Genome-wide gene–environment interaction analysis for asbestos exposure in lung cancer susceptibility

Sheng Wei[1,4], Li-E Wang[1], Michelle K.McHugh[1], Younghun Han[1], Momiao Xiong[2,3], Christopher I.Amos[1,3], Margaret R.Spitz[1] and Qingyi Wei[1,3,*]

[1]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA, [2]Human Genetics Center, School of Public Health, The University of Texas, Houston, TX, 77030, [3]Program in Human and Molecular Genetics, The University of Texas Graduate School of Biomedical Sciences, Houston, TX, 77030, USA and [4]Department of Epidemiology and Health Statistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, 13 Hangkong Road, Wuhan, 430030, China

*To whom correspondence should be addressed. Tel: 713-792-3020;
Fax: 713-563-0999;
E-mail: qwei@mdanderson.org.

Asbestos exposure is a known risk factor for lung cancer. Although recent genome-wide association studies (GWASs) have identified some novel loci for lung cancer risk, few addressed genome-wide gene–environment interactions. To determine gene–asbestos interactions in lung cancer risk, we conducted genome-wide gene–environment interaction analyses at levels of single nucleotide polymorphisms (SNPs), genes and pathways, using our published Texas lung cancer GWAS dataset. This dataset included 317 498 SNPs from 1154 lung cancer cases and 1137 cancer-free controls. The initial SNP-level P-values for interactions between genetic variants and self-reported asbestos exposure were estimated by unconditional logistic regression models with adjustment for age, sex, smoking status and pack-years. The P-value for the most significant SNP rs13383928 was 2.17×10^{-6}, which did not reach the genome-wide statistical significance. Using a versatile gene-based test approach, we found that the top significant gene was C7orf54, located on 7q32.1 ($P = 8.90 \times 10^{-5}$). Interestingly, most of the other significant genes were located on 11q13. When we used an improved gene-set-enrichment analysis approach, we found that the Fas signaling pathway and the antigen processing and presentation pathway were most significant (nominal $P < 0.001$; false discovery rate < 0.05) among 250 pathways containing 17 572 genes. We believe that our analysis is a pilot study that first describes the gene–asbestos interaction in lung cancer risk at levels of SNPs, genes and pathways. Our findings suggest that immune function regulation-related pathways may be mechanistically involved in asbestos-associated lung cancer risk.

Introduction

Asbestos, a term for a group of naturally occurring hydrated silicate fibers, has been widely used in >3000 manufactured products. Exposure to asbestos can result in pleural lung fibrosis, and inhaled asbestos may cause peritoneal malignant mesothelioma and lung cancer (1). It is estimated that occupationally related asbestos exposure contributes to ~5–7% of all lung cancer cases (2). Although the use of asbestos has been banned or severely restricted since early 1970s in many developed countries, asbestos-related lung diseases still pose a great public health threat because of a long-latency period from asbestos exposure to the incidence of asbestos-induced diseases (3). Genetic factors have been suggested to be involved in asbestos-related carcinogenesis and lung genotoxicity by causing oxidative stress,

inflammation, DNA damage-repair response, mutations, chromosomal aberrations, mitochondrial malfunction and apoptosis (4,5). To date, the impact of individual genetic variations on asbestos-related lung cancer risk is still not well understood. A small number of studies using a candidate gene approach reported that polymorphisms in genes encoding xenobiotic metabolizing enzymes (e.g. *GSTM1*, *GSTT1*, *MPO*, *CYP1A1* and *CYP2E1*) and manganese superoxide dismutase (e.g. *SOD2* and *MnSOD*) were associated with asbestos-related lung cancer risk (6–8).

In the past few years, genome-wide association studies (GWASs) have successfully identified some novel genetic susceptibility variants located on 15q21, 5p15.33 and 6p21.33 that are involved in lung cancer etiology (9,10). However, even the most significant single nucleotide polymorphisms (SNPs) identified in GWASs have only accounted for a small proportion of the estimated heritability (11). Besides rare variants and structural variants, gene–environment interaction is expected to be a possible source of missing heritability not explained by current GWASs (12,13). Although environmental factors, such as tobacco smoke and asbestos exposure, play an important role in lung cancer development, few reported studies have investigated the potential of the genome-wide gene–environment interactions following the initial findings of lung cancer GWASs. Because current GWASs have been designed to detect the main effect of genetic variants, which often suffer from inadequate sample sizes, the ability to detect gene–environment interactions even in the single SNP analysis has been limited (14).

To improve the power of current GWASs to detect gene–environment interactions at the SNP level, both gene- and pathway-based analytic approaches, integrating prior biological knowledge and association studies, have been proposed recently (15,16). These approaches combine associations of genetic variants in either the same gene or biological pathway with disease risk and enable us to summarize the associations on a functional basis. Furthermore, these approaches enhance researchers' ability to investigate additional susceptibility genes and pathways to address 'missing heritability', with the potential of unraveling the mechanisms underlying the etiology of complex diseases. Several recent studies have shown that gene-based and pathway-based approaches to gene–environment interactions using existing GWAS datasets could successfully facilitate the mining of biological information and provide additional complementary information (17,18).

Although not formerly possible, researchers are now able to identify evidence in the whole genome-wide scale for potential interactions between genetic variants and asbestos exposure in lung cancer risk. Access to the existing GWAS datasets for lung cancer provided us a unique opportunity to further characterize the gene–environment interaction that may play a critical role in the development of lung cancer. In this study, we determined genetic variations involved in the interactions of gene–asbestos exposure in lung cancer risk by conducting genome-wide gene–environment interaction analyses at levels of SNPs and pathways, using data from the published Texas lung cancer GWAS.

