Carcinogenesis vol.29 no.10 pp.1938–1943, 2008
doi:10.1093/carcin/bgn178
Advance Access publication August 1, 2008

Pathway-based evaluation of 380 candidate genes and lung cancer susceptibility suggests the importance of the cell cycle pathway

H.Dean HosgoodIII[1,2,*], Idan Menashe[1], Min Shen[1], Meredith Yeager[1], Jeff Yuenger[1], Preetha Rajaraman[1], Xingzhou He[3], Nilanjan Chatterjee[1], Neil E.Caporaso[1], Yong Zhu[2], Stephen J.Chanock[1], Tongzhang Zheng[2] and Qing Lan[1]

[1]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA, [2]Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520, USA and [3]Chinese Center for Disease Control and Prevention, Beijing, China

[*]To whom correspondence should be addressed. Tel: +1 301 594 4649;
Fax: +1 301 402 1819;
Email: hosgoodd@mail.nih.gov

Common genetic variation may play an important role in altering lung cancer risk. We conducted a pathway-based candidate gene evaluation to identify genetic variations that may be associated with lung cancer in a population-based case–control study in Xuan Wei, China (122 cases and 111 controls). A total of 1260 single-nucleotide polymorphisms (SNPs) in 380 candidate genes for lung cancer were successfully genotyped and assigned to one of 10 pathways based on gene ontology. Logistic regression was used to assess the marginal effect of each SNP on lung cancer susceptibility. The minP test was used to identify statistically significant associations at the gene level. Important pathways were identified using a test of proportions and the rank truncated product methods. The cell cycle pathway was found as the most important pathway ($P = 0.044$) with four genes significantly associated with lung cancer (*PLA2G6* minP = 0.001, *CCNA2* minP = 0.006, *GSK3*β minP = 0.007 and *EGF* minP = 0.013), after adjusting for multiple comparisons. Interestingly, most cell cycle genes that were associated with lung cancer in this analysis were concentrated in the AKT signaling pathway, which is essential for regulation of cell cycle progression and cellular survival, and may be important in lung cancer etiology in Xuan Wei. These results should be viewed as exploratory until they are replicated in a larger study.

Introduction

Lung cancer is estimated to account for 1.4 million cancer cases and 1.2 million cancer deaths per year in the world (1). In 2000, it was estimated that 85% of lung cancer in men and 47% of lung cancer in women were attributed to tobacco smoking (1). While smoking is the primary risk factor for lung cancer, other environmental, occupational and genetic risk factors have been documented in certain populations (2). Due to the overwhelming risk associated with tobacco smoking, however, other lung cancer risk factors have not been fully elucidated.

Xuan Wei poses a unique opportunity to assess lung cancer susceptibility in a population with substantial in-home coal smoke exposure, a classified group 1 human carcinogen by the International Agency for Research on Cancer (3). Xuan Wei has the highest prevalence of lung cancer in China (4). The age-adjusted lung cancer mortality rates for men and women in Xuan Wei are 27.7 and 25.3 per 100 000, respectively (5). The similarity of lung cancer rates in men and women is of considerable interest because almost all women are non-smokers (6). In Xuan Wei, nearly all women and few men cook, whereas most men and nearly no women smoke tobacco (6). The primary source of indoor air pollution in Xuan Wei is smoke from domestic fuel combustion for heating and cooking with most residents burning smoky coal (bituminous coal) and some using smokeless coal (anthracite coal). Smoky coal use in Xuan Wei homes is associated with very high and comparable risks of lung cancer in both men and women (7,8).

Since the carcinogenic constituent of smoky coal combustion is polycyclic aromatic hydrocarbons (5,9), initial lung cancer susceptibility studies in Xuan Wei focused on individual single-nucleotide polymorphisms (SNPs) in candidate genes associated with polycyclic aromatic hydrocarbon metabolism (10,11). Subsequent studies focused on important biological pathways, such as DNA repair (12). While these studies have provided some promising results, a large-scale candidate gene analysis has not been performed to evaluate genetic susceptibility to lung cancer in Xuan Wei. Therefore, we analyzed 1260 SNPs in 380 candidate genes using an Oligo Pool with an Illumina® GoldenGate Assay. Candidate SNPs were selected from the SNP500Cancer database and genotyped if they were potentially relevant for cancer or other human diseases, had possible functional significance or expanded gene coverage of previously identified candidate genes. We hypothesized that this large-scale candidate gene study would provide insight into the pathways important to lung cancer susceptibility.

