ARTICLE

Pleiotropic Associations of Risk Variants Identified for Other Cancers With Lung Cancer Risk: The PAGE and TRICL Consortia

S. Lani Park, Megan D. Fesinmeyer, Maria Timofeeva, Christian P. Caberto, Jonathan M. Kocarnik, Younghun Han, Shelly-Ann Love, Alicia Young, Logan Dumitrescu, Yi Lin, Robert Goodloe, Lynne R. Wilkens, Lucia Hindorff, Jay H. Fowke, Cara Carty, Steven Buyske, Frederick R. Schumacher, Anne Butler, Holli Dilks, Ewa Deelman, Michele L. Cote, Wei Chen, Mala Pande, David C. Christiani, John K. Field, Heike Bickeböller, Angela Risch, Joachim Heinrich, Paul Brennan, Yufei Wang, Timothy Eisen, Richard S. Houlston, Michael Thun, Demetrius Albanes, Neil Caporaso, Ulrike Peters, Kari E. North, Gerardo Heiss, Dana C. Crawford, William S. Bush, Christopher A. Haiman, Maria Teresa Landi, Rayjean J. Hung, Charles Kooperberg, Christopher I. Amos*, Loïc Le Marchand, Iona Cheng

Correspondence to: S. Lani Park, PhD, MPH, Department of Preventive Medicine, USC Keck School of Medicine, 1450 Biggy St, NRT 1509G, Los Angeles, CA 90033 (e-mail: sungship@med.usc.edu).

Background	Genome-wide association studies have identified hundreds of genetic variants associated with specific cancers. A few of these risk regions have been associated with more than one cancer site; however, a systematic evaluation of the associations between risk variants for other cancers and lung cancer risk has yet to be performed.
Methods	We included 18 023 patients with lung cancer and 60 543 control subjects from two consortia, Population Architecture using Genomics and Epidemiology (PAGE) and Transdisciplinary Research in Cancer of the Lung (TRICL). We examined 165 single-nucleotide polymorphisms (SNPs) that were previously associated with at least one of 16 non–lung cancer sites. Study-specific logistic regression results underwent meta-analysis, and associations were also examined by race/ethnicity, histological cell type, sex, and smoking status. A Bonferroni-corrected P value of 2.5×10^{-5} was used to assign statistical significance.
Results	The breast cancer SNP *LSP1* rs3817198 was associated with an increased risk of lung cancer (odds ratio [OR] = 1.10; 95% confidence interval [CI] = 1.05 to 1.14; $P = 2.8 \times 10^{-6}$). This association was strongest for women with adenocarcinoma ($P = 1.2 \times 10^{-4}$) and not statistically significant in men ($P = .14$) with this cell type ($P_{het\ by\ sex} = .10$). Two glioma risk variants, *TERT* rs2853676 and *CDKN2BAS1* rs4977756, which are located in regions previously associated with lung cancer, were associated with increased risk of adenocarcinoma (OR = 1.16; 95% CI = 1.10 to 1.22; $P = 1.1 \times 10^{-8}$) and squamous cell carcinoma (OR = 1.13; CI = 1.07 to 1.19; $P = 2.5 \times 10^{-5}$), respectively.
Conclusions	Our findings demonstrate a novel pleiotropic association between the breast cancer *LSP1* risk region marked by variant rs3817198 and lung cancer risk.

JNCI J Natl Cancer Inst (2014) 106(4): dju061 doi:10.1093/jnci/dju061

Globally, lung cancer is the most common malignancy and cause of cancer-related deaths (1,2). Tobacco smoking is the primary risk factor for lung cancer, but there is evidence that genetic susceptibility plays a role. Notably, recent genome-wide association studies (GWASs) of lung cancer have identified single-nucleotide polymorphisms (SNPs) in at least 10 independent loci ($P < 5 \times 10^{-8}$) influencing risk in different populations (3). However, these identified loci explain only a small fraction of lung cancer susceptibility and the challenge remains to identify the many additional common risk loci that are expected to have small genetic effects (3).

To date, more than 400 SNPs have been associated with cancer in GWASs (3). The discovery of pleiotropic effects, where a single gene variant is associated with more than one phenotype, may allow for the identification of shared disease pathways. For cancer, this may ultimately lead to the detection of susceptible individuals as well as in the development of regimens for the prevention of multiple cancers and pathway-based treatment. Genetic variants at chromosome 8q24, in *TP53*, and in *TERT*, the telomerase reverse transcriptase gene, are examples of loci with pleiotropic effects for multiple cancer sites and other chronic diseases (4–6). For lung cancer, a systematic evaluation of possible pleiotropic associations for the many risk variants identified with other cancer sites has yet to be conducted.

These genetic associations may have been missed in prior GWASs of lung cancer due to the heavy multiple comparison penalty in surveying the entire genome or due to disease

heterogeneity in factors such as histological cell types or smoking status. For example, *TERT* rs2736100 (7–9) has been primarily associated with risk of adenocarcinoma of the lung, often diagnosed among nonsmokers, whereas SNPs in the 15q25 region, which include nicotinic acetylcholine receptor genes involved in nicotine dependence, have been primarily associated with lung cancer among smokers (10).

Here, we examined the pleiotropic effects of 165 risk variants initially identified for other cancers on lung cancer risk. Our study included a collaboration between two large consortia (11,12), in which we also examined the consistency of associations by race/ethnicity, tumor histology, sex, and smoking status.