Materials and methods

Study population

The Texas lung cancer GWAS population has been described previously (10). Briefly, this study included 1154 patients newly diagnosed with histopathologically confirmed and untreated non-small-cell lung cancer (NSCLC) and 1137 cancer-free controls, who were non-Hispanic white ever smokers and frequency matched by age (±5 years), sex and smoking status. Cases were recruited from The University of Texas MD Anderson Cancer Center, and cancer-free controls were recruited from the Kelsey-Seybold Clinic, Houston's largest multispecialty group practice. All the participants signed an informed

consent, provided a 30 ml blood sample, and completed a personal interview using a questionnaire that included information about environmental exposures including tobacco use and asbestos exposure.

Genotyping and data cleaning

For the Texas lung cancer GWAS, genotyping procedures of Illumina HumanHap300 v1.1 BeadChips (Illumina, San Diego, CA) with genomic DNA and quality control have been described elsewhere (10). In brief, the Chip contained 317 503 tagging SNPs derived from the International HapMap project phase I data. After data cleaning, markers deviated from the Hardy–Weinberg equilibrium in the controls ($P < 0.0001$) and SNPs with a minor allele frequency <0.01 were excluded. The final analysis included 1154 lung cancer cases and 1137 cancer-free controls with 317 498 tagging SNPs.

Exposure assessment

Assessment methods for exposure have been described previously (6). In brief, self-reported exposure to asbestos was assessed by a positive answer to 'have you handled, used, or been in contact with for at least 8 h a week for a year or more'. Former smokers were defined as a person who had quit smoking >1 year before enrollment, and current smokers included recent quitters who had quit smoking within the past 12 months. Pack-years were calculated as the years smoked times the average number of cigarettes per day divided by 20. Pack-years were also classified into three groups (<25 pack-years, 25≤ and <50 pack-years and ≥50 pack-years) by upper quartile and lower quartile of pack-years of the controls.

Statistical analysis

Chi-square tests were used to compare differences in distributions of demographic variables between the cases and controls. Unconditional logistic regression analyses were used to calculate odds ratios (OR) and 95% confidence intervals (CI) to determine the main effect of self-reported asbestos exposure and smoking status on the risk of lung cancer.

SNP-level interaction

For the SNP-level gene–asbestos interaction, we used a standard approach to test gene–environment (G × E) interaction by performing a 1 *df* test of H_0: $\beta_{ge} = 0$ for each SNP based on the model using the equation: logit $P (D = 1 \mid g,e) = \beta_0 + \beta_g G + \beta_e E + \beta_{ge} GE$, where D is an indicator of the disease status (cases, $D = 1$; controls, $D = 0$); G is the genotypic code for each genotype of a SNP (e.g. additive in this study); E is exposure status (exposed, $E = 1$; unexposed, $E = 0$). For a given SNP, an unconditional logistic regression model included the main effects for the genotypes (assuming an additive model), asbestos exposure, all covariates (age, sex, smoking status and pack-years) and the interaction term for genotypes and asbestos exposure was implemented in the PLINK1.07 'G × E' procedure (19). The *P*-value of the interaction term ($\beta_{ge} GE$) was used to assess the significance of the interaction between genetic variants and asbestos exposure. Since there were 307 944 SNPs having *P*-values for interaction term in the logistic regression model, a *P*-value of 1.6×10^{-7} (0.05/307 944) was considered the threshold for a statistically significant test, taking into account the correction for multiple tests.

Gene-level interaction

To determine whether there were genes that might have an excessive interaction with asbestos exposure and lead to an increased risk of lung cancer, we used a versatile gene-based test (VEGAS) to assess the gene-level interaction by assessing the *P*-values of the SNP-level gene–asbestos exposure interaction (20). The gene-based test statistic was the sum of the chi-squared 1 *df* statistics within that gene (with the gene boundaries of 50 kb away from two sides of the gene) and the linkage disequilibrium between markers and number of SNPs per gene; this statistic was based on simulations from the multivariate normal distribution. The empirical gene-based *P*-value was calculated by the proportion of simulated test statistics that had exceeded the observed gene-based test statistic. Since there were 17 572 autosomal genes included in the data analyses, a *P*-value $< 2.8 \times 10^{-6}$ (0.05/17 572) was considered statistically significant.

Pathway-level interaction

To analyze pathway-level interaction, we used the improved gene-set-enrichment analysis approach (*i*-GSEA), an implementation and extension of the original gene-set-enrichment analysis (GSEA) (21) that estimates the pathway-level interactions between genetic variants and asbestos exposure using *P*-values obtained at the SNP level. In the *i*-GSEA analysis, we use the maximum –log (*P*-value) or statistics of all the SNPs mapped to a gene to represent the gene (*t*). Then, we ranked all genes in GWAS by decreasing values, and the genes at the top 5% with the small *P*-values were considered statistically significant. For each given pathway S, significance proportion-based enrichment score was calculated to estimate the enrichment of genotype–phenotype association in a particular gene set S. Then, *i*-GSEA performed label permutations to calculate