Methods

The study population of this population-based case–control study has been described previously (10). Briefly, all residents of Xuan Wei, China, from March 1995 to March 1996 were eligible for inclusion. Lung cancer cases with clinical symptoms and X-ray confirmation were identified at one of five hospitals servicing Xuan Wei County. Of the 135 eligible cases, 133 (99%) agreed to participate. To be enrolled, cases had to be histologically ($n = 14$) or cytologically ($n = 91$) confirmed or have died within 1 year of diagnosis ($n = 17$), since previous studies in Xuan Wei suggest that death within 1 year of clinical diagnosis of lung cancer is a strong indicator of lung cancer diagnosis (13). Based on these criteria, 122 of 133 consenting cases (92%) were enrolled into the study.

Controls were selected from the Xuan Wei general population and were individually matched by sex, age (±2 years), village and type of fuel used for in-home cooking and heating at time of interview. The participation rate for controls was 100%. A detailed questionnaire evaluating smoking history, domestic fuel use history and other demographic information was administered by trained interviewers to cases and controls. This research protocol was approved by a United States Environmental Protection Agency Human Subjects Research Review Official, and informed consent was obtained from all study subjects.

Genotyping was performed on DNA extracted from sputum samples via phenol–chloroform extraction (14). Candidate SNPs were identified through the SNP500Cancer database (http://snp500cancer.nci.nih.gov/) and genotyped if they were potentially relevant for cancer or other human diseases, had possible functional significance or expanded gene coverage of previously identified candidate genes. High-throughput genotyping was successful for 122 (100%) cases and 111 (91%) controls with an Oligo Pool by the Illumina GoldenGate Assay (http://www.illumina.com) at the National Cancer Institute's Core Genotyping Facility (Gaithersburg, MD). Ten controls did not have ample DNA for genotyping. Duplicate samples ($n = 21$) of both cases and controls were randomly distributed throughout study plates to ensure quality control and determine the intra-subject concordance rate for all assays (>98%). Initially, 1442 SNPs were genotyped. Hardy–Weinberg equilibrium for each SNP was tested in controls with a Pearson χ^2 test or a Fisher's exact test if any of the cell counts were less than five. After exclusion of 166 SNPs with low minor allele frequency (<0.01) and 16 SNPs with substantial deviation from Hardy–Weinberg equilibrium ($P < 0.001$), 1260 SNPs in 380 genes were left for analysis.

First, unconditional logistic regression was used to estimate the odds ratio and calculate the 95% confidence interval for the association between lung cancer risk independently for each SNP, using the homozygote of the common allele as the reference group and adjusting for age (<55 and ≥55 years), sex, smoking (0 pack years, >0 and <25 pack years and ≥25 pack years) and

Abbreviations: AKT, v-akt murine thymoma viral oncogene homolog 1; FDR, false discovery rate; LD, linkage disequilibrium; SNP, single-nucleotide polymorphism.

Table I. Study participant characteristics of Xuan Wei case–control study

	Cases		Controls		*P*-value[a]
	n	%	*n*	%	
Gender					0.85
Male	77	67.5	71	66.4	
Female	37	32.5	36	33.6	
Age (years)					0.78
<55	55	45.1	48	43.2	
≥55	67	54.9	63	56.8	
Smoking (pack years)					0.36
Non-smokers	53	43.4	49	44.1	
<25	27	22.1	32	28.8	
≥25	42	34.4	30	27.0	
Lifetime smoky coal use (tons)					0.02
<130	57	46.7	69	62.2	
≥130	65	53.3	42	37.8	

[a] χ^2 test.

Table II. Pathway-based evaluation of gene-based minP and most significant SNP for all significant candidate genes (minP ≤ 0.05) associated with lung cancer

Pathway	Genes	#SNPs	*P*-value for most significant SNP	Permutation *P*-value for gene (minP)[a]	*P*-value for pathway[b]
Cell cycle	49	184			0.044
	PLA2G6	4	0.0003	0.001	
	CCNA2	3	0.0035	0.006	
	GSK3B	34	0.0004	0.007	
	AKT1	1	0.0104	0.010	
	EGF	3	0.0023	0.013	
	TP53I3	5	0.0052	0.017	
	PTEN	2	0.0169	0.018	
	MYBL2	8	0.0103	0.033	
	CCND3	2	0.0225	0.038	

[a] Adjusted for age (<55 and ≥55 years), sex, smoking (0 pack years, >0 and <25 pack years and ≥25 pack years) and lifetime smoky coal exposure (<130 and ≥130 tons).
[b] Proportions test comparing the number of genes in each pathway with a minP ≤0.05 and those >0.05 to the number of genes in all other pathways with a minP ≤0.05 and those >0.05.

lifetime smoky coal exposure (<130 and ≥130 tons). Gene–dose effects for each SNP were estimated by a linear trend test by coding the genotypes based on the number of variant alleles present (0, 1 and 2). Interactions between the dominant model and lifetime smoky coal exposure were tested on the multiplicative scale for significant SNPs in the four significant cell cycle genes while adjusting for age, sex and smoking.