Methods

Study Participants
Two consortia contributed data to this study: the Population Architecture using Genomics and Epidemiology (PAGE) (12) and the Transdiciplinary Research in Cancer of the Lung (TRICL) (11), which is part of the Genetic Associations and MEchanisms in ONcology (GAME-ON) consortium, and is associated with the International Lung Cancer Consortium (ILCCO). This collaboration provided information on 18023 patients with lung cancer and 60543 control subjects from 13 studies (Supplementary Table 1, available online). Details regarding these participating studies are described in the Supplementary Data (available online). All studies were based on primary incident nonsarcoma and nonlymphoma lung cancer cases, and more than 95% of the cases were pathologically confirmed. The majority of these studies utilized patients and control subjects who had no history of another cancer. Among the few studies in which a small proportion of patients and control subjects had a history of another cancer, our findings were similar when excluding these participants. Participants' informed consent and institutional review board approval was obtained for all studies except Epidemiologic Architecture for Genes Linked to Environment, which accesses the Vanderbilt University biorepository (EAGLE-BioVU), which is considered non-human subjects research due to sample de-identification (13).

SNP Selection and Genotyping
A total of 165 SNPs associated with 16 malignancies excluding lung cancer and smoking-related SNPs were selected as of January 2010 from the National Human Genome Research Institute GWAS catalog (3) and review of the cancer GWASs and fine-mapping literature review (Supplementary Table 3, available online). Additionally, we studied 18 lung cancer risk variants to replicate their associations with lung cancer risk in populations of European ancestry (Supplementary Table 2, available online) (11). The risk allele for each SNP was defined as the allele associated with an increased risk of cancer in the initial report. For PAGE, candidate SNP genotyping was performed using Illumina BeadXpress (Women's Health Initiative [WHI]), Sequenom (EAGLE-BioVU), and the TaqMan OpenArray platform (Multiethnic Cohort study [MEC]). Atherosclerosis Risk in Communities Study [ARIC] (in PAGE) and TRICL extracted genotypes from GWAS data and were comprised of only European-ancestry populations. The ARIC samples were genotyped using the Affymetrix 6.0 platform. Genotypes were called with Birdseed and only SNPs with call rate equal to or greater than 90%, MAF equal to or greater than 1%,

and Hardy-Weinberg equilibrium $P > 1 \times 10^{-6}$ were considered for imputation. Untyped and missing SNPs were imputed using Mach1 v1.00.16 based on HapMap release 2 (build 36) and a European ancestry (CEU) reference panel (14). Imputed SNPs with a quality threshold of r^2 greater than or equal to 0.3 were included in this analysis. MEC, EAGLE-BioVU, and WHI could not impute missing SNPs due to the reduced number of variants genotyped. For TRICL, genotyping was performed using the Illumina HumanHap300 BeadChips or Human Hap550 or 610 Quad arrays. At the time of this analysis, imputed SNPs were not available for TRICL.

All PAGE studies, with the exception of ARIC, genotyped a panel of 128 ancestry informative markers (15) and used principal components analysis to estimate principal components of genetic ancestry (16). ARIC (17) and TRICL (11) estimated principal components of genetic ancestry based on GWAS data using EIGENSTRAT (16). These principal components of genetic ancestry were included in regression models to adjust for population substructure.

Standard quality-assurance and quality-control measures were utilized to ensure genotyping quality. In PAGE (12), samples and SNPs were included based on call rates (≥90%), concordance of blinded replicates (>98%), and departures from Hardy-Weinberg equilibrium ($P < .001$). More than 97.9% of samples and more than 99% of SNPs had a call rate equal to or greater than 95% in all four PAGE studies. In TRICL (11), samples were excluded if the average call rate was less than 90%; if there was sex discrepancy (threshold of heterozygosity >10% for men and <20% for women), unexpected duplicates, evidence of first-degree relatedness, or heterozygosity rates for autosomal chromosomes exceeding six standard deviations of the mean; samples with less than 80% European ancestry based on STRUCTURE (18) analysis, and outliers based on principal component analysis using EIGENSTRAT (16), were also excluded.

Statistical Analyses
For each study, we estimated the association between each SNP and risk of lung cancer using unconditional logistic regression and an additive genetic model of the risk allele. Models were adjusted for age, sex, country/study center (as appropriate), principal components of genetic ancestry, and smoking status (never, former, current). The Liverpool and Institute of Cancer Research (ICR) studies, which used generic control subjects, were not adjusted for age, sex, or smoking status. Studies with more than 85 lung cancer cases per racial/ethnic group were retained for race/ethnicity-stratified analysis. Associations by tumor histology were estimated based on logistic models of World Health Organization–defined histological cell type (adenocarcinoma, squamous cell carcinoma [SCC], and small cell lung cancer) compared to all control subjects. Large cell lung cancers were not included in the histology-specific analysis due to their limited sample size and heterogeneous nature. Stratified analyses by sex and smoking status (never and ever) were also performed.

To examine whether the associations with SNPs in *TERT* were independent of the known lung cancer risk variant in *TERT* (rs2736100) (7), conditional analysis was performed.