nominal *P*-values to assess the significance of the pathway-based enrichment score and the false discovery rate (FDR) to correct multiple testing. *i*-GSEA was implemented by using SNP-label permutations instead of phenotype-label permutations in a classical GSEA in order to calculate nominal *P*-values. Moreover, *i*-GSEA focuses on the pathways or gene sets with the highest proportions of significant genes instead of relying solely on the total significance generated from either a few or many significant genes, thus improving the sensitivity to identify pathways or gene sets that represent the combined effects of all possibly modest SNPs/genes. Since *i*-GSEA uses *P*-values at both SNP and gene levels as input data, it was convenient to perform the pathway-based analysis in the existing GWAS dataset, especially for the GWAS gene–environment interaction analysis. For *P*-values at the SNP level, we mapped all SNPs within 20 kb around a gene estimated precisely for the association of a gene. Considering the relatively low coverage of Illumina HumanHap300 v1.1 BeadChips, we also compared the results by mapping SNPs within 20 kb around the gene and by mapping SNPs within 100 kb around the gene. Next, the canonical pathways or gene sets were used for further pathway analyses. These canonical pathways were extracted and curated from Molecular Signatures Database (MSigDB) v2.5 (http://www.broadinstitute.org/gsea/msigdb/) (22). MSigDB included gene sets denoting canonical pathways integrated from a variety of online resources like KEGG (23), signal transduction knowledge environment (Science Signaling, [24]), BioCarta (25), GO (gene ontology) terms with high confidence (26). To reduce the multiple-testing issue and to avoid too narrow or too broadly defined pathways, we restricted our analysis to pathways with 20–200 genes as the default (27). Pathways or gene sets with FDR < 0.25 were regarded as mild confidence that the G × E interactions were enriched in a pathway; FDR < 0.05 were regarded as high confidence that the interactions were enriched in a gene or pathway.

Results

Demographics of subjects and main effect of asbestos exposure

A total of 1154 lung cancer cases and 1137 controls were included in the study. All of the cases were non-small-cell lung cancer (NSCLC) including adenocarcinoma (51.6%), squamous (26.8%) and other NSCLCs (21.5%). As shown in Table I, age distributions were significantly different but sex distribution was similar between cases

Table I. Distribution of demographic, smoking, self-reported asbestos exposure and histology in subjects in Texas lung cancer GWAS

Characteristics	Cases (n = 1154)		Controls (n = 1137)		OR (95% CI)[a]	P
	No.	%	No.	%		
Age (years)						<0.000[a]
≤50	195	16.9	137	12.1		
51–60	286	24.8	359	31.6		
61–70	391	33.9	491	43.2		
>70	282	24.4	150	13.2		
Sex						0.855[a]
Male	658	57.0	644	54.6		
Female	496	43.0	493	43.4		
Smoking Status						
Current smoker	600	52.0	657	57.8	1.00	
Former smoker	554	48.0	480	42.2	1.26 (1.07–1.49)	0.005[b]
Pack-years						
<25	209	18.1	288	25.3	1.00	
26–50	458	36.7	492	43.3	1.28 (1.03–1.60)	0.026[b]
>50	487	42.2	357	31.4	1.88 (1.50–2.35)	<0.000[b]
Asbestos exposure[c]						
No	754	67.2	873	77.0	1.00	
Yes	368	32.8	260	23.0	1.64 (1.36–1.97)	<0.000[b]
Histology						
Adenocarcinoma	597	51.7				
Squamous	309	26.8				
Other NSCLC	248	21.5				

[a]Two-sided chi-square test.
[b]OR, 95% CI and *P*-values were estimated by unconditional logistic regression model.
[c]Asbestos exposure status for 32 cases and 4 controls were missed.

and controls. There were more current smokers in controls than in cases ($P = 0.005$), but cases had smoked more pack-years than the controls ($P < 0.001$). In the models adjusted for smoking status and pack-years, we found that self-reported asbestos exposure was significantly associated with increased lung cancer risk (OR = 1.64, 95% CI = 1.36–1.97).

SNP-level interaction

Overall, 307 944 SNPs were included in our analysis. The *P*-values for the interaction between each of SNPs and asbestos exposure are shown in a Manhattan plot (Figure 1). The top 20 significant SNPs are

listed in Table II. SNP rs13383928, located in the 3′ flanking region of *PTHR2*, was the most significant SNP and had a *P*-value of 2.17×10^{-6} for the SNP-level interaction. The other top significant SNPs were distributed in several chromosomes with similar *P*-values; however, no SNPs reached the genome-wide significance (10^{-7}).

Gene-level interaction

In the gene-based gene–asbestos interaction analysis, 17 572 genes were mapped according to positions on the University of California, Santa Cruz Genome Browser hg18 assembly (Table III). *C7orf54*, located on 7q32.1, had a *P*-value of 8.90×10^{-5} and was the most
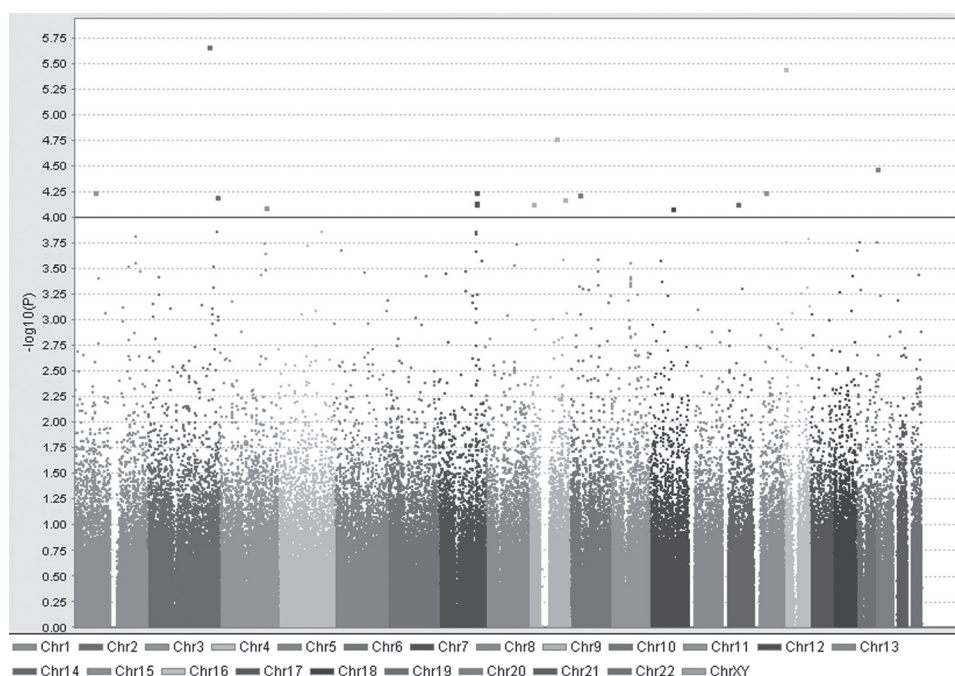