Second, gene-based analyses were performed on 380 genes. To assess the significance of the association between each gene and lung cancer, we used MatLab to perform a minP test that assesses the significance of the minimal *P*-value in each gene using a permutation-based resampling procedure (1000 permutations) that takes into account the number of SNPs genotyped in each gene and their underlying linkage disequilibrium (LD) structure (15). A gene was significantly associated with lung cancer if it had a minP ≤0.05, after adjustment for age (<55 and ≥55 years), sex, smoking (0 pack years, >0 and <25 pack years and ≥25 pack years) and lifetime smoky coal exposure (<130 and ≥130 tons). False discovery rates (FDRs) were calculated using the Benjamini–Hochberg method to evaluate the significance of the minP results within the cell cycle pathway (16).

Third, haplotype blocks and structure were determined with Haploview using data from controls for the four significant genes in the cell cycle pathway with more than one SNP (17). Haplotype frequencies were estimated using the expectation–maximization algorithm (18). Haplotypes with frequencies <1% were excluded. The overall difference in haplotype frequencies between cases and controls was assessed using a global score test (19). Haplotype odds ratios and 95% confidence intervals were calculated and adjusted for age

Table III. Risk of lung cancer associated with individual SNPs of significant cell cycle genes

Gene	SNP	Alleles	Heterozygotes			Homozygotes			Dominant model			Trend		
			OR[a]	95% CI[a]	*P*	OR[a]	95% CI[a]	*P*	OR[a]	95% CI[a]	*P*	OR[a]	95% CI[a]	*P*
CCNA2—cyclin A2	rs3217773	G > A	2.26	1.19–4.29	**0.0130**	Not applicable			2.40	1.27–4.54	**0.0070**	2.42	1.31–4.48	**0.0049**
EGF—epidermal growth factor	rs2237051	A > G	0.55	0.30–0.99	**0.0480**	0.26	0.10–0.68	**0.0061**	0.47	0.27–0.83	**0.0095**	0.52	0.34–0.80	**0.0029**
GSK3B—glycogen synthase kinase 3 beta	rs6781942	A > G	1.62	0.88–3.00	0.1221	5.34	2.07–13.79	**0.0005**	2.11	1.18–3.79	**0.0122**	2.10	1.37–3.22	**0.0007**
PLA2G6—phospholipase A2, group VI	rs84473	A > G	1.87	0.98–3.58	0.0590	5.13	2.07–12.71	**0.0004**	2.37	1.27–4.41	**0.0064**	2.19	1.42–3.38	**0.0004**

Values in bold signify $P \leq 0.05$.
[a] Odds ratios (ORs) and 95% confidence intervals (CIs) are adjusted for age (<55 and ≥55 years), sex, smoking (0 pack years, >0 and <25 pack years and ≥25 pack years) and lifetime smoky coal exposure (<130 and ≥130 tons).