The regression estimates were combined across studies using inverse-variance weighted, fixed-effect meta-analysis using the METAL program, tool for meta-analysis genomewide association scans (19). The Cochran Q statistic was used to test for

heterogeneity by study and whether the meta-analyzed odds ratios (ORs) were heterogeneous by race/ethnicity, histological cell type, sex, and smoking status. To account for multiple testing of 165 SNPs and 11 stratified analyses (four race/ethnicities, three histological cell types, two sexes, two levels of smoking status), we used a Bonferroni-corrected P value to assign statistical significance ($\alpha = .05/[165$ SNPs*12 above mentioned tests$] = 2.5 \times 10^{-5}$). No additional associations were detected at a less stringent P value (e.g., $.05/165$ SNPs $= 3 \times 10^{-4}$). Statistical tests were two-sided.

Results

The main characteristics of the 18 023 patients with lung cancer and 60 543 control subjects are presented in Supplementary Table 1 (available online). The PAGE study was comprised of European-ancestry, African American, Hispanic, Asian, Pacific Islander, and American Indian populations. The TRICL study was comprised only of individuals of European descent. The great majority (96%) of subjects were of European ancestry. Also, the majority of patients and control subjects were older than 50 years, with the exception of the Helmholtz-Gemeinschaft Deutscher Forschungszentren Lung Cancer GWAS (HGF) Germany study, where all subjects were 50 years of age or younger (3%). All studies, except WHI, were comprised of both sexes. In all studies, patients were more likely to be ever smokers and control subjects were more likely to be never smokers. Histology information was available for all studies, with the exception of ARIC. Among the studies with histology information, adenocarcinoma (34.0%) was the most common cell type, with the exception of the International Agency for Research on Cancer (IARC) GWAS, where SCC was more common (35.6%).

We evaluated the association between 18 known lung risk variants located in previously identified lung cancer risk loci and risk of lung cancer among European-ancestry populations (Supplementary Table 2, available online). Of the 18 lung cancer risk variants, 16 replicated at $P < .05$.

Among the 165 risk variants, 15 were nominally associated with lung cancer at $P < .05$ (Figure 1; Supplementary Table 3, available online), which is notably more than the eight associations expected by chance (i.e., 165 SNPs*.05 = 8.3). Using a binomial distribution with a $P = .05$ and n = 165 SNPs, the probability of observing 15 or more associations is .009. These 15 associations included eight prostate cancer variants, four glioma variants, one breast cancer variant, one childhood acute lymphocytic leukemia variant, and one follicular lymphoma variant. Twelve of the 15 SNPs were associated with an increased risk of lung cancer in the same direction of the known GWAS association. No heterogeneity by race/ethnicity ($P > .05$) was noted for the 15 nominally associated SNPs (Supplementary Table 4, available online).

The breast cancer SNP *LSP1* rs3817198 was associated with an increased risk of lung cancer (OR = 1.10; 95% confidence interval [CI] = 1.05 to 1.14) and remained statistically significant ($P = 2.8 \times 10^{-6}$) after correction for multiple comparisons (Figure 2). This association was heterogeneous by cell type ($P_{het} = .03$) and sex ($P_{het} = .01$), where it appeared to be limited to adenocarcinoma (OR = 1.11; 95% CI = 1.05 to 1.17; $P = 1.14 \times 10^{-4}$) (Supplementary Table 5, available online) and women (OR = 1.16; 95% CI = 1.09 to 1.23; $P = 4.31 \times 10^{-6}$) (Supplementary Table 6, available online). This association was not observed in SCC or small cell carcinoma ($P \geq .35$) or in men ($P = .16$). In stratified analysis by both sex and histology cell type (data not shown), among studies with available
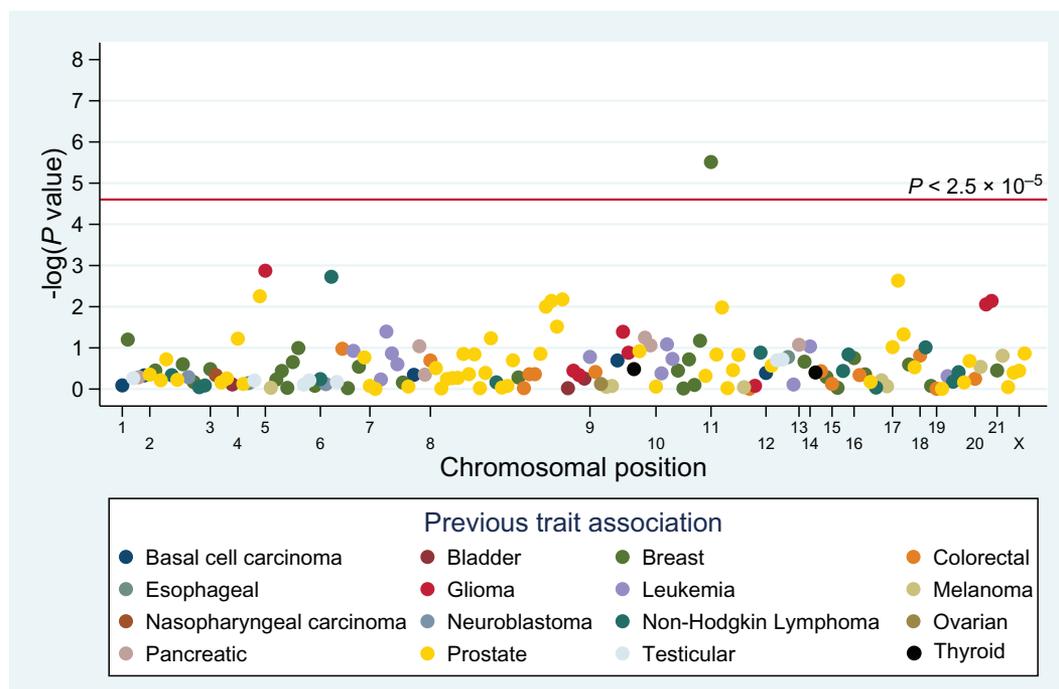

Figure 1. Manhattan plot of the meta-analysis association between risk variants of 16 other cancers and lung cancer. The solid line is the Bonferroni-corrected significance threshold. Each association is colored according to the cancer for which the single-nucleotide polymorphism was originally reported, and positioned on the *x*-axis according to its genomic position.