Fig. 1. SNP-level genome-wide gene self-reported asbestos exposure interactions with –log10 *P*-value for genome-wide gene–asbestos exposure interaction and lung cancer risk. Each of *P*-values was calculated by the interaction term of an additive model for risk associated with self-reported asbestos exposure in the unconditional logistic regression model with adjustment for age, sex, smoking status and smoking pack-years.

Table II. The top 20 SNPs in the SNP-level genome-wide interactions between genetic variants and self-reported asbestos exposure in lung cancer risk

SNP	Chromosome	Position	Gene	Location	Allele change	MAF[a]	*P*-inter[b]
rs13383928	2	209404106	*PTHR2*	Flanking 3UTR	A > C	0.04	2.17E-06
rs9635542	16	4941381	*PPL*	Flanking 5UTR	A > G	0.10	3.52E-06
rs10993705	9	90678411	*SYK*	Intron	C > A	0.09	1.67E-05
rs6056505	20	9191782	*PLCB4*	Intron	G > A	0.08	3.35E-05
rs10520058	15	36337292	*SPRED1*	Intron	C > A	0.03	5.63E-05
rs502376	1	66934528	*TCTEX1D1*	Intron	A > G	0.08	5.65E-05
rs6467157	7	127254714	*LRRC4*	Flanking 3UTR	A > G	0.28	5.65E-05
rs4749711	10	32475194	*KIF5B*	Flanking 5UTR	C > A	0.27	6.00E-05
rs10174028	2	233821934	*INPP5D*	Flanking 5UTR	G > A	0.05	6.22E-05
rs1179766	9	120903526	*C5*	Flanking 5UTR	G > A	0.06	6.64E-05
rs6962635	7	127282576	*LRRC4*	Flanking 5UTR	G > A	0.25	7.11E-05
rs12345299	9	15237714	*C9orf52*	Intron	G > A	0.09	7.30E-05
rs6953231	7	127275506	*LRRC4*	Flanking 5UTR	G > A	0.25	7.39E-05
rs3007165	14	50701569	*TRIM9*	Flanking 5UTR	G > A	0.06	7.40E-05
rs4432598	3	156234614	*MME*	Flanking 5UTR	A > G	0.03	7.91E-05
rs2037248	12	75154511	*C12orf58*	Flanking 3UTR	G > A	0.07	8.15E-05
rs12706823	7	127160429	*SND1*	Intron	G > A	0.26	1.37E-04
rs10200271	2	233822562	*INPP5D*	Flanking 5UTR	A > C	0.05	1.37E-04
rs4835011	4	146917841	*MMAA*	5UTR	A > G	0.08	1.38E-04
rs17151466	7	127100031	*SND1*	Intron	G > A	0.27	1.44E-04

[a]MAF, minor allele frequency.
[b]*P*-values were estimated by the interaction term in unconditional logistic regression model with adjustment by age, sex, smoking status and pack-years.

Table III. The top 20 genes in the gene-level genome-wide interactions between genetic variants and self-reported asbestos exposure and lung cancer risk

Gene	Chromosome location	Gene description	Number of SNPs	P^a
C7orf54	7q32.1	Chromosome 7 open reading frame 54	7	8.90E-05
LRRC4	7q31.3	Leucine rich repeat containing 4	9	1.87E-04
SND1	7q31.3	Staphylococcal nuclease and tudor domain containing 1	38	2.87E-04
RNF121	11q13.4	Ring finger protein 121	10	4.22E-04
NUMA1	11q13	Nuclear mitotic apparatus protein 1	5	7.28E-04
C11orf59	11q13.4	Chromosome 11 open reading frame 59	5	7.94E-04
LRRC51	11q13.4	Leucine rich trans-membrane and *O*-methyltransferase domain containing	5	8.26E-04
C11orf51	11q13.4	Chromosome 11 open reading frame 51	8	1.16E-03
IL18BP	11q13	Interleukin 18 binding protein	2	1.95E-03
ADRBK1	11q13.1	Adrenergic, beta, receptor kinase 1	7	2.80E-03
FBXL11	11q13.2	Lysine (K)-specific demethylase 2A	13	3.03E-03
NGDN	14q11.2	Neuroguidin, EIF4E-binding protein	14	3.34E-03
ODF4	17p13.1	Outer dense fiber of sperm tails 4	16	3.35E-03
ZNF559	19p13.2	Zinc finger protein 559	6	4.56E-03
EXOC6	10q23.33	Exocyst complex component 6	30	4.89E-03
TADA1L	1q24.1	Transcriptional adaptor 1	6	5.62E-03
AGTR1	3q21–q25	Angiotensin II receptor, type 1	18	5.76E-03
ZNF177	19p13.2	Zinc finger protein 177	6	5.94E-03
ARHGEF15	17p13.1	Rho guanine nucleotide exchange factor (GEF) 15	16	6.10E-03
TXN2	22q13.1	Thioredoxin 2	19	6.19E-03

[a]*P*-values were estimated by a versatile gene-based test approach.

significant gene. In contrast, most of the other top 20 significant genes were located at a narrow region of chromosome 11q13, but none reached statistical significance, after multiple test corrections.

