Published OnlineFirst April 26, 2012; DOI: 10.1158/1055-9965.EPI-11-1123-T

Cancer Epidemiology, Biomarkers & Prevention

Research Article

Association of the 15q25 and 5p15 Lung Cancer Susceptibility Regions with Gene Expression in Lung Tumor Tissue

Gord Fehringer[1], Geoffrey Liu[2,4], Melania Pintilie[3], Jenna Sykes[3], Dangxiao Cheng[2], Ni Liu[2], Zhuo Chen[2], Lesley Seymour[7], Sandy D. Der[2], Frances A. Shepherd[6], Ming-Sound Tsao[2,5], and Rayjean J. Hung[1]

Abstract

Background: Genome-wide association studies have identified two independent lung cancer susceptibility loci at chromosome 15q25 and one locus at 5p15. We examined the association of genetic variants in these regions with gene expression in lung tumor tissue, in an effort to elucidate carcinogenic mechanisms by which these variants influence lung cancer risk.

Methods: We used data from 2 independent studies of non–small cell lung carcinoma patients: the JBR.10 clinical trial ($n = 131$) and a University Health Network (UHN) patient sample in Toronto ($n = 181$). We genotyped seven 15q25 and five 5p15 variants and examined their association with expression profiles of genes in the corresponding regions, measured by Affymetrix HG-U133A.

Results: The minor allele (C) of a variant representing one of the two loci at 15q25 (rs2036534) was associated with increased iron-responsive element binding protein 2 (*IREB2*) expression in both studies (JBR.10 $P = 0.042$; UHN $P = 0.002$). A false discovery rate of 0.05 or less in the UHN sample increased our confidence in this association. The association appears to be more prominent among lung adenocarcinoma patients. We did not detect an association between genotype and expression profile for the other 15q25 locus or for 5p15 variants.

Conclusions: In contrast to previous studies that indicate 15q25 variants are associated with lung cancer risk through an effect on smoking behavior, our results suggest these variants may influence risk through a second mechanism, involving modulation of *IREB2* expression.

Impact: This finding expands on potential mechanisms through which 15q25 variants influence lung cancer risk and may have implications for future research on chemoprevention strategies. *Cancer Epidemiol Biomarkers Prev; 21(7); 1097–104. ©2012 AACR.*

Introduction

Lung cancer, one of the most common cancers worldwide, with approximately 1.6 million cases diagnosed each year, ranks first as a cause of annual global cancer deaths. The 5-year survival rate remains low, at about 15% (1). The primary cause of lung cancer is tobacco smoking (2), but inherited genetic variations also influence lung cancer risk. Genome-wide association studies (GWAS) have identified lung cancer susceptibility regions at chromosomes 15q25 and 5p15 (3–6). Currently, it is not clear how variants in these regions influence risk. Risk-associated variants at 15q25 lie in a region of strong linkage disequilibrium (LD) that comprises several genes including the nicotinic acetylcholine receptor genes (*CHRNB4*, *CHRNA5*, *CHRNA3*, *LOCI123688*, *PSMA4*, and *IREB2*). Because the variants at 15q25 are also associated with nicotine dependence (7–10), it has been suggested they influence lung cancer risk at least, in part, through an effect on smoking behavior. Risk-associated variants at 5p15 are also found in a region of high LD and localize to the *TERT* and *CLPTM1L* genes. *TERT* is expressed in 80% of non–small cell lung carcinoma (NSCLC) and may influence carcinogenesis through its role in telomere maintenance. *CLPTM1L* may play a role in carcinogenesis through an influence on apoptosis (11–14).

In an effort to elucidate the carcinogenic mechanisms by which genetic variants at these regions influence disease risk, we examined the association of genetic variants in lung cancer susceptibility regions with gene expression in tumor tissue of NSCLC patients. Because the main objective was to investigate whether expression of specific genes in these regions mediated the observed associations

Authors' Affiliations: [1]Prosserman Centre for Health Research, Samuel Lunenfeld Research Institute, Mount Sinai Hospital; [2]University Health Network, Ontario Cancer Institute/Princess Margaret Hospital; [3]Department of Biostatistics, Princess Margaret Hospital; Departments of [4]Medicine and [5]Laboratory Medicine and Pathobiology, University of Toronto; [6]Department of Medical Oncology and Hematology, University Health Network, Princess Margaret Hospital, Toronto; and [7]National Cancer Institute-Canada Clinical Trials Group, Queens University, Kingston, Ontario, Canada

