Received: 2 March 2017 | Revised: 21 July 2017 | Accepted: 29 September 2017
DOI: 10.1002/mc.22748

RESEARCH ARTICLE

WILEY Molecular Carcinogenesis

Novel genetic variants in the P38MAPK pathway gene *ZAK* and susceptibility to lung cancer

Yun Feng[1,2,3] | Yanru Wang[2,3] | Hongliang Liu[2,3] | Zhensheng Liu[2,3] | Coleman Mills[2,3] | Kouros Owzar[2,4] | Jichun Xie[4] | Younghun Han[5] | David C. Qian[5] | Rayjean J. Hung RJ[6] | Yonathan Brhane[6] | John McLaughlin[7] | Paul Brennan[8] | Heike Bickeböller[9] | Albert Rosenberger[9] | Richard S. Houlston[10] | Neil Caporaso[11] | Maria Teresa Landi[11] | Irene Brüske[12] | Angela Risch[13] | Yuanqing Ye[14] | Xifeng Wu[14] | David C. Christiani[15,16] | Christopher I. Amos[5] | Qingyi Wei[2,3,17]

[1] Department of Respiration, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China
[2] Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina
[3] Department of Medicine, Duke University School of Medicine, Durham, North Carolina
[4] Duke Cancer Institute and Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, North Carolina
[5] Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire
[6] Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada
[7] Public Health Ontario, Toronto, Ontario, Canada
[8] Genetic Epidemiology Group, International Agency for Research on Cancer (IARC), Lyon, France
[9] Department of Genetic Epidemiology, University Medical Center, Georg-August-University Göttingen, Göttingen, Germany
[10] Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK
[11] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland
[12] Helmholtz Centre Munich, German Research Centre for Environmental Health, Institute of Epidemiology I, Neuherberg, Germany
[13] Department of Molecular Biology, University of Salzburg, Salzburg, Austria
[14] Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas
[15] Massachusetts General Hospital, Boston, Massachusetts
[16] Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts
[17] Department of Population Health Sciences, Duke University School of Medicine, Durham, North Carolina

Correspondence
Qingyi Wei, MD, PhD, Duke Cancer Institute, Duke University Medical Center, 905 S. LaSalle Street, Durham, NC.
Email: qingyi.wei@duke.edu

Funding information
NIH, Grant number: CA014236; P30 Cancer Center Support Grant NIH, Grant number: CA014236

The P38MAPK pathway participates in regulating cell cycle, inflammation, development, cell death, cell differentiation, and tumorigenesis. Genetic variants of some genes in the P38MAPK pathway are reportedly associated with lung cancer risk. To substantiate this finding, we used six genome-wide association studies (GWASs) to comprehensively investigate the associations of 14 904 single nucleotide polymorphisms (SNPs) in 108 genes of this pathway with lung cancer risk. We identified six significant lung cancer risk-associated SNPs in two genes (*CSNK2B* and

Abbreviations: AD, Adenocarcinoma; CI, confidence interval; eQTL, expression quantitative trait loci; FDR, false discovery rate; GWAS, genome-wide association study; ILCCO, International Lung Cancer Consortium; LD, linkage disequilibrium; OR, odds ratio; SC, squamous cell carcinoma; SNP, single nucleotide polymorphisms; TCGA, The Cancer Genome Atlas; TRICL, Transdisciplinary Research in Cancer of the Lung.

Yun Feng, Yanru Wang, and Qingyi Wei contributed equally to this work.

ZAK) after correction for multiple comparisons by a false discovery rate (FDR) <0.20. After removal of three *CSNK2B* SNPs that are located in the same locus previously reported by GWAS, we performed the LD analysis and found that rs3769201 and rs7604288 were in high LD. We then chose two independent representative SNPs of rs3769201 and rs722864 in *ZAK* for further analysis. We also expanded the analysis by including these two SNPs from additional GWAS datasets of Harvard University (984 cases and 970 controls) and deCODE (1319 cases and 26 380 controls). The overall effects of these two SNPs were assessed using all eight GWAS datasets (OR = 0.92, 95%CI = 0.89-0.95, and $P = 1.03 \times 10^{-5}$ for rs3769201; OR = 0.91, 95% CI = 0.88-0.95, and $P = 2.03 \times 10^{-6}$ for rs722864). Finally, we performed an expression quantitative trait loci (eQTL) analysis and found that these two SNPs were significantly associated with *ZAK* mRNA expression levels in lymphoblastoid cell lines. In conclusion, the *ZAK* rs3769201 and rs722864 may be functional susceptibility loci for lung cancer risk.

KEYWORDS
lung cancer risk, pathway analysis, SNP, *ZAK*

1 | INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths among adults worldwide. In the United States, it is estimated that 224,390 new lung cancer cases will occur in 2016.[1] Both environmental and genetic factors contribute to the risk of lung cancer.[2,3] Single nucleotide polymorphisms (SNPs) are the most common genetic variants that are found to be associated with cancer risk, including lung cancer.[4,5] Although genome-wide association studies (GWASs) have identified multiple SNPs to be associated with lung cancer risk, most of these SNPs have no annotated biological functions.[6,7] Therefore, we sought to perform a hypothesis-driven pathway-based analysis to identify possible functional SNPs that may be associated with lung cancer risk but have not been reported by previous single GWAS analysis. This approach has helped us successfully identify additional unreported susceptibility loci in those genes involved in centrosome,[8] DNA repair,[9] lncRNA,[10] and RNA degradation.[11] In the present study, we investigated the associations between genetic variants of genes in the P38 mitogen-activated protein kinase (P38MAPK) pathway and lung cancer risk.

