F, Compernolle V, et al. (1999) Targeted deficiency or cytosolic truncation of the VE-cadherin gene in mice impairs VEGF-mediated endothelial survival and angiogenesis. Cell 98: 147–157.
40. Roeckel N, Woerner SM, Kloor M, Yuan YP, Patsos G, et al. (2009) High frequency of LMAN1 abnormalities in colorectal tumors with microsatellite instability. Cancer Res 69: 292–299.
41. Lee SH, Jang AS, Woo Park S, Park JS, Kim YK, et al. (2011) Genetic effect of single-nucleotide polymorphisms in the PPARGC1B gene on airway hyperreactivity in asthma patients. Clin Exp Allergy 41: 1533–1544.
42. Korman BD, Selkin MF, Taylor KE, Le JM, Lee AT, et al. (2009) The chromosome 7q region association with rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. Nat Genet 41: 1156–1159.
43. Kohno T, Takayama H, Yamaguchi M, Takano H, Yamaguchi N, et al. (1993) Deletion mapping of chromosome 3p in human uterine cervical cancer. Oncogene 8: 1823–1832.
44. Korthagen NM, van Mooreel CH, Barlo NP, Kazemier KM, Ruiten HJ, et al. (2012) Association between variations in cell cycle genes and idiopathic pulmonary fibrosis. PLoS One 7: e30442.
45. Imboden M, Schwartz J, Schindler C, Curjuric I, Berger W, et al. (2009) Decreased PM10 exposure attenuates age-related lung function decline: genetic variants in p33, p21, and CCND1 modify this effect. Environ Health Perspect 117: 1420–1427.
46. Guan XJ, Song L, Han FF, Cui ZL, Chen X, et al. (2013) Mesenchymal stem cells protect cigarette smoke-damaged lung and pulmonary function partly via VEGF/VEGFR receptors. J Cell Biochem 114: 325–335.
47. Sugiyama T, Wingeader E, Reuter I, Hermjakob H, Kel AE, et al. (1998) Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res 26: 362–367.
48. German Z, Chambliss KL, Pace MC, Arnett UA, Lowenstein CJ, et al. (2000) Molecular basis of cell-specific endothelial nitric-oxide synthase expression in airway epithelium. J Biol Chem 275: 8103–8109.
49. Gao W, Shen H, Liu L, Xu J, Liu J, et al. (2011) MRK-21 overexpression in human primary squamous cell lung carcinoma is associated with poor patient prognosis. J Cancer Res Clin Oncol 137: 557–566.
50. Ishiguro H, Onaivi ES, Horiiuchi Y, Imai K, Komaki G, et al. (2011) Functional polymorphism in the GPR55 gene is associated with anorexia nervosa. Synapse 65: 103–108.
51. Andradas C, Caffarel MM, Perez-Gomez E, Salazar M, Llorente M, et al. (2011) The orphan G protein-coupled receptor GPR55 promotes cancer cell proliferation via ERK. Oncogene 30: 245–252.