Pathway-level interaction

As shown in Table IV, 171 112 SNPs of 307 944 SNPs were mapped to 15 961 genes that were assigned to 250 pathways. When mapping SNPs were limited within 20 kb around a gene, only 4 pathways were significantly enriched with association signals and had a FDR <0.25 and nominal *P*-value <0.01. The top two significant pathways were the Fas signaling pathway (nominal *P* < 0.001 and FDR = 0.034) and the antigen processing and presentation pathway (nominal *P* = 0.001 and FDR = 0.055). After we performed expanded *i*-GSEA analyses (i.e. mapping SNPs on a gene within a 100 kb flanking range around the gene), we found that the Fas signaling pathway was still the top significant pathway with a nominal *P* = 0.001 and FDR = 0.16 (Supplementary Table I is available at *Carcinogenesis* Online).

The Fas signaling pathway was annotated by the Signal Transduction Knowledge Environment database (28), and there were 57 mapped genes covered by the Texas lung cancer GWAS dataset in this pathway, of which 23 genes were significant for the pathway-level interaction (Figure 2A). The antigen processing and presentation pathway was annotated by the KEGG pathway database, and there were 59 mapped genes covered by the Texas lung cancer GWAS dataset, of

Table IV. Pathways with nominal *P* ≤ 0.05 and FDR < 0.25 from pathway-based analysis for the genome-wide interactions between genetic variants and self-reported asbestos exposure and lung cancer risk by mapping SNPs in the region with 20 kb around a gene

Pathway/Gene set name	Description	P-value[a]	FDR	Significant genes/selected genes/all genes
Fas signaling pathway	The Fas receptor induces apoptosis and nuclear factor-kappaB activation when bound to Fas ligand	<0.0010	0.034	23/57/68
Antigen processing and presentation	Genes involved in antigen processing and presentation	0.0010	0.055	19/59/83
Type I diabetes mellitus	Genes involved in type I diabetes mellitus	<0.0010	0.089	19/41/45
Granule cell survival pathway	The survival and differentiation of granule cells in the brain is controlled by pro-growth PACAP and pro-apoptotic ceramides.	0.0030	0.113	12/25/28

[a]*P*-values were computed based on 10 000 permutations.

which 19 genes were significant for the pathway-level interaction (Figure 2B). Although the two pathways were significant at the pathway level, most of the significant genes in the two pathways only had relatively high *P*-values at the SNP level (Supplementary Tables II and III are available at *Carcinogenesis* Online).

Discussion

To the best of our knowledge, this is the first pilot study to have investigated the genome-wide gene–asbestos interactions in lung cancer based on the current GWAS design. Although we did not find statistical evidence for the hypothesized gene–asbestos interaction in the etiology of lung cancer at levels of SNP and gene in our published Texas lung cancer GWAS dataset, the pathway-based analysis suggested a clue that two pathways of immune function regulation (i.e. the Fas signaling and the antigen processing and presentation pathways) might play a role in the etiology of asbestos-related lung cancer.

It is not surprising that we failed in identifying statistical evidence of gene–asbestos exposure interaction at the SNP level, because this finding is similar to what we have seen in most published GWASs that reported results of gene–environment interactions. For example, a GWAS of metabolic traits in the Northern Finland Birth Cohort 1966 (NFBC1966) reported few SNPs that were significant in genome-wide gene–environment interaction analyses (29). Another study of the influence of genetic and environmental factors on susceptibility to rheumatoid arthritis also failed to find genome-wide significant gene–environment interaction [30]. Of >900 GWASs that have been published to date, very few have reported significant gene–environment interactions (11), which suggests that the current GWAS design does not provide enough statistical power to detect interactions at the single SNP level, because such a design mainly focuses on the main effect of a signal SNP on risk of common, complex diseases.