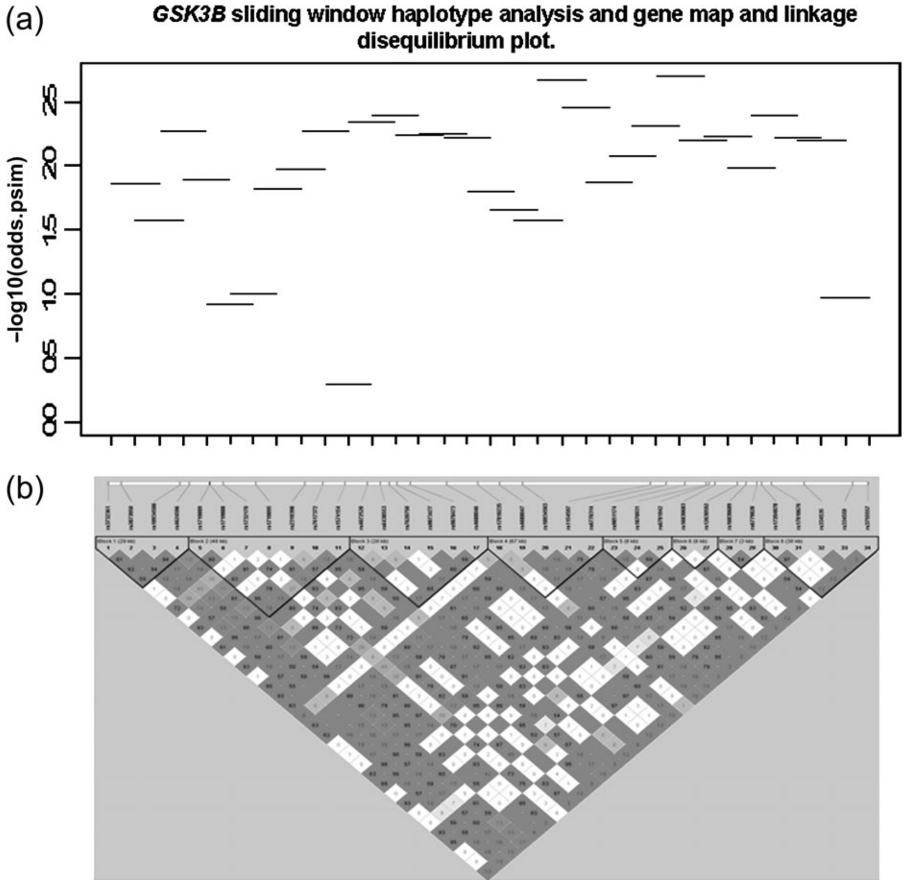

Fig. 1. *GSK3B* sliding window haplotype analysis and gene map and LD blot. (**a**) Sliding window: *y*-axis is odds of lung cancer associated with each 3-SNP window. *X*-axis is the corresponding SNP number seen in the gene map. (**b**) Gene map and LD plot: color scheme is based on D' and the logarithm of the odds of linkage (LOD) values (white: $D' < 1$ and LOD < 2, blue: $D' = 1$ and LOD < 2, pink/red: $D' < 1$ and LOD ≥ 2 and bright red: $D'1 = 1$and LOD ≥ 2). Numbers in squares are $D'1$ values (1.0 not shown).

(<55 and ≥55 years), sex, smoking (0 pack years, >0 and <25 pack years and ≥25 pack years) and lifetime smoky coal exposure (<130 and ≥130 tons). A sliding window 3-SNP haplotype approach was also performed for the two significant genes in the cell cycle pathway with more than three SNPs to comprehensively evaluate potential disease loci in small genetic regions that may have been overlooked with the single-locus analysis (20).

Finally, pathway-based analysis was performed on all 1260 SNPs available for analysis. Genes were categorized into biological pathways using the Go-Miner software (http://discover.nci.nih.gov/gominer/), which utilizes the Gene Ontology database (http://www.geneontology.org) to identify the biological processes and functions of the genes and classify them into biologically coherent categories. To test the significance of each pathway, the proportion of statistically significant genes versus non-significant genes for each pathway was compared with the proportion of significant genes versus non-significant genes in all the remaining pathways using the one-sample test for proportions (21). Exact methods were used when cell counts were less than five. In addition, we used the rank truncated product method to evaluate the excess of highly significant SNPs within each pathway (22). Since the rank truncated product test yielded similar results as the one-sample test for proportions, only the one-sample test for proportions results are reported.

All statistical methods were performed using SAS software, version 9.1 (SAS Institute, Cary, NC), unless stated elsewhere.

Results

Cases and controls were comparable in age, sex and smoking status (Table I). As in previous reports of this study population (10), cases tended to use significantly more in-home smoky coal over the course of their lifetimes than controls ($P = 0.02$).

A total of 1260 functional or likely functional SNPs in 380 genes were successfully genotyped. Pathway-based analysis categorized the 380 genes into 10 pathways (supplementary Table 1 is available at *Carcinogenesis* Online): cell cycle genes (including apoptosis) ($n = 49$), DNA repair genes ($n = 49$), telomere maintenance genes ($n = 5$), immune response genes ($n = 63$), molecular transport genes ($n = 27$), signal transduction genes ($n = 35$), one-carbon metabolism genes ($n = 11$), xenobiotic metabolism genes ($n = 29$), other metabolism genes ($n = 57$) and other uncategorized genes ($n = 55$). While all pathways had at least one significant gene, including polycyclic aromatic hydrocarbon metabolism-related *EPHX1* and *GSTM3* in the xenobiotic pathway and telomere maintenance genes *TERT* and *TERF2* (supplementary Table 1 is available at *Carcinogenesis* Online), only the cell cycle pathway had a significantly increased proportion of significant genes ($P = 0.044$) (Table II).