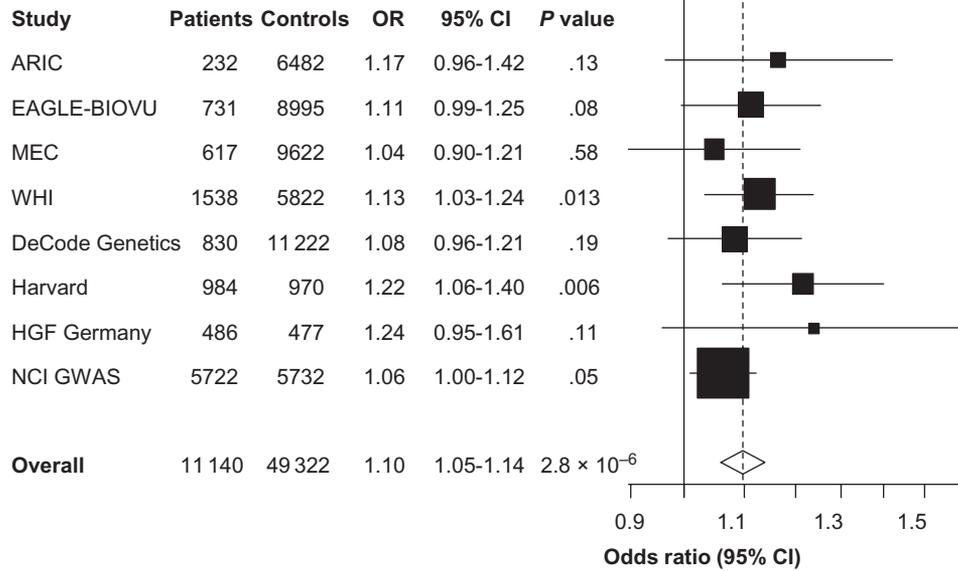**Breast Cancer SNP *LSP1* rs3817198, region 11p15.5**

Study	Patients	Controls	OR	95% CI	P value
ARIC	232	6482	1.17	0.96-1.42	.13
EAGLE-BIOVU	731	8995	1.11	0.99-1.25	.08
MEC	617	9622	1.04	0.90-1.21	.58
WHI	1538	5822	1.13	1.03-1.24	.013
DeCode Genetics	830	11 222	1.08	0.96-1.21	.19
Harvard	984	970	1.22	1.06-1.40	.006
HGF Germany	486	477	1.24	0.95-1.61	.11
NCI GWAS	5722	5732	1.06	1.00-1.12	.05
Overall	11 140	49 322	1.10	1.05-1.14	2.8×10^{-6}

Figure 2. Forest plot of the association between lymphocyte-specific protein 1 (*LSP1*) rs3817198 and lung cancer risk. Study-specific and meta-analysis associations are plotted, modeling the C risk allele for breast cancer. Squares represent odds ratios (ORs); size of the square represents inverse of the variance of the log ORs; horizontal lines represent 95% confidence intervals (CIs); diamonds represent summary estimate combining the study-specific estimates with a fixed-effects model; solid vertical lines represent OR = 1; dashed vertical lines represent the overall ORs. The single-nucleotide polymorphism (SNP) rs3817198 was genotyped in all studies. GWAS = genome-wide association study.