P38MAPK belongs to the MAP kinase family and is involved in cell cycle, inflammation, development, cell death, cell differentiation, and tumorigenesis.[12,13] Many transcription factors, including p53, activating transcription factor 1/2/6 (ATF-1/2/6), C/EBP, SRF accessory protein (Sap1), MEF2A, DDIT3 and NFAT, can be activated by P38 MAPKs.[14–19] Studies have shown that lack of P38MAPK functions may lead to cell cycle deficiency and tumorigenesis.[20,21] On the other hand, other published studies showed that the oncogenic potential of this pathway may lead to tumor growth, angiogenesis, and metastasis.[22,23]

Several studies have shown that *TP53* [24,25] and *ATM* [26] in the P38MAPK signaling pathway are associated with lung cancer risk, but these studies did not include many other candidate genes and SNPs of this pathway. In the present study, we were further motivated to comprehensively investigate this pathway, because associations between genetic variants in the P38MAPK pathway genes and lung cancer risk are collectively more significant than what would be expected by chance (see section 2).

2 | METHODS

2.1 | Study populations

We used the summary data from the Transdisciplinary Research in Cancer of the Lung and The International Lung Cancer Consortium (TRICL-ILCCO), which included six GWASs of 16 838 controls and 12 160 lung cancer cases.[27,28] These six GWASs included The University of Texas MD Anderson Cancer Center (MDACC) study, Institute of Cancer Research (ICR) study, National Cancer Institute (NCI) study, International Agency for Research on Cancer (IARC) study, Toronto study from Samuel Lunenfeld Research Institute (Toronto) study, and German Lung Cancer (GLC) study. The expanded analysis included two additional GWASs from ILCCO: the Harvard Lung Cancer (Harvard) study (984 cases and 970 controls)[29] and the Icelandic Lung Cancer (deCODE) study (1319 cases and 26 380 controls).[30] A written informed consent was obtained from all participating subjects in the original GWASs. All methods were performed in accordance with the relevant guidelines and regulations for each of the participating institutions. The present study also followed the study protocols approved by the Duke University Health System Institutional Review Board.

2.2 | Selection of genes and SNPs from the P38MAPK pathway

Multiple genotyping platforms were used in these GWASs, including Illumina HumanHap 317, 317 + 240S, 370Duo, 550, 610 or 1M arrays for all the GWAS datasets. We used IMPUTE2 v2.1.1[6] or MaCH v1.0[7] software to perform the imputation of untyped SNPs using the 1000 Genomes Project (phase I integrated release 3, March 2012) as the reference. Genes in the P38MAPK pathway were identified from the Molecular Signatures Database (C2).[31] Overall, 108 genes located on autosomal chromosomes were selected (details presented in Supplementary Table S1). There were 14 904 SNPs within these selected genes with 2 kb upstream and 2 kb downstream with the following selection criteria: (1) minor allele frequency (MAF) ≥5%; (2) genotyping rate ≥95%; and (3) Hardy-Weinberg Equilibrium (HWE) exact *P*-value ≥ 10^{-5}. The detailed workflow is shown in Figure 1. *P* values from lung cancer GWAS meta-analysis of these SNPs significantly deviated from the null distribution, with Kolmogorov-Smirnov test $P = 4.67 \times 10^{-6}$.

2.3 | In silico functional prediction and validation

SNPinfo,[32] RegulomeDB,[33] and HaploReg[34] were used to predict SNP-associated potential functions. The expression quantitative trait loci (eQTL) analysis was performed by using the genotyping and expression data available from the lymphoblastoid cell data of 373 European individuals from Genetic European Variation in Health and Disease Consortium (GEUVADIS) and the 1000 Genomes Project (phase I integrated release 3, March 2012).[35] We also performed eQTL and differential expression analyses by using expression data of both tumor and adjacent normal tissues from the Cancer Genome Atlas (TCGA) database (dbGaP Study Accession: phs000178.v1.p1).[36,37] The TCGA level 3 RNAseq data (LUSC_rnaseqv2_Level_3_RSEM_genes_normalized_data.2016012800.0.0.tar.gz and LUAD_Level_3_RSEM_genes_normalized_data_2016012800.0.0.tar.gz) was obtained from the Broad TCGA GDAC site (http://gdac.broadinstitute.org).

2.4 | Statistical analysis

We performed an unconditional logistic regression to estimate odds ratios (ORs) and 95% confidence intervals (CIs) per effect allele by using R (v2.6), Stata (v10, State College, Texas) and PLINK (v1.06) for each GWAS dataset, and a meta-analysis was also performed on the selected 14 904 SNPs. We tested the heterogeneity among the GWASs by using the Cochran's Q statistic and investigated the proportion of the total variation by the I^2 statistic. When there was no heterogeneity among GWASs (Q-test $P > 0.100$ and $I^2 < 25\%$), we used the fixed-effects model; otherwise, we used the random-effects model. We controlled for multiple testing with a threshold of a false discovery rate (FDR) <0.20. The paired Student *t*-test was used to test for the differences in gene mRNA expression levels between lung cancer and adjacent normal tissues from the TCGA database. LocusZoom (http://locuszoom.sph.umich.edu/locuszoom/) (reference version: 1000 Genomes, Nov 24, 2014; EUR)[38] was employed to generate the regional association plots.[38] The Manhattan plot and linkage disequilibrium (LD) plots were generated by Haploview v4.2. We used the LD analysis in choosing representative SNPs of the