Gene-based analysis identified 9 of 49 genes as significantly associated with lung cancer risk (*PLA2G6* minP = 0.001, *CCNA2* minP = 0.006, *GSK3*β minP = 0.007, v-akt murine thymoma viral oncogene homolog 1 (AKT) minP = 0.010, *EGF* minP = 0.013, *TP53I3* minP = 0.017, *PTEN* minP = 0.018, *MYBL2* minP = 0.033 and *CCND3* minP = 0.038) in the cell cycle pathway (Table II). When adjusting for the number of tested genes, only *PLA2G6* (FDR = 0.0098), *GSK3*β (FDR = 0.0098), *EGF* (FDR = 0.0376) and *CCNA2* (FDR = 0.0426) remained significant.

Individual SNP analyses, using logistic regression, found 24 of 44 individual SNPs genotyped in the four significant genes of the cell cycle pathway to be associated with lung cancer risk ($P_{\text{trend}} \leq 0.05$). After accounting for SNPs in high LD ($D' \geq 0.97$) and highly correlated ($r^2 \geq 0.90$) within each gene, only one SNP in each gene was identified as the important SNP associated with lung cancer risk (Table III). Variant carriers of *CCNA2* rs3217773, *GSK3*β rs6781942 and *PLA2G6*

Table IV. Glycogen synthase kinase 3 beta haplotypes and risk of lung cancer[a]

Block[b]	Haplotype	Cases	Controls	OR[a]	95% CI[a]	Global test omnibus[a]	Order of SNPs
1	GAAA	109	123	1.00	Reference	**0.0100**	rs3732361 (G > A), rs2873950 (A > C), rs10934500 (A > G) and rs4624596 (A > G)
	ACGG	106	71	1.89	1.24–2.89		
	GCGG	25	24	1.05	0.54–2.03		
2	TGAAAAA	96	109	1.00	Reference	**0.0150**	rs1719889 (T > A), rs1719888 (G > A), rs1732170 (A > G), rs1719895 (A > G), rs2319398 (A > C), rs7617372 (A > G) and rs1574154 (A > G)
	TGGACAA	109	71	1.88	1.22–2.88		
	AAAGAGG	21	22	0.99	0.49–2.01		
	TGGAAAA	12	14	0.57	0.21–1.51		
	AGAAAGG	2	4	0.54	0.09–3.40		
3	ATGAGA	110	124	1.00	Reference	**0.0074**	rs4072520 (A > C), rs6438553 (T > A), rs7620750 (G > A), rs9873477 (A > G), rs9878473 (G > A) and rs4688046 (A > G)
	AAGGAG	107	70	1.89	1.24–2.88		
	CTAAGG	24	28	0.99	0.52–1.90		
4	AAAAA	108	123	1.00	Reference	**0.0157**	rs17810235 (A > G), rs4688047 (A > G), rs10934503 (A > G), rs1154597 (A > G) and rs6770314 (A > G)
	AGGAA	91	59	1.89	1.22–2.92		
	AGGGG	23	24	1.06	0.54–2.10		
	GGGAA	16	10	2.68	1.04–6.87		
	AGGAG	2	4	0.46	0.06–3.45		
5	GGA	108	124	1.00	Reference	**0.0032**	rs9851174 (G > A), rs1870931(G > C) and rs6781942 (A > G)
	GCG	108	69	2.03	1.32–3.12		
	ACA	25	27	1.03	0.53–1.97		
6	GA	112	127	1.00	Reference	**0.0005**	rs16830683 (G > A) and rs12630592 (A > C)
	GC	110	69	2.05	1.35–3.13		
7	CA	207	176	1.00	Reference	0.4718	rs16830689 (C > G) and rs6779828 (A > G)
	GG	25	28	0.79	0.42–1.50		
8	AGGAT	108	70	1.00	Reference	**0.0101**	rs17204878 (C > A), rs17810676 (A > G), rs334535 (G > A), rs334559 (A > G) and rs3755557 (T > A)
	CAGAT	76	78	0.57	0.36–0.91		
	CAGAA	35	45	0.43	0.25–0.77		
	CAAGT	23	27	0.51	0.25–1.04		

Values in bold signify $P \leq 0.05$.

[a] Odds ratios (ORs), 95% confidence intervals (CIs) and omnibus tests are adjusted for age (<55 and ≥55 years), sex, smoking (0 pack years, >0 and <25 pack years and ≥25 pack years) and lifetime smoky coal exposure (<130 and ≥130 tons).