Glioma SNP *TERT* rs2853676, region 5p15.33

Study	Patients	Controls	OR	95% CI	P value
MEC	252	9587	1.19	0.96-1.47	.11
WHI	760	5825	1.11	0.98-1.25	.10
DeCode Genetics	346	11 225	1.03	0.86-1.22	.77
Harvard	488	970	1.26	1.06-1.51	.01
HGF Germany	188	479	1.18	0.84-1.67	.34
IARC GWAS	586	3174	1.25	1.08-1.44	.003
MDACC	619	1133	1.09	0.93-1.28	.31
NCI GWAS	1836	5686	1.19	1.09-1.29	1×10^{-4}
SLRI/Toronto	89	488	1.07	0.76-1.51	.69
Overall	5164	38 567	1.16	1.10-1.22	1.1×10^{-8}
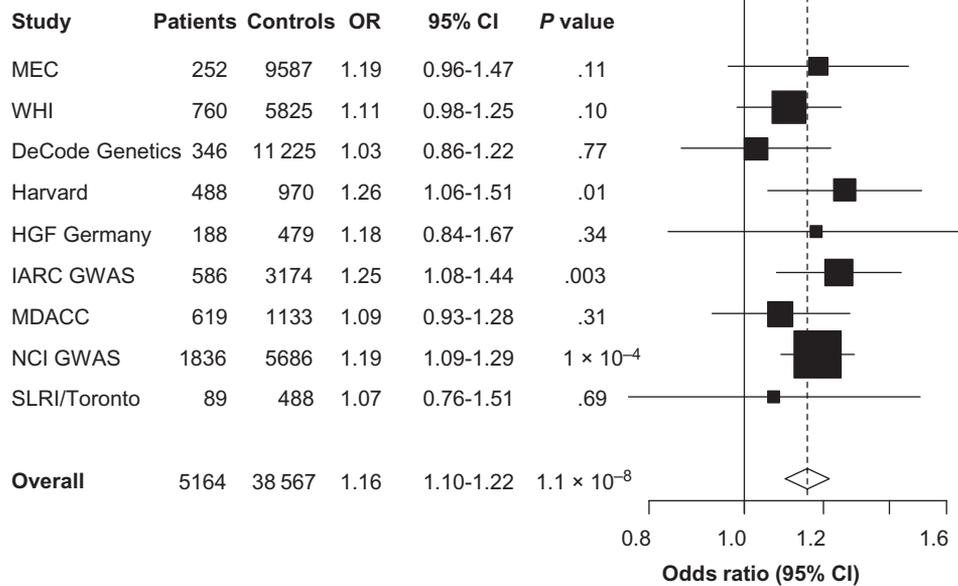
Figure 3. Forest plot of the association between telomerase reverse transcriptase gene (*TERT*) rs2853676 and lung adenocarcinoma risk. Study specific and meta-analysis associations are plotted, modeling the A risk allele for glioma. Squares represent odds ratios (ORs); size of the square represents inverse of the variance of the log ORs; horizontal lines represent 95% confidence intervals (CIs); diamonds represent summary estimate combining the study-specific estimates with a fixed-effects model; solid vertical lines represent OR = 1; dashed vertical lines represent the overall ORs. The single-nucleotide polymorphism (SNP) rs2853676 was genotyped in all studies. GWAS = genome-wide association study.

data, we found that the association was present for female adenocarcinoma (n = 1,607 cases, 4 studies: EAGLE-BioVU, MEC, National Cancer Institute Lung GWAS (NCI), and WHI; OR = 1.19; $P = 1.2 \times 10^{-4}$). This association was not observed for male adenocarcinoma (n = 1507, 3 studies: EAGLE-BioVU, MEC, NCI; $P = .14$). However, the test for heterogeneity in effects between rs3817198 and adenocarcinoma by sex was not statistically significant ($P = .10$).

Whereas the *TERT* rs2853676 variant was only nominally associated with overall lung cancer ($P = .001$) (Supplementary Table 3, available online), a statistically significant association with adenocarcinoma (OR = 1.16; 95% CI = 1.10 to 1.22; $P = 1.1 \times 10^{-8}$) was observed among 5164 patients and 38 567 control subjects (Figure 3; Supplementary Table 5 and Supplementary Figure 1A, available online). This SNP was not associated with either SCC

or small cell carcinoma ($P \geq .18$) (P_{het} by cell type = 3.9×10^{-4}). In a subset of six studies with available data (IARC, MD Anderson Cancer Center (MDACC), MEC, NCI, Samuel Lunenfeld Research Institute study (SLRI), and WHI), when conditioning on the known *TERT* risk variant for lung cancer (rs2736100; linkage disequilibrium [LD] with rs2853676 in European CEU: $r^2 = 0.17$), the association with adenocarcinoma was attenuated (OR = 1.06; $P = .09$). Alternatively, the meta-analyzed result among these six studies when not conditioned on rs2736100 was similar to the main adenocarcinoma finding (OR = 1.16; $P = 1.3 \times 10^{-7}$).

The *CDKN2BAS1* glioma SNP, rs4977756, was not associated with overall lung cancer risk ($P = .13$) but was associated with SCC (OR = 1.11; 95% CI = 1.07 to 1.19; $P = 2.5 \times 10^{-5}$) (Figure 4; Supplementary Figure 1B, available online). This SNP was not associated with adenocarcinoma ($P = .68$) or small cell carcinoma ($P = .48$) (P_{het} by cell type = .0006) (Supplementary Table 5, available online). Independent effects between rs4977756 and the previously reported lung cancer risk variant in 9p21.3 (11) could not be determined because only a small subset of data on the later variant was available.

Among the 15 nominally statistically significant associations, only two associations were heterogeneous by smoking status (8q24 rs10090154 and 6p21.33 rs6457327) at $P_{het} < .05$ (Supplementary Table 7, available online).

Discussion

In this large meta-analysis of 18 023 lung cancer patients and 60 543 control subjects, we examined 165 established cancer risk variants (excluding lung cancer and smoking-related risk variants)

and their associations with lung cancer. This is the first study to systematically examine pleiotropic effects from risk variants identified in GWASs of other malignancies on the risk of lung cancer. We found that the breast cancer risk allele "C" of *LSP1* rs3817198 was associated with an increased risk of lung cancer.