The gene-based analysis of gene–environment interaction considers interactions between environmental exposures and potential biomarkers (usually SNPs) within a gene, thus leading to an increased statistical power for detecting an association (31). We identified three top significant genes, and their functions are either unknown or not well studied. For example, *C7orf54* is a gene with unknown functions; *LRRC4* is a tumor suppressor gene that has not been well studied but may be involved in modulating the extracellular signal-regulated
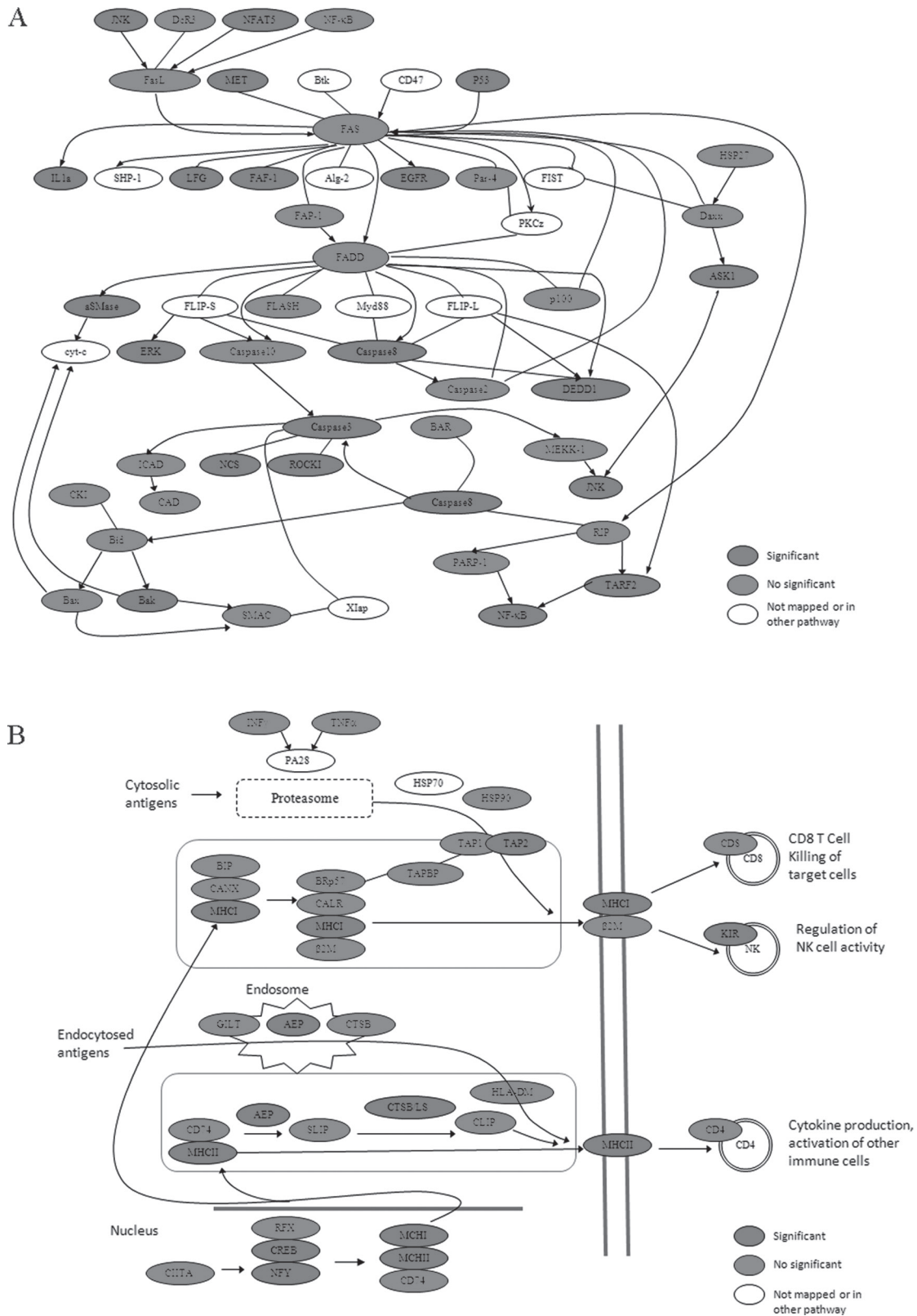
Fig. 2. Schematic illustrations of the two significant pathways in the genome-wide gene self-reported asbestos exposure interaction analysis. (**A**) Modified diagram of the Fas signaling pathway from Science Signaling database (http://stke.sciencemag.org/cgi/cm/stkecm;CMP_7966). (**B**) Modified diagram of the antigen processing and presentation pathway from KEGG (http://www.kegg.jp/kegg/pathway/hsa/hsa04612.html). Genes are marked according to their *P*-value that assesses the significance of the most strongly associated SNP in each gene. Lines indicate the interactions among the genes. Genes with dark black color were significant, whereas genes with lighter color were not significant. Genes without color were not mapped by the Texas lung cancer GWAS dataset or not included in the current pathways.

kinase, protein kinase B, nuclear factor-kappaB pathway in glioma cells (32); *SND1* is a component of the RNA-induced splicing complex that may lead to the degradation of specific mRNAs (33). However, there was no statistical evidence of gene–environment interaction for these three top significant genes. Interestingly, most of the other significant genes are located in a region at 11q13. For example, SNPs in *RNF121* and *NUMA1* have been reported to be associated with risk of Crohn's disease, an autoimmune disease, in a large GWAS (34). *C11orf59* was reported to play a role in the process of host against coronavirus infection (35). Few studies have reported any associations of SNPs at 11q13 with the development of asbestos-induced diseases; however, genetic variants in this region have been reported to be associated with risk of Crohn's disease, type 2 diabetes and breast cancer (36–38). It is biologically plausible that top significant genes identified in our gene–asbestos interaction analyses might be related to immune function regulation, and the findings from our pathway-based analysis further supported this hypothesis.

Of the two pathways we identified, the top significant pathway was the Fas signaling pathway, in which the participating components are mainly expressed on the cell membrane of lymphocytes and involved in regulation of tissue homeostasis in the immune system by inducing apoptosis (39). Fas signaling could trigger the apoptosis of T lymphocytes and play a critical role against autoimmunity and tumor development (40,41). Although polymorphisms in the Fas signaling pathway have been reported to be associated with an increased risk of lung cancer (42), few studies have reported an interaction between genetic variations in the Fas signaling pathway and asbestos exposure in the etiology of lung cancer. But previous studies showed that mice deficient for *Arf* were susceptible to accelerated asbestos-induced malignant mesothelioma and that homozygous loss of the *Faf1* (FAS-associated factor 1) locus deregulated the tumor necrosis factor-α-induced nuclear factor-kappaB signaling, which had previously been implicated in asbestos-induced oncogenesis (43,44). Recent work also suggested that asbestos exposure might increase FAS expression in lung tissue (45). So our finding provides a possible biological explanation that genetically determined function of the Fas-mediated apoptotic pathway may be involved in asbestos-related lung carcinogenesis.