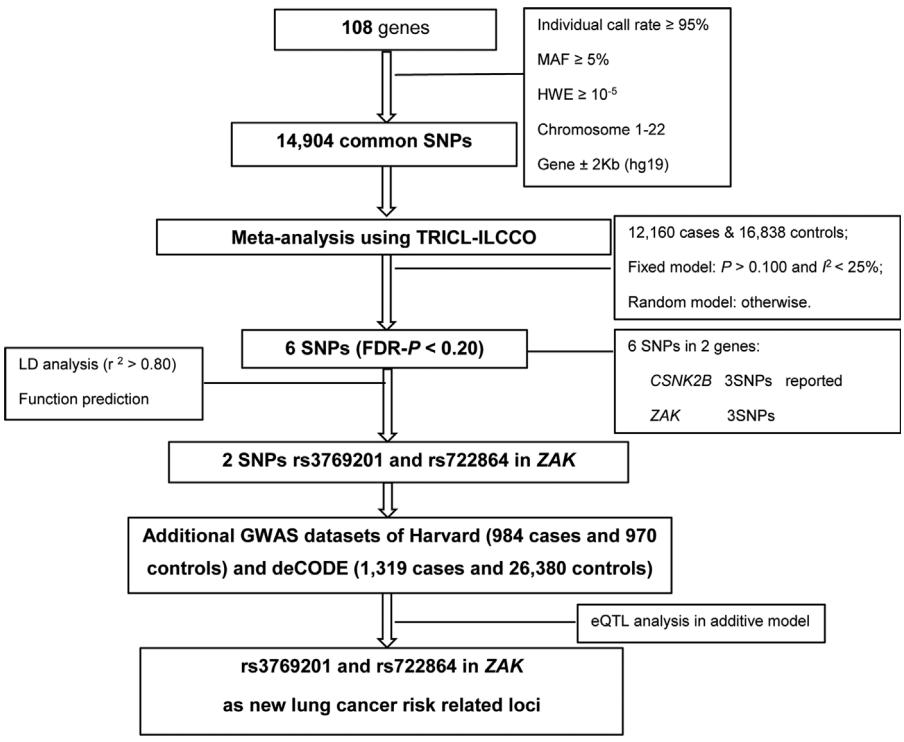

FIGURE 1 Flowchart of SNP selection among the P38MAPK pathway genes

TABLE 1 Associations between six SNPs in the P38MAPK pathway genes and lung cancer risk with FDR <0.20 in six GWASs

SNP	Gene	Chr.	Allele[a]	Position (hg19)	I^2	EAF	OR (95%CI)	*P*	FDR
rs114487324	*CSNK2B*	6	A/G	31636742	20	0.11	1.20 (1.13-1.28)	6.31E-11	9.41E-07
rs115609040	*CSNK2B*	6	C/A	31632134	8	0.19	1.11 (1.06-1.16)	2.76E-06	0.015
rs116442837	*CSNK2B*	6	G/T	31633496	8	0.19	1.11 (1.06-1.16)	2.95E-06	0.015
rs3769201	***ZAK***	**2**	**C/T**	**173956541**	**15**	**0.22**	**0.91 (0.88-0.96)**	**2.48E-05**	**0.092**
rs7604288	*ZAK*	2	T/G	173998431	0	0.22	0.92 (0.88-0.95)	3.21E-05	0.096
rs722864	***ZAK***	**2**	**G/A**	**173983204**	**17**	**0.20**	**0.91 (0.87-0.97)**	**5.41E-05**	**0.134**

SNP, single nucleotide polymorphism; Chr, chromosome; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval; FDR, false discovery rate. Newly identified tagSNPs are marked in bold.
[a]Reference allele/effect allele.

harboring genes. All other analyses were conducted with SAS (Version 9.3; SAS Institute, Cary, NC), unless specified otherwise. All the statistical methods and codes were checked and reproduced by one of the co-authors.

3 | RESULTS

3.1 | Analysis of the six GWAS datasets

In total, 14 904 SNPs from 108 genes were available from the six GWAS datasets of the TRICL-ILCCO consortium. A Manhattan plot demonstrating the associations of SNPs of these genes and lung cancer risk as identified by the single locus analysis are presented in Supplementary Figure S1A. Overall, six SNPs in two genes (*CSNK2B* and *ZAK*) remained significantly associated with lung cancer risk after multiple-testing correction by FDR <0.20. Their locations and associations with lung cancer risk are presented in Table 1. We excluded three SNPs in *CSNK2B*, because they are located on the same locus (6p21.33) previously reported by a GWAS.[39] Based on LD analysis ($r^2 > 0.80$) (Supplementary Figure S2A) and in silico SNP functional prediction (SNPinfo, RegulomeDB, and HaploReg) (Supplementary Table S2), we chose two representative SNPs: rs3769201 and rs722864 of *ZAK* for further analyses. Regional association plots for rs3769201 and rs722864 in 500 kb up- and downstream region are shown in Supplementary Figures S2B and S2C. The regional association plots demonstrated that the top SNP rs3769201 was in high LD with rs7604288 and a medium LD with rs722864. The two representative SNPs of *ZAK* had no LD with previously reported GWAS loci.

3.2 | Functional validation by the eQTL analysis

We performed the eQTL analysis to assess the associations between SNPs and their gene mRNA expression levels in the lymphoblastoid cell lines from 373 subjects of European ancestry in the 1000 Genomes project, and we found that *ZAK* rs3769201 and rs722864 were associated with *ZAK* mRNA expression levels in an additive model (Figure 2). *ZAK* mRNA expression levels significantly decreased with an increased number of the rs3769201T allele in additive ($P = 2.86 \times 10^{-4}$) (Figure 2A). The eQTL analysis results of rs722864 were also significant in an additive model ($P = 1.68 \times 10^{-4}$) (Figure 2B). By using the genotyping data and expression data in tumor tissues of lung cancer patients in the TCGA database, we performed SNP-mRNA correlation analysis for these two SNPs and found no significant correlation for the eQTL analysis in lung adeno carcinoma for both SNPs (Supplementary Figures S2A and S2B). The two SNPs showed a