LSP1 encodes the lymphocyte-specific protein 1, an F-actin bundling cytoskeletal protein. In GWAS, common variants in or near the gene have been associated with risk of breast cancer in women (20) and ulcerative colitis in men and women (21). This *LSP1* region is conserved in mice, and studies have found loss of heterozygosity in this region in breast and lung cancers (22,23). We found that this association was stronger in women for overall lung cancer and for adenocarcinoma. When stratifying on both histology and sex, we observed an association in women with adenocarcinoma but not in men with adenocarcinoma. Furthermore, epidemiologic studies of familial aggregation of cancers found an excess of breast cancer among relatives of nonsmokers with lung cancer (24) and relatives of early-onset lung cancer (25), suggesting a genetic susceptibility across these two cancers. To confirm that this association was not a result of excess breast cancer cases, we excluded lung cancer cases with previous history of breast cancer and obtained similar results. The underlying biological mechanism through which *LSP1* may influence cancer development remains to be elucidated. *LSP1* is expressed in lymphocytes, neutrophils, macrophages, and endothelial cells and may regulate neutrophil motility, adhesion to fibrinogen matrix proteins, and transendothelial migration (26).

Risk variants in or near the *TERT-CLPTM1L* locus have been associated with risk of several cancer sites (6), including adenocarcinoma of the lung (6,7,9,27). *TERT* encodes for telomerase

Glioma SNP *CDKN2BAS1* rs4977756, region 9p21.3

Study	Patients	Controls	OR	95% CI	P value
MEC	114	9647	0.97	0.73-1.28	.81
WHI	212	5826	1.12	0.91-1.37	.29
DeCode Genetics	179	11 226	1.15	0.92-1.43	.22
Harvard	215	970	1.05	0.83-1.33	.68
HGF Germany	99	478	0.93	0.61-1.43	.74
IARC GWAS	961	3240	1.13	1.01-1.26	.04
MDACC	306	1134	1.18	0.98-1.42	.09
NCI GWAS	1448	5662	1.15	1.05-1.25	.002
SLRI/Toronto	50	498	1.33	0.82-2.18	.25
Overall	3584	38 681	1.13	1.07-1.19	2.5×10^{-5}
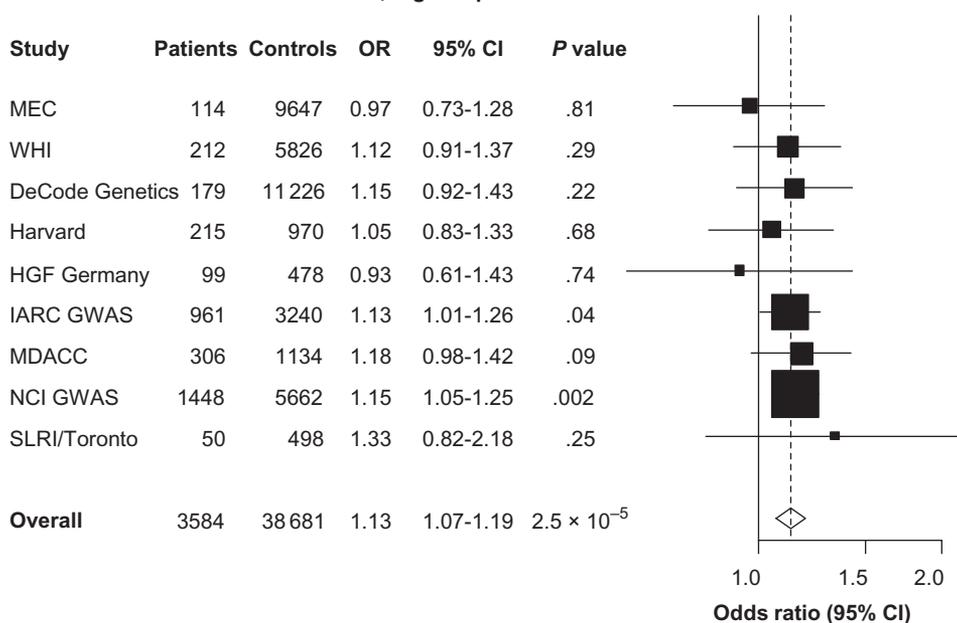
Figure 4. Forest plot of the association between cyclin-dependent kinase 4 inhibitor B antisense RNA 1 (*CDKN2BAS1*) rs4977756 and lung squamous cell carcinoma risk. Study specific and meta-analysis associations are plotted, modeling the G risk allele for glioma. Squares represent odds ratios (ORs); size of the square represents inverse of the variance of the log ORs; horizontal lines represent 95% confidence intervals (CIs); diamonds represent summary estimate combining the study-specific estimates with a fixed-effects model; solid vertical lines represent OR = 1; dashed vertical lines represent the overall ORs. The single-nucleotide polymorphism (SNP) rs4977756 was genotyped in all studies. GWAS = genome-wide association study.

reverse transcriptase, which maintains telomere length through each cell division. Telomere shortening is associated with increased genomic instability, thereby increasing the risk of cancer development. The "A" allele of rs2853676, located in intron 2 of *TERT*, was initially reported to be associated with an increased risk of glioma (28). In our study, we found a strongly statistically significant association with adenocarcinoma and notable heterogeneity by histological cell type. Consistent with our findings, the NCI study, which is part of TRICL, reported a modest association between rs2853676 and adenocarcinoma ($P = 3.4 \times 10^{-4}$) (7). This same study identified *TERT* rs2736100, also located in intron 2, to be associated with a 12% increase in lung cancer risk ($P = 1.6 \times 10^{-10}$) (7). Whereas rs2853676 is in low LD with rs2736100 (European CEU: $r^2 = .17$), results from our conditional analysis suggest that the association between rs2853676 and adenocarcinoma may not be independent of rs2736100. In addition, a recent Japanese study found that *TERT* rs2853677 (CEU: $r^2 = 0.59$) is associated with lung adenocarcinoma ($P = 3.1 \times 10^{-40}$) (29). However, because this SNP was not genotyped in our study, we were unable to condition on rs2853677. It is possible that the association between rs2853676 and adenocarcinoma may be influenced by rs2736100 and rs2853677.