[b] Corresponds to blocks in Figure 1.

rs84473 were associated with a significantly increased risk of lung cancer ($P_{\text{trend}} \leq 0.05$), whereas variant carriers of *EGF* rs2237051 were associated with a significantly decreased risk of lung cancer ($P_{\text{trend}} \leq 0.05$). The magnitude and direction of risk associated with *EGF* rs2237051, *CCNA2* rs3217773, *GSK3*β rs6781942 and *PLA2G6* rs84473 and lung cancer were similar between men and women (data not shown). Lifetime smoky coal use did not interact significantly with any significant SNP in *PLA2G6*, *CCNA2*, *GSK3*β and *EGF* (data not shown). Supplementary Table 2 (available at *Carcinogenesis* Online) provides lung cancer risk associated with all genotyped SNPs.

Haplotype analysis of *GSK3*β was based on 34 SNPs that covered 61 of 75 SNPs in this gene (Figure 1). Of the eight blocks defined by the LD in the controls, the seven blocks significantly associated with lung cancer risk were consistent with the individual SNP results (Table IV). Sliding window analysis identified the *GSK3*β genomic regions of blocks 4 and 5 to have a slightly more increased risk of lung cancer than the other genomic regions (Figure 1), identifying *GSK3*β rs6781942 as important, similar to individual SNP analyses. Haplotype analyses for *PLA2G6*, *EGF* and *CCNA2* did not provide any additional information beyond individual SNP analyses.

Discussion

Through an exploratory analysis of 10 different biological pathways that are potentially important to carcinogenic processes, only the cell cycle pathway had a significantly increased proportion of significant genes compared with all other pathways. Gene-based analyses identified nine cell cycle genes that were significantly associated with lung cancer susceptibility: *AKT1*, *CCNA2*, *CCND3*, *EGF*, *GSK3*β, *MYBL2*, *PLA2G6*, *PTEN* and *TP53I3*. Only four genes remained significant after adjustment for multiple comparisons and individual SNP analyses identified the most important variant of each gene (*CCNA2* rs3217773, *EGF* rs2237051, *GSK3*β rs6781942 and *PLA2G6* rs84473). Of the four genes identified in the cell cycle pathway, three (*CCNA2*, *EGF* and *GSK3*β) are closely interconnected through the AKT signaling pathway, which is an important regulator of apoptosis and is essential to help cells manage apoptotic stimuli, by regulating cell cycle progression and cellular survival (23). Similar to our findings, a recent study found cell cycle genes to be important in the expression signature of smoking-related lung cancers (24). AKT-dependent apoptosis and cell cycle disruption are triggered by the binding of epidermal growth factor to EGFR on the cell surface (25). The activation of AKT leads to phosphorylation of GSK3β, which inhibits cyclin D, a regulator of entry into the S phase of the cell cycle (26).

The importance of AKT in the regulation of apoptosis and the cell cycle makes its expression essential to homeostasis. The dysregulation of AKT has been seen in many cancers, including lung cancer (27). AKT-dependent apoptosis and cell cycle disruption are triggered by extracellular growth factors that activate AKT (28), such as the binding of epidermal growth factor to EGFR on the cell surface (25). Similar to AKT, epidermal growth factor has also been shown to be upregulated in lung cancer tumors (29,30). In our study, *EGF* rs2237051 was associated with decreased risk of lung cancer risk. Although this particular variant has not been reported previously to be associated with lung cancer, it was in LD ($D' = 0.86$) with *EGF* rs4444903, which has been associated with increased risk of lung cancer in one Korean population (31), but not another (32).

The activation of AKT leads to the phosphorylation of GSK3β, which inhibits cyclin D (23). Recently, it has been hypothesized that EGFR, and not just AKT, might also be involved with the phosphorylation of GSK3β lung adenocarcinomas (33). Decreased GSK3β expression has been seen in bronchial and tracheobronchial epithelial cells exposed to cigarette smoke (34,35). Whereas GSK3β has been strongly implicated as an important etiologic factor of colorectal and pancreatic cancer (36,37), mutations in the *GSK3*β have not been reported previously in lung cancer. We observed a significant gene-based association between *GSK3*β and risk of lung cancer. Haplotype analyses identified a genomic region, which included rs6781942, with an increased risk of lung cancer. *GSK3*β rs6781942 was significantly associated with an increased risk of lung cancer in homozygote variant carriers. The strength and consistent associations between lung cancer and *GSK3*β found in our study warrant further investigation.