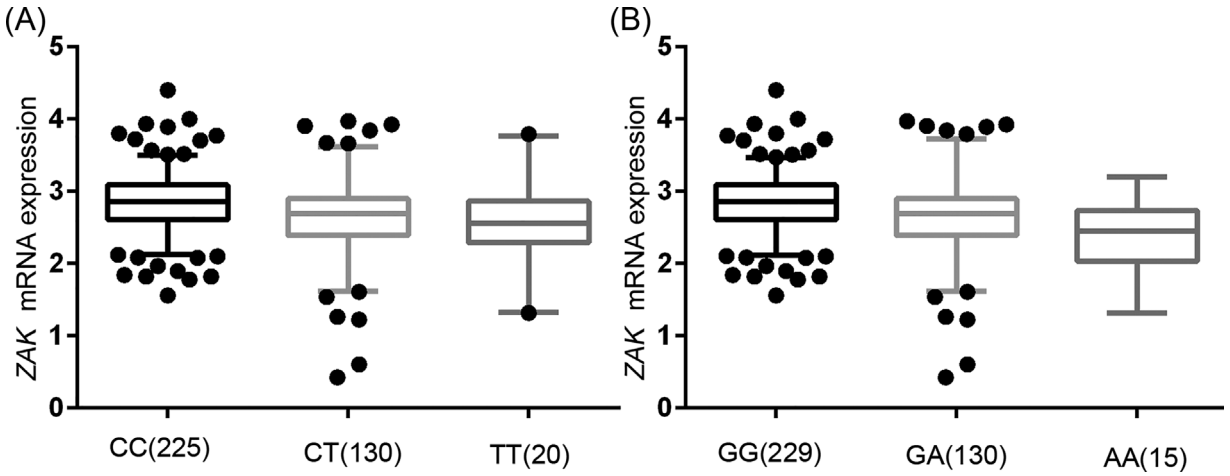

FIGURE 2 The correlations between identified SNPs and *ZAK* mRNA expression levels. A, rs3769201 in an additive model, $P = 2.86 \times 10^{-4}$; B, rs722864 in an additive model, $P = 1.68 \times 10^{-4}$.

marginally significantly correlation with *ZAK* mRNA expression levels in the primary tumor tissues of patients with lung squamous cell carcinoma ($P = 0.067$ and 0.069 for rs3769201 and rs722864, respectively. Supplementary Figures S2C and S2D). In addition, we compared mRNA expression levels of *ZAK* in 109 paired target tissue samples with normal adjacent tissue samples from the TCGA database and found that *ZAK* mRNA expression levels were also significantly decreased in the tumor tissues compared to the normal tissues ($P = 6.29E-08$), as well as stratified by adenocarcinoma (AD) and squamous cell lung carcinoma (SC) (Figure 3). Therefore, rs3769201 and rs722864 were chosen as the representative SNPs for further analyses because they were significantly associated with lung cancer risk as assessed in the overall association analysis and had potential functions according to the eQTL analysis.

3.3 | Expanded analysis by additional two GWASs

We sought to expand our analysis by two additional independent lung cancer GWASs, Harvard Lung Cancer Study (Harvard) and Icelandic Lung Cancer Study (deCODE). We subsequently performed an overall meta-analysis to evaluate associations between the two *ZAK* SNPs and lung cancer risk in all eight GWASs. We found the overall effect of these two SNPs from among all eight GWASs remained significant (OR = 0.92, 95%CI = 0.89-0.95, *P*-value of heterogeneity test [*P*het] = 0.471, and $P = 1.03E-05$ for rs3769201 and OR = 0.91, 95% CI = 0.88-0.95, *P*het = 0.504, and $P = 2.03E-06$ for rs722864) (Table 2 and Supplementary Figures S3A and S3B).

In subgroup analysis by histology (Table 2, Supplementary Figure S3), we found that the rs3769201T allele was significantly associated with SC risk (OR = 0.91, 95%CI = 0.85-0.96, $P = 0.002$), but not with AD risk (OR = 0.97, 95%CI = 0.89-1.05, $P = 0.401$). Similarly, we also found that the rs722864A allele was associated with SC risk (OR = 0.91, 95%CI = 0.86-0.97, $P = 0.004$), but not with AD risk (OR = 0.94, 95%CI = 0.87-1.02, $P = 0.139$) as well. In subgroup analysis by smoking status, there was a significant decrease in lung cancer risk for the rs3769201T allele among ever smokers (OR = 0.90, 95%CI = 0.85-0.94, $P = 1.79E-05$), but not among never smokers (OR = 0.97, 95%CI = 0.84-1.13, $P = 0.725$) (Table 3, Supplementary Figure S3A). We also found that the rs722864A allele was associated with lung cancer risk among ever smokers (OR = 0.89, 95%CI = 0.85-0.94, $P = 5.58E-05$), but not among never smokers (OR = 0.95, 95%CI = 0.80-1.12, $P = 0.520$) (Table 3, Supplementary Figure S3B). However, heterogeneity test showed that the effect difference between ever smokers and never smokers was statistically non-significant for both SNPs (*P*het = 0.349 for rs3769201 and *P*het = 0.467 for rs722864). There is also no significant difference between AD and SC by heterogeneity test (*P*het = 0.223 for rs3769201 and *P*het = 0.524 for rs722864).

4 | DISCUSSION

In the present study, we used eight published GWASs from the TRICL-ILCCO consortium to investigate the associations between genetic variants in P38MAPK pathway genes and lung cancer risk. We found that two novel, potentially functional SNPs, that is, rs3769201T and rs722864A alleles of *ZAK*, were associated with both a decreased lung cancer risk and a decreased mRNA expression level of ZAK. We also demonstrated that the rs3769201T and rs722864A alleles were significantly associated with risk of both lung AD and SC among ever smokers.