The second significant pathway, the antigen processing and presentation pathway, is involved in the regulation of immune function, consisting of a subgroup of human immune-system genes set in the major histocompatibilty complex (46). The genes in this pathway help to discriminate and present antigens to T-cell receptors, natural killer cells and other immune cells (47), and studies have found that asbestos can disturb autoimmunity and tumor immunity, while inducing DNA damage and apoptosis in alveolar epithelial cells (48,49). This is because asbestos possesses the super antigenic capacity against T cells and the restricted overexpression of the T-cells receptor V beta without clonal expansion (50). In addition, the function of natural killer cells and regulatory T cells was depressed when they were exposed to asbestos (51). If the immune function is impaired as a result of the asbestos-induced pathogenesis, genetic variations in these pathways will modulate the risk of asbestos-related diseases, including lung cancer. The findings of gene–asbestos interactions at levels of SNP, gene and pathway are seemingly not consistent for specific genes or specific chromosome regions in this study, but it is of interest that all the interactions at these three levels seem to tend to point to the genes or pathways involved in immune function regulation, which might be one of the possible mechanisms involved in the asbestos-related carcinogenesis.

Some limitations in this study should be addressed. One limitation was the use of self-reported asbestos exposure. Our previous study had shown that risk estimates for interactions of genetic variants with self-reported asbestos exposure, self-reported occupational exposure and self-reported employment in asbestos-related industries were similar, in which all these asbestos exposure measurements were reviewed and confirmed by a certified industrial hygienist (6), although reporting bias may still exist. Second, the coverage of the genotyping chip in this study is relatively low. A single SNP may be simultaneously mapped on nearby genes in the gene–asbestos interaction analyses, which may lead to errors in the analyses at the gene and pathway levels. Third, we did not have the access to similar datasets available for us to replicate our findings. This may be due to a myriad of reasons, some of which include lack of asbestos exposure data in publicly available lung cancer GWAS datasets and/or the lack of common or comparable definitions for asbestos exposure. Furthermore, the current pathway-based analysis assumes that the local SNP only modify the function of the local gene, which is only partially true. The *cis* and *trans* regulation of the genes should be considered in the future pathway-based analysis. In addition, for the limitation of relatively small sample size, we have not performed the stratified G × E interactions analyses by histology, but the most informative gene–environment interactions between asbestos exposure and lung cancer risk could only be determined in large cohorts of patients with the histology of interest associated with that exposure. Despite these limitations, our study demonstrated the advantage of performing the pathway-based analysis that integrates biological knowledge and biostatistics approaches, thus forming a useful and important tool for analyzing data from existing GWASs and the forthcoming whole genome sequence association studies (52,53).

In summary, we performed a first pilot study to describe the genome-wide gene–asbestos exposure interaction on lung cancer risk at levels of SNP, gene and pathway. Our findings suggest that the current designed GWASs may have limited power to detect gene–environment interactions at the SNP and gene levels, but the pathway-based approach may be more powerful in performing genome-wide gene–environment analyses.

Supplementary material

Supplementary Tables I–III can be found at http://carcin.oxfordjournals.org/

References

1. Kamp,D.W. (2009) Asbestos-induced lung diseases: an update. *Transl. Res.*, **153**, 143–152.
2. LaDou,J. (2004) The asbestos cancer epidemic. *Environ. Health Perspect.*, **112**, 285–290.
3. Lin,R.T. *et al.* (2007) Ecological association between asbestos-related diseases and historical asbestos consumption: an international analysis. *Lancet*, **369**, 844–849.
4. Nymark,P. *et al.* (2008) Molecular and genetic changes in asbestos-related lung cancer. *Cancer Lett.*, **265**, 1–15.
5. Heintz,N.H. *et al.* (2010) Asbestos, lung cancers, and mesotheliomas: from molecular approaches to targeting tumor survival pathways. *Am. J. Respir. Cell Mol. Biol.*, **42**, 133–139.
6. Schabath,M.B. *et al.* (2002) Association between asbestos exposure, cigarette smoking, myeloperoxidase (MPO) genotypes, and lung cancer risk. *Am. J. Ind. Med.*, **42**, 29–37.
7. Neri,M. *et al.* (2008) Genetic susceptibility to malignant pleural mesothelioma and other asbestos-associated diseases. *Mutat. Res.*, **659**, 126–136.
8. Hiraku,Y. *et al.* (2010) The role of iNOS-mediated DNA damage in infection- and asbestos-induced carcinogenesis. *Ann. N. Y. Acad. Sci.*, **1203**, 15–22.

9. Wang,Y. *et al.* (2008) Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.*, **40**, 1407–1409.

10. Amos,C.I. *et al.* (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.*, **40**, 616–622.

11. Hindorff,L.A. *et al.* (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 9362–9367.

12. Manolio,T.A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.

13. Vineis,P. *et al.* (2010) Missing heritability in genome-wide association study research. *Nat. Rev. Genet.*, **11**, 589.

14. Engelman,C.D. *et al.* (2009) Detecting gene-environment interactions in genome-wide association data. *Genet. Epidemiol.*, **33** (suppl 1), S68–S73.