We found that rs4977756 at 9p21.3 was associated with SCC. This SNP is located in *CDKN2BAS1*, a long noncoding RNA region, and near the cluster of two tumor suppressor genes, *CDKN2A* and *CDKN2B*. *CDKN2BAS1* has been implicated in the development of multiple chronic diseases and cancers, due to the role of *CDKN2A* and *CDKN2B* in cell cycle inhibition, senescence, and stress-induced apoptosis (30). Furthermore, three *CDKN2BAS1* spliced variant transcripts expressed in lung cancer cell lines (31) have been shown to have various enhancer activities (32). The SNP rs4977756 has been previously associated with glioma (28,33) and glaucoma (34). A recent meta-analysis of lung cancer GWASs by TRICL found rs1333040, which is approximately 74 kb upstream from *CDKN2B*, to be associated with lung cancer (OR = 1.06; $P = 9.4 \times 10^{-5}$), with a stronger association for SCC (OR = 1.14; $P = 2.9 \times 10^{-7}$) (11). Among European-ancestry populations, there is little LD between rs1333040 and rs4977756 (CEU + Toscans in Italy [TSI]: $r^2 = 0.27$). However, because only two studies had genotype data for rs1333040, we were unable to examine the independent effects of the two SNPs. Further evaluation of rs4977756 and SCC risk is needed.

Our finding of pleiotropy between the breast cancer risk locus at *LSP1* and lung cancer risk points toward shared etiologic mechanisms for these two cancer sites. Concurrently, we observed cell type–specific effects for lung cancer with two variants located in cancer pleiotropic regions (*TERT* and risk of lung adenocarcinoma and *CDKN2BAS1* with risk of lung SCC), indicating distinct etiological processes for these two subtypes. These observations of shared and distinct effects with particular genetic loci are consistent with other studies of lung cancer. For example, *EGFR* kinase domain mutations are frequent in lung adenocarcinoma of nonsmokers and extremely rare in lung SCC (35). Alternatively, the *EGFR* variant III mutations have been found in lung SCC and gliomas (36), but not in lung adenocarcinoma (35). These findings demonstrate the complexity of carcinogenesis and the need to study both shared and distinct etiological pathways.

Study limitations include reduced power to detect effects for some of the 165 SNPs. Nonetheless, 72% of the SNPs were genotyped in more than 50% of studies. Due to the limited size of the populations of non-European descent, we were unable to fully examine the generalizability of effects across these populations. Additionally, with the available data, we could only test in a subset of studies the independence of the *TERT* rs2853676 association from the previously reported *TERT* associations. Thus, the associations that we observed with *TERT* rs2853676 and *CDKN2BAS1* rs4977756 may reflect weak LD with previously identified lung cancer risk variants in these regions. However, because the functional SNPs for these regions remain unknown, our findings are informative for future studies (e.g., fine-mapping, functional and population-specific generalizability studies). Furthermore, we recognize the need to study the additional risk loci identified by more recent cancer GWASs. Last, as the majority of our controls excluded all cancers, there may have been a greater likelihood of observing associations with the cancer risk variants studied. However, in the MEC, where control selection allowed inclusion of subjects with other cancers than lung cancer, the associations for the top SNPs were consistent with the overall findings. Our study strengths include the systematic "candidate-SNP" approach based on strong prior evidence of an association from GWASs of cancer, the large sample size from well-characterized epidemiologic lung cancer studies, and the power to examine these associations by cell type, smoking status, and sex.

In conclusion, the breast cancer SNP *LSP1* rs3817198 was associated with lung cancer risk. Our results support the influence of non–lung cancer risk variants on the risk of lung cancer, and these associations may differ by histological cell type and sex. Molecular studies are needed to better characterize these genetic effects and to increase our understanding of the apparent heterogeneity of effects across sex and histological cell type.

References

1. Ferlay J, Shin H-R, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*. 2010;127(12):2893–2917.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin*. 2005;55(2):74–108.
3. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106(23):9362–9367.
4. Stacey SN, Sulem P, Jonasdottir A, et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet*. 2011;43(11):1098–1103.
5. Ghoussaini M, Song H, Koessler T, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst*. 2008;100(13):962–966.
6. Mocellin S, Verdi D, Pooley KA, et al. Telomerase reverse transcriptase locus polymorphisms and cancer risk: a field synopsis and meta-analysis. *J Natl Cancer Inst*. 2012;104(11):840–854.
7. 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(5):679–691.
8. Miki D, Kubo M, Takahashi A, et al. Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. *Nat Genet*. 2010;42(10):893–896.
9. Hsiung CA, Lan Q, Hong Y-C, et al. The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS Genet*. 2010;6(8):e1001051.
10. Truong T, Hung RJ, Amos CI, et al. Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled

analysis from the International Lung Cancer Consortium. *J Natl Cancer Inst*. 2010;102(13):959–971.

11. 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(22):4980–4995.

12. Matise TC, Ambite JL, Buyske S, et al. The next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. *Am J Epidemiol*. 2011;174(7):849–859.