The P38 signaling cascade activation is triggered by several MAP3Ks. *ZAK* (sterile alpha motif and leucine zipper-containing kinase AZK) is a subfamily of MAP3Ks. ZAK participate in cell cycle, apoptosis, neoplastic cell transformation, and several other cancer-related pathways.[40,41] ZAK has two major different transcript variants, ZAK-α and ZAK-β. In the present study, we found that *ZAK* mRNA expression levels were significantly decreased in tumor tissues, compared with normal tissues in 109 paired target tissue samples from TCGA. These findings suggest that ZAK may be a suppressor gene, because ZAK has been shown to behave as a tumor suppressor by inhibiting lung cancer growth.[42] However, more evidence supports that ZAK may have a pro-oncogenic function. For example, the TCGA data showed significantly higher ZAK mRNA expression levels in lung cancer tissues than adjacent normal tissues in cancers of the bladder, breasts, and stomach. Others showed that the overexpression of ZAK-α activated several cancer-related signaling genes, such as AP1 and NF-kB[43] and that ZAK also participate in cell proliferation in gastric cell lines[44] as well as enhanced human colon cancer HCT116 cell EGF-dependent motility and migration.[45] These studies were

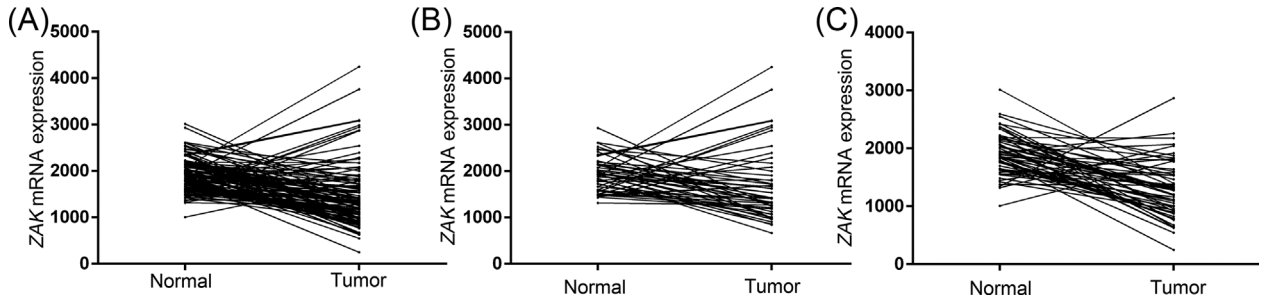

FIGURE 3 The mRNA expression levels of *ZAK* in the 109 paired lung cancer and normal adjacent tissue samples from the TCGA database (A, over all, $P = 6.29 \times 10^{-8}$; B, squamous cell carcinoma, $P = 0.069$; C, adenocarcinoma, $P = 1.55 \times 10^{-9}$).

TABLE 2 Associations between two tagSNPs and lung cancer risk stratified by histologic types in eight lung cancer GWASs

	Overall				AD				SC			
Study	Case	Control	OR (95%CI)	*P*	Case	Control	OR (95%CI)	*P*	Case	Control	OR (95%CI)	*P*
rs3769201 C > T												
ICR	1952	5200	0.98 (0.90, 1.07)	0.700	465	5200	1.04 (0.89, 1.23)	0.596	611	5200	0.96 (0.83, 1.11)	0.591
MDACC	1150	1134	0.91 (0.79, 1.06)	0.218	619	1134	0.95 (0.80, 1.13)	0.577	306	1134	0.91 (0.73, 1.14)	0.426
IARC	2533	3791	0.89 (0.81, 0.97)	0.009	517	2824	0.99 (0.84, 1.16)	0.856	911	2968	0.89 (0.79, 1.02)	0.088
NCI	5713	5736	0.88 (0.83, 0.94)	1.05E-04	1841	5736	0.85 (0.77, 0.93)	4.58E-04	1447	5736	0.89 (0.81, 0.98)	0.023
Toronto	331	499	1.04 (0.81, 1.33)	0.783	90	499	0.85 (0.58, 1.27)	0.437	50	499	1.00 (0.61, 1.65)	0.999
GLC	481	478	1.01 (0.81, 1.27)	0.924	186	478	1.30 (0.97, 1.75)	0.081	97	478	0.95 (0.63, 1.43)	0.793
Harvard	984	970	0.98 (0.84, 1.15)	0.802	597	970	1.02 (0.85, 1.21)	0.868	216	970	0.91 (0.70, 1.19)	0.513
deCODE	1319	26380	[a]	[a]	547	26380	[a]	[a]	259	26380	[a]	[a]
Overall	14463	44188	0.92 (0.89, 0.95)	1.03E-05	4862	43221	0.97 (0.89, 1.05)	0.401	3897	43365	0.91 (0.85, 0.96)	0.002
rs722864 G > A												
ICR	1952	5200	0.98 (0.89, 1.07)	0.623	465	5200	1.07 (0.90, 1.26)	0.461	611	5200	0.97 (0.83, 1.12)	0.664
MDACC	1150	1134	0.90 (0.78, 1.05)	0.179	619	1134	0.93 (0.78, 1.11)	0.436	306	1134	0.94 (0.74, 1.19)	0.604
IARC	2533	3791	0.90 (0.82, 0.99)	0.033	517	2824	1.00 (0.84, 1.18)	0.982	911	2968	0.95 (0.83, 1.08)	0.414
NCI	5713	5736	0.87 (0.82, 0.93)	8.40E-05	1841	5736	0.85 (0.77, 0.94)	0.001	1447	5736	0.87 (0.78, 0.97)	0.009
Toronto	331	499	1.01 (0.78, 1.33)	0.915	90	499	0.76 (0.50, 1.16)	0.207	50	499	1.15 (0.67, 1.98)	0.609
GLC	481	478	1.07 (0.84, 1.35)	0.585	186	478	1.27 (0.93, 1.73)	0.136	97	478	0.96 (0.63, 1.47)	0.855
Harvard	984	970	0.88 (0.76, 1.03)	0.117	597	970	0.90 (0.75, 1.07)	0.219	216	970	0.76 (0.58, 1.00)	0.051
deCODE	1319	26380	[a]	[a]	547	26380	[a]	[a]	259	26380	[a]	[a]
Overall	14463	44188	0.91 (0.88, 0.95)	2.03E-06	4862	43221	0.94 (0.87, 1.02)	0.139	3897	43365	0.91 (0.86, 0.97)	0.004

GWAS, genome-wide association study; OR, odds ratio; CI, confidence interval; AD, adenocarcinoma; SC, squamous cell carcinoma.