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Author: Rayjean J. Hung, Samuel Lunenfeld Research Institute of Mount Sinai Hospital, 60 Murray St., Toronto, M5T 3L9, Canada. Phone: 416-586-4750; Fax: 416-586-8404; E-mail: Rayjean.hung@lunenfeld.ca

doi: 10.1158/1055-9965.EPI-11-1123-T

Published OnlineFirst April 26, 2012; DOI: 10.1158/1055-9965.EPI-11-1123-T

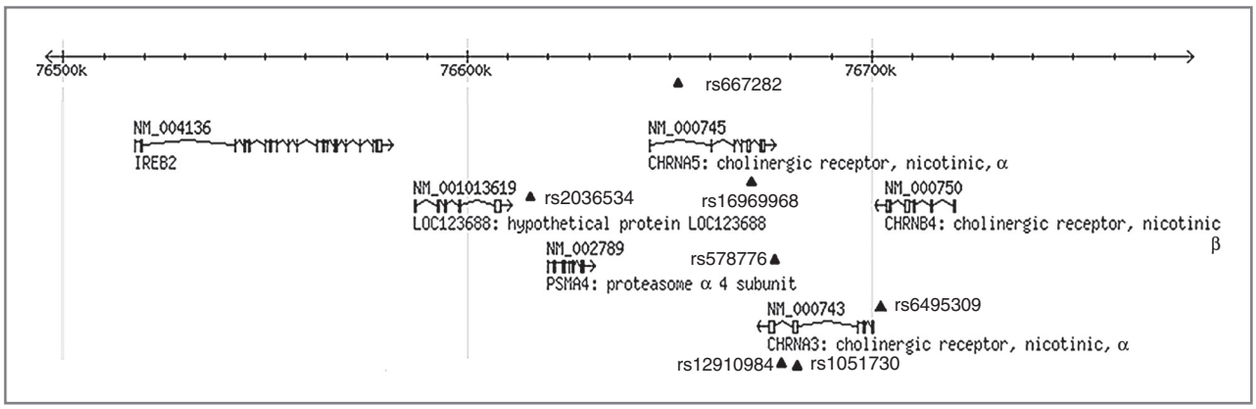

Figure 1. 15q25 region showing genes and locations of genotyped lung cancer risk variants.

with lung cancer risk, we specifically focused on genetic variants that produced the strongest associations in GWAS and related studies (3, 5, 15, 16). We chose 7 genetic variants at 15q25 to represent the 2 loci known to be associated with disease at this region (2 variants tagged loci 1 and 5 tagged loci 2) and 5 genetic variants to represent 5p15. We used data from 2 independent studies, comparing associations across studies to assess significance after accounting for multiple comparisons.

Materials and Methods

To investigate the association between the genetic variants and gene expression in the corresponding regions, we used data from 2 independent studies: the JBR.10 clinical trial and a patient sample from the University Health Network (UHN) in Toronto. We genotyped seven 15q25 and five 5p15 variants (see Figs. 1 and 2 for variant locations) that were identified by GWAS as associated with lung cancer risk and examined their association with the expression profiles of genes in the corresponding regions measured by Affymetrix HG-U133A. Detailed methods describing recruitment and gene expression microarray profiling for JBR.10 study subjects have been described elsewhere (17, 18). Relevant aspects of methods used for the JBR.10 study, and methods for the UHN study, are described here.

Study subjects

The JBR.10 study was a randomized trial of patients with stage IB or II NSCLC assigned to vinorelbine plus cisplatin (adjuvant treatment arm, $n = 242$) or to observation (no adjuvant treatment arm, $n = 240$; ref. 17). Eligible patients were those >18 years of age with completely resected stage IB or II NSCLC, with an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1. Subject recruitment began in April of 1994 and the study was closed as of April 2004 (17). Gene expression profiling using lung tumor tissue samples was completed for 133 patients, all with tumor cellularity greater than 20% (18). DNA, which was obtained primarily from normal tissue, was not available for 2 of these samples leaving 131 JBR.10 patients with both gene expression profiles and DNA available for genotyping.

The UHN NSCLC study included patients with stage I or II primary NSCLC retrospectively identified from the UHN tumor bank. These patients had undergone surgery for their cancer at the UHN between 1996 and 2004 but had not received chemotherapy. Only patients with tissue samples of 20% or greater tumor cellularity were included in the study. A total of 190 patients met these criteria and gene expression profiles were obtained for 183 of these (RNA quality was unsuitable for expression profiling for the rest). DNA (obtained mainly from normal tissue) was unavailable for 2 patients, leaving 181 UHN patients with expression profiles and genotype data.

Clinical data, including information on age, sex, smoking status, stage, and histology, was available for both studies.