13. Roden DM, Pulley JM, Basford MA, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther*. 2008;84(3):362–369.

14. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34(8):816–834.

15. Kosoy R, Nassir R, Tian C, et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation*. 2009;30(1):69–78.

16. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–909.

17. Rasmussen-Torvik LJ, Alonso A, Li M, et al. Impact of repeated measures and sample selection on genome-wide association studies of fasting glucose. *Genet Epidemiol*. 2010;34(7):665–673.

18. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164(4):1567–1587.

19. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190–2191.

20. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447(7148):1087–1093.

21. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet*. 2011;43(3):246–252.

22. Tran YK, Newsham IF. High-density marker analysis of 11p15.5 in non-small cell lung carcinomas reveals allelic deletion of one shared and one distinct region when compared to breast carcinomas. *Cancer Res*. 1996;56(13):2916–2921.

23. Bepler G, Fong KM, Johnson BE, et al. Association of chromosome 11 locus D11S12 with histology, stage, and metastases in lung cancer. *Cancer Detect Prev*. 1998;22(1):14–19.

24. Gorlova OY, Weng S-F, Zhang Y, Amos CI, Spitz MR. Aggregation of cancer among relatives of never-smoking lung cancer patients. *Int J Cancer*. 2007;121(1):111–118.

25. Schwartz AG, Siegfried JM, Weiss L. Familial aggregation of breast cancer with early onset lung cancer. *Genet Epidemiol*. 1999;17(4):274–284.

26. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: integrating information about genes, proteins and diseases. http://www.genecards.org/cgi-bin/carddisp.pl?gene=LSP1 Accessed March 1, 2013.

27. Broderick P, Wang Y, Vijayakrishnan J, et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res*. 2009;69(16):6633–6641.

28. Shete S, Hosking FJ, Robertson LB, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet*. 2009;41(8):899–904.

29. Shiraishi K, Kunitoh H, Daigo Y, et al. A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. *Nat Genet*. 2012;44(8):900–903.

30. Pasmant E, Sabbagh A, Vidaud M, Bièche I. ANRIL, a long, non-coding RNA, is an unexpected major hotspot in GWAS. *FASEB J*. 2011;25(2):444–448.

31. Folkersen L, Kyriakou T, Goel A, et al. Relationship between CAD risk genotype in the chromosome 9p21 locus and gene expression. Identification of eight new *ANRIL* splice variants. *PLoS One*. 2009;4(11):e7677.

32. Jarinova O, Stewart AF, Roberts R, et al. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol*. 2009;29(10):1671–1677.

33. Rajaraman P, Melin B, Wang Z, et al. Genome-wide association study of glioma and meta-analysis. *Hum Genet*. 2012;131(12):1877–1888.

34. Burdon KP, Macgregor S, Hewitt AW, et al. Genome-wide association study identifies susceptibility loci for open angle glaucoma at TMCO1 and CDKN2B-AS1. *Nat Genet*. 2011;43(6):574–578.

35. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med*. 2008;359(13):1367–1380.

36. Hatanpaa KJ, Burma S, Zhao D, Habib AA. Epidermal growth factor receptor in glioma: signal transduction, neuropathology, imaging, and radioresistance. *Neoplasia*. 2010;12(9):675–684.

Notes

Affiliations of authors: Department of Preventive Medicine, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA (SLP, FRS, CAH); Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA (MDF, JMK, AY, YL, UP, CC, CK); Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh and Medical Research Council Human Genetics Unit, Edinburgh, UK (MT); Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI (CPC, LRW, LL); Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH (YH, CIA); Department of Epidemiology (S-AL, AB, KEN, GH) and Carolina Center for Genome Sciences (KEN), University of North Carolina, Chapel Hill, NC; Molecular Physiology and Biophysics (LD, DCC), Center for Human Genetics Research (LD, RG, HD, DCC, WSB), Vanderbilt Epidemiology Center (JHF), and Biomedical Informatics (WSB), Vanderbilt University, Nashville, TN; Division of Genomic Medicine, National Human Genome Research Institute (LH), and Division of Cancer Epidemiology and Genetics, National Cancer Institute (DA, NC, MTL), National Institutes of Health, Bethesda, MD; Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ (SB); Information Sciences Institute, University of Southern California, Marina del Rey, CA (ED); Department of Oncology, School of Medicine, Wayne State University, Detroit, Michigan (MLC); M.D. Anderson Cancer Center, Houston, TX (WC, MP); Harvard University School of Public Health, Boston, MA (DCC); Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, The University of Liverpool Cancer Research Centre, Institute of Translational Medicine, The University of Liverpool, Liverpool, UK (JKF); Department of Genetic Epidemiology, University Medical Centre Göttingen, Göttingen, Germany (HB); DKFZ–German Cancer Research Center and Translational Lung Research Centre Heidelberg, Member of the German Center for Lung Research, Heidelberg, Germany (AR); Institute of Epidemiology I, Helmholtz Zentrum Munchen, Munich, Germany (JH); International Agency for Research on Cancer, Lyon, FR (PB); Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey, UK (YW, RSH); Cambridge University Health Partners, Cambridge, UK (TE); American Cancer Society, Epidemiology and Surveillance Research, Atlanta, GA (MT); Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Toronto, ON (RJH); Cancer Prevention Institute of California, Fremont, CA (IC).

*For the Transdisciplinary Research in Cancer of the Lung Research Team.