Finally, at the heart of the cell cycle control system and the end of the AKT pathway is a family of protein kinases known as cyclin-dependent kinases. Changes in cyclin levels result in activation of cyclin–cdk complexes, triggering cell cycle events (38). Cyclin D, which is inhibited by GSK3β, regulates entry into the S phase of the cell cycle, where it is essential for cyclin A–Cdk2 to phosphorylate E2F and inhibit its bindings to DNA, thus inactivating its function as a transcription factor (39). Variant carriers of *CCNA2* (rs3217773) were associated with increased risk of lung cancer in our study.

One of the major strengths of our population-based case–control study is the high participation rate. The moderate sample size, on the other hand, may lead to false-positive and false-negative findings (40). Therefore, our findings should be viewed as hypothesis generating until they are replicated in a larger study. We accounted for possible spurious findings due to multiple comparisons by using a permutation method for the gene-based analyses and then evaluating FDRs. The use of a gene-based permutation analysis identifies genes significantly associated with disease status by comparing the observed association with the distribution of gene–disease associations seen in 1000 randomly generated populations. This robust identification allows for sequential honing of important genomic regions and subsequently SNPs associated with lung cancer. Although functionality is not known for all genotyped SNPs, our results are biologically plausible given that variants in cell cycle pathway genes could contribute to lung cancer risk. However, associations with any specific SNP should be cautiously interpreted until these results are replicated, and functionality is determined, especially given that associations with a particular SNP in this study may be attributed to another SNP in LD.

In summary, our findings provide evidence of genetic variation that may be important to lung cancer susceptibility. Our results implicate the cell cycle pathway, particularly the *CCNA2*, *EGF*, *GSK3*β and *PLA2G6* genes. The strongest findings in our study were disproportionally concentrated in the biologically interconnected AKT signaling pathway that regulates cell cycle and apoptosis. Our results should be viewed as exploratory until they are replicated in larger studies with more substantial genomic coverage.

Funding

Intramural National Cancer Institute (N01-CO-12400) program; Yale University–National Cancer Institute Partnership Predoctoral Fellowship Training Program (NCI TU2 CA105666).

Supplementary material

Supplementary Tables 1 and 2 can be found at http://carcin.oxfordjournals.org/

Acknowledgements

Conflict of Interest Statement: None declared.