15. Peng,G. *et al.* (2010) Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur. J. Hum. Genet.*, **18**, 111–117.

16. Wang,K. *et al.* (2010) Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.*, **11**, 843–854.

17. Menashe,I. *et al.* (2010) Pathway analysis of breast cancer genome-wide association study highlights three pathways and one canonical signaling cascade. *Cancer Res.*, **70**, 4453–4459.

18. Luo,L. *et al.* (2010) Genome-wide gene and pathway analysis. *Eur. J. Hum. Genet.*, **18**, 1045–1053.

19. Purcell,S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.

20. Liu,J.Z. *et al.* (2010) A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.*, **87**, 139–145.

21. Zhang,K. *et al.* (2010) i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res.*, **38**(suppl.), W90–W95.

22. Subramanian,A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15545–15550.

23. Kanehisa,M. *et al.* (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

24. Science Signaling. Available from http://stke.sciencemag.org/.

25. BioCarta. Available from: http://www.biocarta.com/genes/allpathways.asp.

26. Ashburner,M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

27. Wang,K. *et al.* (2007) Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.*, **81**, 1278–1283.

28. Harald Wajant Fas Signaling Pathway. Science Signaling.*Connections Map in the Database of Cell Signaling*. http://stke.sciencemag.org/cgi/cm/stkecm;CMP_7966 (6 February 2011).

29. Murcray,C.E. *et al.* (2009) Gene-environment interaction in genome-wide association studies. *Am. J. Epidemiol.*, **169**, 219–226.

30. Arya,R. *et al.* (2009) Effects of covariates and interactions on a genome-wide association analysis of rheumatoid arthritis. *BMC Proc.*, **3** (suppl 7), S84.

31. Neale,B.M. *et al.* (2004) The future of association studies: gene-based analysis and replication. *Am. J. Hum. Genet.*, **75**, 353–362.

32. Wu,M. *et al.* (2006) LRRC4, a putative tumor suppressor gene, requires a functional leucine-rich repeat cassette domain to inhibit proliferation of glioma cells *in vitro* by modulating the extracellular signal-regulated kinase/protein kinase B/nuclear factor-kappaB pathway. *Mol. Biol. Cell*, **17**, 3534–3542.

33. Li,C.L. *et al.* (2008) Structural and functional insights into human Tudor-SN, a key component linking RNA interference and editing. *Nucleic Acids Res.*, **36**, 3579–3589.

34. WTCCC (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.

35. Vogels,M.W. *et al.* (2011) Identification of host factors involved in coronavirus replication by quantitative proteomics analysis. *Proteomics*, **11**, 64–80.

36. Franke,A. *et al.* (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.*, **42**, 1118–1125.

37. Voight,B.F. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.

38. Eeles,R.A., *et al.* (2009) Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet*, **41**, 1116–1121.

39. Krammer,P.H. (2000) CD95's deadly mission in the immune system. *Nature*, **407**, 789–795.

40. Pluygers,E. *et al.* (2001) The impact of immune responses on lung cancer and the development of new treatment modalities. *Lung Cancer*, **34** (suppl 2), S71–S77.

41. Strasser,A. *et al.* (2009) The many roles of FAS receptor signaling in the immune system. *Immunity*, **30**, 180–192.

42. Ter-Minassian,M. *et al.* (2008) Apoptosis gene polymorphisms, age, smoking and the risk of non-small cell lung cancer. *Carcinogenesis*, **29**, 2147–2152.

43. Altomare,D.A. *et al.* (2009) Activated TNF-alpha/NF-kappaB signaling via down-regulation of Fas-associated factor 1 in asbestos-induced mesotheliomas from Arf knockout mice. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 3420–3425.

44. Yang,H. *et al.* (2006) TNF-alpha inhibits asbestos-induced cytotoxicity via a NF-kappaB-dependent pathway, a possible mechanism for asbestos-induced oncogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 10397–10402.

45. Matsuzaki,S. *et al.* (2010) Analysis on the co-localization of asbestos bodies and Fas or CD163 expression in asbestos lung tissue by in-air micro-pixe. *Int. J. Immunopathol. Pharmacol.*, **23**, 1–11.

46. Horton,R. *et al.* (2004) Gene map of the extended human MHC. *Nat. Rev. Genet.*, **5**, 889–899.

47. Vyas,J.M. *et al.* (2008) The known unknowns of antigen processing and presentation. *Nat. Rev. Immunol.*, **8**, 607–618.

48. Otsuki, T. *et al.* (2007) Immunological effects of silica and asbestos. *Cell. Mol. Immunol.*, **4**, 261–268.

49. Maeda,M. *et al.* (2010) Dysregulation of the immune system caused by silica and asbestos. *J. Immunotoxicol.*, **7**, 268–278.

50. Nishimura,Y. *et al.* (2006) Expression of the T cell receptor Vbeta repertoire in a human T cell resistant to asbestos-induced apoptosis and peripheral blood T cells from patients with silica and asbestos-related diseases. *Int. J. Immunopathol. Pharmacol.*, **19**, 795–805.

51. Nishimura,Y. *et al.* (2009) Impairment in cytotoxicity and expression of NK cell- activating receptors on human NK cells following exposure to asbestos fibers. *Int. J. Immunopathol. Pharmacol.*, **22**, 579–590.

52. Thomas,D.C. *et al.* (2009) Use of pathway information in molecular epidemiology. *Hum. Genomics*, **4**, 21–42.

53. Moore,J.H. *et al.* (2010) Bioinformatics challenges for genome-wide association studies. *Bioinformatics*, **26**, 445–455.