[a]The deCORE data was allowed to be used for calculation in the meta-analysis but cannot be shown according to the deCODE's requirement.

TABLE 3 Associations between two tagSNPs and lung cancer risk stratified by smoking status in eight lung cancer GWASs

			rs3769201 C > T		rs722864 G > A	
Study	**Case**	**Control**	**OR (95%CI)**	***P***	**OR (95%CI)**	***P***
Ever smokers						
IARC	2367	2508	0.87 (0.79, 0.96)	0.005	0.88 (0.79, 0.98)	0.017
Toronto	236	272	1.04 (0.75, 1.43)	0.820	1.05 (0.75, 1.48)	0.770
GLC	433	258	0.97 (0.72, 1.31)	0.844	0.97 (0.71, 1.32)	0.835
Harvard	892	809	0.97 (0.82, 1.15)	0.733		
MDACC	1150	1134	0.91 (0.79, 1.06)	0.218	0.90 (0.78, 1.05)	0.179
ATBC	1732	1270	0.88 (0.77, 1.00)	0.048	0.90 (0.78, 1.04)	0.143
CPSII	600	383	0.79 (0.62, 1.00)	0.051	0.86 (0.68, 1.10)	0.241
EAGLE	1767	1339	0.90 (0.80, 1.02)	0.107	0.90 (0.79, 1.02)	0.104
PLCO	1243	1344	0.90 (0.77, 1.03)	0.133	0.86 (0.74, 1.00)	0.045
Overall	10420	9317	0.90 (0.85, 0.94)	1.79E-05	0.89 (0.85, 0.94)	5.58E-05
Never smokers						
IARC	159	1253	0.94 (0.70, 1.25)	0.663	0.95 (0.70, 1.31)	0.769
Toronto	95	217	0.95 (0.61, 1.50)	0.832	0.89 (0.55, 1.44)	0.642
GLC	35	220	1.15 (0.61, 2.18)	0.663	1.16 (0.59, 2.29)	0.675
Harvard	92	161	1.03 (0.66, 1.60)	0.896		
CPSII	86	275	1.03 (0.67, 1.60)	0.880	1.09 (0.70, 1.70)	0.710
EAGLE	138	634	1.07 (0.77, 1.50)	0.675	1.00 (0.70, 1.43)	0.987
PLCO	126	470	0.76 (0.49, 1.17)	0.208	0.72 (0.46, 1.12)	0.145
Overall	731	3230	0.97 (0.84, 1.13)	0.725	0.95 (0.80, 1.12)	0.520

GWAS, genome-wide association study; OR, odds ratio; CI, confidence interval. NCI GWAS includes four sub-studies: the alpha-tocopherol, beta-carotene cancer prevention study (ATBC), the cancer prevention study II nutrition cohort (CPS-II), the environment and genetics in lung cancer etiology (EAGLE), and the prostate, lung, colon, ovary screening trial (PLCO).

consistent with the results of the present study, in which the two SNPs had a loss of function, as evident by a decreased mRNA expression level of the gene, which was associated with a decreased lung cancer risk.

We also performed subgroup analysis by histology and smoking status, and we demonstrated that both rs3769201T and rs722864A alleles were associated with risk of SC and ever smokers, but not with AD and never smokers. Cigarette smoke is the major risk factor for lung cancer, especially for SC. It has been reported that the transcription factor (TF) STAT can be activated after tobacco smoke exposure.[46] Functional prediction analysis from HaploReg showed that rs3769201 had a STAT binding site motif, and the results of eQTL in the present study also demonstrated that the SNP may alter the mRNA expression of *ZAK*. However, there is no significant difference between ever smokers and never smokers as tested by the heterogeneity test.

The present study has some limitations. First of all, although we found five relevant pathways from the Molecular Signatures Database, perhaps some newly discovered genes may not have been included yet. Second, some published studies support that *ZAK* is an oncogene,[44] but others have shown that ZAK may have the function of suppressing cancer growth.[42] More biological and molecular experiments should be performed to reveal the mechanisms underlying the observed associations. Finally, the analyses have not been adjusted for potentially important baseline risk covariates including family history.

In conclusion, the present study of eight published GWASs revealed two novel, potentially functional susceptibility loci in *ZAK* associated with lung cancer risk in European populations. Further functional evaluations of these genetic variants are warranted to verify our findings.

ACKNOWLEDGMENTS

As Duke Cancer Institute members, QW, and KO acknowledge support from the Duke Cancer Institute as part of the P30 Cancer Center Support Grant (Grant ID: NIH CA014236). QW was also supported by a start-up fund from Duke Cancer Institute, Duke University Medical Center.

FUNDING INFORMATION

TRICL-ILCCO: This work was supported by the Transdisciplinary Research in Cancer of the Lung (TRICL) Study, U19-CA148127 on behalf of the Genetic Associations, and Mechanisms in Oncology (GAME-ON) Network. The Toronto study was supported by Canadian Cancer Society Research Institute (020214), Ontario Institute of Cancer and Cancer Care Ontario Chair Award to RH The ICR study was

supported by Cancer Research UK (C1298/A8780 andC1298/A8362—Bobby Moore Fund for Cancer Research UK) and NCRN, HEAL, and Sanofi-Aventis. Additional funding was obtained from NIH grants (5R01CA055769, 5R01CA127219, 5R01CA133996, and 5R01CA121197). The Liverpool Lung Project (LLP) was supported by The Roy Castle Lung Cancer Foundation, UK. The ICR and LLP studies made use of genotyping data from the Wellcome Trust Case Control Consortium 2 (WTCCC2); a full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Sample collection for the Heidelberg lung cancer study was in part supported by a grant (70-2919) from the Deutsche Krebshilfe. The work was additionally supported by a Helmholtz-DAAD fellowship (A/07/97379 to MNT) and by the NIH (U19CA148127). The KORA Surveys were financed by the GSF, which is funded by the German Federal Ministry of Education, Science, Research, and Technology and the State of Bavaria. The Lung Cancer in the Young study (LUCY) was funded in part by the National Genome Research Network (NGFN), the DFG (BI576/2-1; BI 576/2-2), the Helmholtzgemeinschaft (HGF) and the Federal office for Radiation Protection (BfS: STSch4454). Genotyping was performed in the Genome Analysis Center (GAC) of the Helmholtz Zentrum Muenchen. Support for the Central Europe, HUNT2/Tromsø and CARET genome-wide studies was provided by Institute National du Cancer, France. Support for the HUNT2/Tromsø genome-wide study was also provided by the European Community