Laboratory

The Affymetrix HG-U133A GeneChip was used to obtain gene expression profiles for tumor tissue from

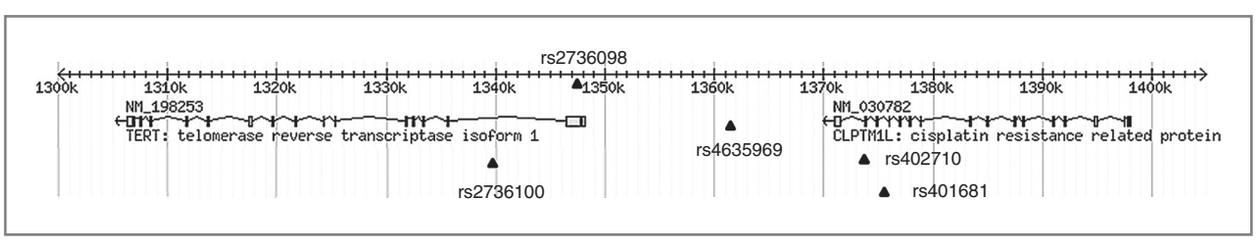

Figure 2. 5q15 region showing genes and locations of genotyped lung cancer risk variants.

Published OnlineFirst April 26, 2012; DOI: 10.1158/1055-9965.EPI-11-1123-T

Table 1. Variants genotyped, closest gene, assay and genotyping success rate for JBR.10 and UHN patients

dbSNP # polymorphism ID	Type	Gene symbol	Assay	Number (% genotyped) JBR.10	UHN
rs16969968	SNP	*CHRNA5*	Sequenom	129/131 (98.5)	173/181 (95.6)
rs6495309	SNP	*CHRNA3*	Sequenom	129/131 (98.5)	174/181 (96.1)
rs1051730	SNP	*CHRNA3*	Sequenom	129/131 (98.5)	174/181 (96.1)
rs12910984	SNP	*CHRNA3*	Sequenom	129/131 (98.5)	176/181 (97.2)
rs578776	SNP	*CHRNA3*	Sequenom	124/131 (94.7)	179/181 (98.9)
rs667282	SNP	*CHRNA5*	Sequenom	129/131 (98.5)	177/181 (97.8)
rs2036534	SNP	*LOCI123688*	Sequenom	129/131 (98.5)	172/181 (95.0)
rs401681	SNP	*CLPTM1L*	Sequenom	129/131 (98.5)	174/181 (96.1)
rs4635969	SNP	*CLPTM1L*	Sequenom	125/131 (95.4)	166/181 (91.7)
rs402710	SNP	*CLPTM1L*	TaqMan	131/131 (100.0)	180/181 (99.4)
rs2736098	SNP	*TERT*	Sequenom	128/131 (97.7)	174/181 (96.1)
rs2736100	SNP	*TERT*	Sequenom	131/131 (100.0)	175/181 (96.7)

JBR.10 patients following RNA extraction from frozen tissue [details are available from a previous publication (18)]. The chip provided expression profiles for all genes in the regions of interest (see Figs. 1 and 2) except *LOCI123688* and *CLPTM1L* for which probe sets are not available. DNA used for genotyping of JBR.10 subjects was extracted from normal snap-frozen or formalin-fixed and embedded lung tissue using phenol chloroform isoamyl alcohol and genotyped using Sequenom MassArray or TaqMan assay (Table 1 provides a list of genetic variants genotyped, assays used, and genotyping success rate for each variant). For one patient sample, tumor tissue DNA was used for genotyping some variants, as genotyping in normal tissue failed.

For the UHN study, RNA isolation from frozen tissue was carried out with guanidinium thiocyanate-phenol-chloroform reagent. DNA was extracted from normal tissue formalin-fixed, paraffin-embedded samples with chloroform isoamyl alcohol. Expression profiling (including use of the Affymetrix HG-U133A platform) and genotyping for UHN samples (Table 1) were identical to methods used for JBR.10 samples. For 2 patients, normal tissue was not available and DNA from tumor tissue was used for genotyping instead.

Data analysis

Microarray data were preprocessed with RMAexpress (version 0.3; ref. 19) for the JBR.10 study and Affymetrix R/Bioconductor (version 2.8.1) for UHN data. For the JBR.10 study, probe sets were annotated with the NetAffx version 4.2 annotation tool and only probe sets with grade A annotation (NA22; ref. 20), corresponding to the genes of interest, were included. Corresponding probe sets were selected for the UHN study. Distance weighted discrimination (21) was used to adjust systematic differences found between JBR.10 batches identified by unsupervised heuristic K-means clustering (Genesis version 1.7.5). Expression data for both studies were transformed to a Z-score by centering to the mean and scaling to the SD.