References

1. Parkin,D.M. *et al.* (2005) Global cancer statistics, 2002. *CA Cancer J. Clin.*, **55**, 74–108.
2. Pass,H.I. (2005) *Lung Cancer: Principles and Practice*. Lippincott Williams & Wilkins, Philadelphia, PA.
3. Straif,K. *et al.* (2006) Carcinogenicity of household solid fuel combustion and of high-temperature frying. *Lancet Oncol.*, **7**, 977–978.
4. Lan,Q. *et al.* (2004) Molecular epidemiological studies on the relationship between indoor coal burning and lung cancer in Xuan Wei, China. *Toxicology*, **198**, 301–305.
5. Mumford,J.L. *et al.* (1987) Lung cancer and indoor air pollution in Xuan Wei, China. *Science*, **235**, 217–220.
6. Chapman,R.S. *et al.* (1988) The epidemiology of lung cancer in Xuan Wei, China: current progress, issues, and research strategies. *Arch. Environ. Health.*, **43**, 180–185.
7. Lan,Q. *et al.* (2002) Household stove improvement and risk of lung cancer in Xuanwei, China. *J. Natl Cancer Inst.*, **94**, 826–835.
8. Lan,Q. *et al.* (2008) Variation in lung cancer risk by smoky coal subtype in Xuanwei, China. *Int. J. Epidemiol.* in press.
9. Mumford,J.L. *et al.* (1995) Human exposure and dosimetry of polycyclic aromatic hydrocarbons in urine from Xuan Wei, China with high lung cancer mortality associated with exposure to unvented coal smoke. *Carcinogenesis*, **16**, 3031–3036.
10. Lan,Q. *et al.* (2000) Indoor coal combustion emissions, GSTM1 and GSTT1 genotypes, and lung cancer risk: a case-control study in Xuan Wei, China. *Cancer Epidemiol. Biomarkers Prev.*, **9**, 605–608.
11. Lan,Q. *et al.* (2004) Oxidative damage-related genes AKR1C3 and OGG1 modulate risks for lung cancer due to exposure to PAH-rich coal combustion emissions. *Carcinogenesis*, **25**, 2177–2181.
12. Shen,M. *et al.* (2005) Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. *Int. J. Cancer.*, **116**, 768–773.
13. Hu,F. *et al.* (1989) Natural survival of lung cancer observed in the general investigation in Xuan Wei county. *Chin. J. Cancer.*, **8**, 42–44. [In Chinese].
14. Garcia-Closas,M. *et al.* (2001) Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 687–696.
15. Chen,B.E. *et al.* (2006) Resampling-based multiple hypothesis testing procedures for genetic case-control association studies. *Genet. Epidemiol.*, **30**, 495–507.
16. Benjamini,Y. *et al.* (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B.*, **57**, 289–300.
17. Barrett,J.C. *et al.* (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
18. Excoffier,L. *et al.* (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.*, **12**, 921–927.
19. Schaid,D.J. *et al.* (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am. J. Hum. Genet.*, **70**, 425–434.
20. Huang,B.E. *et al.* (2007) Detecting haplotype effects in genomewide association studies. *Genet. Epidemiol.*, **31**, 803–812.
21. Gauvreau,K. (2006) Hypothesis testing: proportions. *Circulation*, **114**, 1545–1548.
22. Dudbridge,F. *et al.* (2003) Rank truncated product of *P*-values, with application to genomewide association scans. *Genet. Epidemiol.*, **25**, 360–366.
23. Duronio,V. *et al.* (1998) Downstream signalling events regulated by phosphatidylinositol 3-kinase activity. *Cell. Signal.*, **10**, 233–239.
24. Landi,M.T. *et al.* (2008) Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS ONE*, **3**, e1651.
25. Laurence,D.J. *et al.* (1990) The epidermal growth factor. A review of structural and functional relationships in the normal organism and in cancer cells. *Tumour Biol.*, **11**, 229–261.
26. Kandel,E.S. *et al.* (1999) The regulation and activities of the multifunctional serine/threonine kinase Akt/PKB. *Exp. Cell Res.*, **253**, 210–229.
27. Lee,S.H. *et al.* (2002) Non-small cell lung cancers frequently express phosphorylated Akt; an immunohistochemical study. *APMIS*, **110**, 587–592.
28. Franke,T.F. *et al.* (2003) PI3K/Akt and apoptosis: size matters. *Oncogene*, **22**, 8983–8998.
29. Tateishi,M. *et al.* (1990) Immunohistochemical evidence of autocrine growth factors in adenocarcinoma of the human lung. *Cancer Res.*, **50**, 7077–7080.
30. Gorgoulis,V. *et al.* (1992) Expression of EGF, TGF-alpha and EGFR in squamous cell lung carcinomas. *Anticancer Res.*, **12**, 1183–1187.
31. Lim,Y.J. *et al.* (2005) Epidermal growth factor gene polymorphism is different between schizophrenia and lung cancer patients in Korean population. *Neurosci. Lett.*, **374**, 157–160.
32. Kang,H.G. *et al.* (2007) +61A>G polymorphism in the EGF gene does not increase the risk of lung cancer. *Respirology*, **12**, 902–905.

33. Zheng,H. *et al*. (2007) Phosphorylated GSK3beta-ser9 and EGFR are good prognostic factors for lung carcinomas. *Anticancer Res.*, **27**, 3561–3569.
34. Chari,R. *et al*. (2007) Effect of active smoking on the human bronchial epithelium transcriptome. *BMC Genomics*, **8**, 297.
35. Tian,D. *et al*. (2006) Role of glycogen synthase kinase 3 in squamous differentiation induced by cigarette smoke in porcine tracheobronchial epithelial cells. *Food Chem. Toxicol.*, **44**, 1590–1596.
36. Ougolkov,A.V. *et al*. (2006) Aberrant nuclear accumulation of glycogen synthase kinase-3beta in human pancreatic cancer: association with kinase activity and tumor dedifferentiation. *Clin. Cancer Res.*, **12**, 5074–5081.
37. Shakoori,A. *et al*. (2005) Deregulated GSK3beta activity in colorectal cancer: its association with tumor cell survival and proliferation. *Biochem. Biophys. Res. Commun.*, **334**, 1365–1373.
38. Alberts,B. (2002) *Molecular Biology of the Cell*. Garland Science, New York, NY.
39. Hartl,D.L. *et al*. (2002) *Essential Genetics: A Genomics Perspective*. Jones and Bartlett, Boston, MA.
40. Wacholder,S. *et al*. (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J. Natl Cancer Inst.*, **96**, 434–442.

Received June 2, 2008; revised July 20, 2008; accepted July 25, 2008