(Integrated Project DNA repair, LSHG-CT- 2005-512113), the Norwegian Cancer Association and the Functional Genomics Programme of Research Council of Norway. Support for the Central Europe study, Czech Republic, was also provided by the European Regional Development Fund and the State Budget of the Czech Republic (RECAMO, CZ.1.05/2.1.00/03.0101). Support for the CARET genome-wide study was also provided by grants from the US National Cancer Institute, NIH (R01 CA111703 and UO1 CA63673), and by funds from the Fred Hutchinson Cancer Research Center. Additional funding for study coordination, genotyping of replication studies, and statistical analysis was provided by the US National Cancer Institute (R01 CA092039). The lung cancer GWAS from Estonia was partly supported by a FP7 grant (REGPOT245536), by the Estonian Government (SF0180142s08), by EU RDF in the frame of Centre of Excellence in Genomics and Estoinian Research Infrastructure's Roadmap and by University of Tartu (SP1GVARENG). The work reported in this paper was partly undertaken during the tenure of a Postdoctoral Fellowship from the IARC (for MNT). The Environment and Genetics in Lung Cancer Etiology (EAGLE), the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC), and the Prostate, Lung, Colon, Ovary Screening Trial (PLCO)studies and the genotyping of ATBC, the Cancer Prevention Study II Nutrition Cohort (CPS-II) and part of PLCO were supported by the Intramural Research Program of NIH, NCI, Division of Cancer Epidemiology, and Genetics. ATBC was also supported by US Public Health Service contracts (N01-CN-45165, N01-RC-45035, and N01-RC-37004) from the NCI. PLCO was also supported by individual contracts from the NCI to the University of Colorado Denver (NO1-CN-25514), Georgetown University(NO1-CN-25522), Pacific Health Research Institute (NO1-CN-25515), Henry Ford Health System (NO1-CN-25512), University of Minnesota(NO1-CN-25513), Washington University (NO1-CN-25516), University of Pittsburgh (NO1-CN-25511), University of Utah (NO1-CN-25524), Marshfield Clinic Research Foundation (NO1-CN-25518), University of Alabama at Birmingham (NO1-CN-75022, Westat, Inc. NO1-CN-25476), University of California, Los Angeles (NO1-CN-25404). The Cancer Prevention Study II Nutrition Cohort was supported by the American Cancer Society. The NIH Genes, Environment and Health Initiative (GEI) partly funded DNA extraction and statistical analyses (HG-06-033-NCI-01 andRO1HL091172-01), genotyping at the Johns Hopkins University Center for Inherited Disease Research (U01HG004438 and NIH HHSN268200782096C) and study coordination at the GENEVA Coordination Center (U01 HG004446) for EAGLE and part of PLCO studies. Funding for the MD Anderson Cancer Study was provided by NIH grants (P50 CA70907, R01CA121197, R01CA127219, U19 CA148127, R01 CA55769, and K07CA160753) and CPRIT grant (RP100443). Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is funded through a federal contract from the NIH to The Johns Hopkins University (HHSN268200782096C). The Harvard Lung Cancer Study was supported by the NIH (National Cancer Institute) grants CA092824, CA090578, and CA074386.

deCODE: The project was funded in part by GENADDICT: LSHMCT-2004-005166), the National Institutes of Health (R01-DA017932)

TCGA: The results published here are in whole or part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute The Cancer Genome Atlas (TCGA) Research Network can be found at "http://cancergenome.nih.gov." The TCGA SNP data analyzed here are requested through dbGAP (accession#: phs000178.v1.p1).

CONFLICTS OF INTEREST

The authors declare no potential conflicts of interest.

ORCID

Qingyi Wei http://orcid.org/0000-0002-3845-9445

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA. Cancer J. Clin.* 2016;66:7–30.
2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc*. 2008;83:584–594.
3. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. *Clin Chest Med*. 2011;32:605–644.
4. Smith C. Genomics: sNPs and human disease. *Nature*. 2005;435:993.
5. Wang Z, Seow WJ, Shiraishi K, et al. Meta-analysis of genome-wide association studies identifies multiple lung cancer susceptibility loci in never-smoking Asian women. *Hum Mol Genet*. 2016;25:620–629.
6. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet*. 2008;17:R156–R165.
7. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93:779–797.

8. Kang X, Liu H, Onaitis MW, et al. Polymorphisms of the centrosomal gene (FGFR1OP) and lung cancer risk: a meta-analysis of 14,463 cases and 44,188 controls. *Carcinogenesis* 2016;37:280–289.
9. Wang M, Liu H, Liu Z, et al. Genetic variant in DNA repair gene GTF2H4 is associated with lung cancer risk: a large-scale analysis of six published GWAS datasets in the TRICL consortium. *Carcinogenesis* 2016;37:888–896.
10. Yuan H, Liu HL, Liu ZS, et al. A novel genetic variant in long non-coding RNA gene NEXN-AS1 is associated with risk of lung cancer. *Sci Rep-Uk*. 2016;6:34234.
11. Zhou F, Wang Y, Liu H, et al. Susceptibility loci of CNOT6 in the general mRNA degradation pathway and lung cancer risk-A re-analysis of eight GWASs. *Mol Carcinog*. 2016;56:1227–1238.
12. Wagner EF, Nebreda AR. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat Rev Cancer*. 2009;9:537–549.
13. Zarubin T, Han J. Activation and signaling of the p38 MAP kinase pathway. *Cell Res*. 2005;15:11–18.
14. Hazzalin CA, Cano E, Cuenda A, Barratt MJ, Cohen P, Mahadevan LC. P38/RK is essential for stress-induced nuclear responses: jNK/SAPKs and c-Jun/ATF-2 phosphorylation are insufficient. *Curr Biol*. 1996;6:1028–1031.
15. Whitmarsh AJ, Yang SH, Su MS, Sharrocks AD, Davis RJ. Role of p38 and JNK mitogen-activated protein kinases in the activation of ternary complex factors. *Mol Cell Biol*. 1997;17:2360–2371.
16. Janknecht R, Hunter T. Convergence of MAP kinase pathways on the ternary complex factor Sap-1a. *EMBO J*. 1997;16:1620–1627.
17. Zhao M, New L, Kravchenko VV, et al. Regulation of the MEF2 family of transcription factors by p38. *Mol Cell Biol*. 1999;19:21–30.
18. Pereira RC, Delany AM, Canalis E. CCAAT/enhancer binding protein homologous protein (DDIT3) induces osteoblastic cell differentiation. *Endocrinology*. 2004;145:1952–1960.
19. Gomez del Arco P, Martinez-Martinez S, Maldonado JL, Ortega-Perez I, Redondo JM. A role for the p38 MAP kinase pathway in the nuclear shuttling of NFATp. *J Biol Chem*. 2000;275:13872–13878.
20. Todd DE, Densham RM, Molton SA, et al. ERK1/2 and p38 cooperate to induce a p21CIP1-dependent G1 cell cycle arrest. *Oncogene*. 2004;23:3284–3295.
21. Faust D, Dolado I, Cuadrado A, et al. P38alpha MAPK is required for contact inhibition. *Oncogene*. 2005;24:7941–7945.
22. Simon C, Goepfert H, Boyd D. Inhibition of the p38 mitogen-activated protein kinase by SB 203580 blocks PMA-induced Mr 92,000 type IV collagenase secretion and in vitro invasion. *Cancer Res*. 1998;58:1135–1139.
23. Houle F, Huot J. Dysregulation of the endothelial cellular response to oxidative stress in cancer. *Mol Carcinog*. 2006;45:362–367.
24. Mostaid MS, Ahmed MU, Islam MS, Bin Sayeed MS, Hasnat A. Lung cancer risk in relation to TP53 codon 47 and codon 72 polymorphism in Bangladeshi population. *Tumour Biol*. 2014;35:10309–10317.
25. Li Y, Chang SC, Niu R, et al. TP53 genetic polymorphisms, interactions with lifestyle factors and lung cancer risk: a case control study in a Chinese population. *BMC Cancer*. 2013;13:607.
26. Bhowmik A, Nath S, Das S, Ghosh SK, Choudhury Y. ATM rs189037 (G>A) polymorphism and risk of lung cancer and head and neck cancer: a meta-analysis. *Meta Gene*. 2015;6:42–48.
27. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet*. 2014;46:736–741.
28. Timofeeva MN, Hung RJ, Rafnar T, et al. Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet*. 2012;21:4980–4995.
29. Su L, Zhou W, Asomaning K, et al. Genotypes and haplotypes of matrix metalloproteinase 1, 3 and 12 genes and the risk of lung cancer. *Carcinogenesis*. 2006;27:1024–1029.
30. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008;452:638–642.
31. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1:417–425.
32. Xu ZL, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009;37:W600–W605.
33. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22:1790–1797.
34. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012;40:D930–D934.
35. Lappalainen T, Sammeth M, Friedlander MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501:506–511.
36. Cancer Genome Atlas Research N Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511:543–550.
37. Cancer Genome Atlas Research N Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;489:519–525.
38. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–2337.
39. Landi MT, Chatterjee N, Yu K, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet*. 2009;85:679–691.
40. Tosti E, Waldbaum L, Warshaw G, Gross EA, Ruggieri R. The stress kinase MRK contributes to regulation of DNA damage checkpoints through a p38 gamma-independent pathway. *J Biol Chem*. 2004;279:47652–47660.
41. Cho YY, Bode AM, Mizuno H, Choi BY, Choi HS, Dong Z. A novel role for mixed-lineage kinase-like mitogen-activated protein triple kinase alpha in neoplastic cell transformation and tumor development. *Cancer Res*. 2004;64:3855–3864.
42. Yang JJ, Lee YJ, Hung HH, et al. ZAK inhibits human lung cancer cell growth via ERK and JNK activation in an AP-1-dependent manner. *Cancer Sci*. 2010;101:1374–1381.
43. Liu TC, Huang CJ, Chu YC, et al. Cloning and expression of ZAK, a mixed lineage kinase-like protein containing a leucine-zipper and a sterile-alpha motif. *Biochem Bioph Res Commun*. 2000;274:811–816.
44. Liu J, McCleland M, Stawiski EW, et al. Integrated exome and transcriptome sequencing reveals ZAK isoform usage in gastric cancer. *Nat Commun*. 2014;5:3830.
45. Rey C, Faustin B, Mahouche I, et al. The MAP3K ZAK, a novel modulator of ERK-dependent migration, is upregulated in colorectal cancer. *Oncogene*. 2016;35:3190–3200.
46. Niimori-Kita K, Ogino K, Mikami S, et al. Identification of nuclear phosphoproteins as novel tobacco markers in mouse lung tissue following short-term exposure to tobacco smoke. *FEBS Open Bio* 2014;4:746–754.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

How to cite this article: Feng Y, Wang Y, Liu H, et al. Novel genetic variants in the P38MAPK pathway gene *ZAK* and susceptibility to lung cancer. *Molecular Carcinogenesis*. 2017;1–9. https://doi.org/10.1002/mc.22748