All genes were profiled with a single probe set except for *CHRNA3*, which was profiled with 3 probe sets. Pairs of probe sets for *CHRNA3* were positively correlated ($r \geq 0.34$), and specificity of probe sets for this gene was high (≥ 0.95). Therefore, the mean of expression results of all 3 probe sets was used to represent expression for this gene in primary analyses, followed by a generalized estimating equation (GEE) approach to verify our results. Multivariate linear regression models were used to analyze the associations between genetic variants and expression of genes in the same region (e.g., 15q25 variants with 15q25 genes) under an additive model using SAS 9.2 (SAS Institute Inc.). This resulted in 40 statistical tests for association. Statistical analyses were first conducted for all patients combined and then stratified by histology (adenocarcinoma and nonadenocarcinoma). Covariates included in models were (i) age (modeled as a continuous variable) and sex and (ii) age, sex, and stage. Because including stage in the model had little influence on regression coefficients and *P* values, results for models with age and sex are presented here. Genotype was treated as a continuous variable with 0, 1, and 2 representing number of copies of the minor allele. We also conducted stratified analyses by sex and smoking status (current, former, and never) focusing on models in which iron-responsive element binding protein 2 (*IREB2*) expression was the outcome. The smoking stratification was carried out in UHN only, as the number of never-smokers in JBR10 was not sufficient to conduct a meaningful analysis.

The association of 15q25 haplotypes with *IREB2* expression was examined with the haplotype-based association test with GLMs [PLINK version 1.06 (22)]. Genotyped SNPs that best tagged HapMap database SNPs in *IREB2* (*IREB2* SNPs were not genotyped as we had focused on SNPs that best captured associations with lung cancer in

Published OnlineFirst April 26, 2012; DOI: 10.1158/1055-9965.EPI-11-1123-T

Table 2. Comparison of JBR.10 and UHN patient characteristics

	JBR.10 N (%)	UHN N (%)
Number of patients	131	181
Sex		
Male	89 (67.9)	98 (54.1)
Female	42 (32.1)	83 (45.9)
Age		
Mean (range)	60.6 (35–81)	68.8 (40–88)
Histology age		
Adenocarcinoma	71 (54.2)	128 (70.7)
Squamous cell carcinoma	50 (38.2)	43 (23.8)
Other	19 (7.6)	10 (5.5)
Stage		
1	73 (55.7)	128 (70.7)
2	58 (44.3)	53 (29.3)
Smoking status		
Current	8 (6.1)	58 (32.0)
Ex-smoker	117 (89.3)	78 (43.1)
Never	5 (3.8)	25 (13.8)
Unknown	1 (0.8)	20 (11.0)

GWAS) were selected to construct haplotypes using the Tagger program (23) incorporated into Haploview version 4.2.

An association was considered to be robust if it reached the conventional level of significance of $P \leq 0.05$ in JBR.10, was consistent for direction of effect in both studies, and achieved an false discovery rate (FDR) of $\leq$0.05 in the larger UHN study [using the Benjamini-Hochberg-Yekutieli procedure for controlling the FDR under dependence assumptions (24)]. This assessment was made for primary analyses only [main genotype effects for all NSCLC patients combined (i.e., all histologies)].

Results

The basic characteristics of the 2 study populations are summarized in Table 2. Most patients were male (JBR.10: 68%; UHN: 54%) and were current or former smokers (JBR.10: 95%, UHN: 75%). Adenocarcinoma was the predominant histology (JBR.10: 54% UHN: 71%). Mean age was 60.6 in JBR.10 and 68.8 in the UHN study (Table 2).

Linear regression analysis testing the association between genetic variants and tumor tissue gene expression found the minor allele of variant rs2036534 (C allele) to be associated with higher expression of *IREB2* in both studies ($P \leq 0.05$ in each study; Table 3). The other gene expression–variant associations did not replicate across the 2 studies. The FDR for the association between *IREB2* and rs2036534 was less than 0.05 [Benjamini-Hochberg-Yekutieli procedure (24)] in the larger UHN sample, providing confidence that this association is not likely to be a false positive finding resulting from multiple statistical tests. In additional analyses using a GEE approach to account for multiple probe sets (*CHRNA3* gene only), we found nearly identical results relative to main analyses in which we averaged the 3 probe sets (e.g., UHN rs16969968: $\beta = -0.06$, $P = 0.45$).

Before accounting for multiple comparisons (i.e., based on $P \leq 0.05$), all of the other 15q25 variants (in addition to rs2036534) were associated with *IREB2* expression in the UHN study but not in the JBR.10 study (Table 3). Variants at 15q25 represent 2 distinct LD bins [which account for 2 lung cancer association signals in this region (4)], with one bin tagged by rs16969968 and rs1051730 (denoted as Bin 1) and the other by rs6495309, rs12910984, rs578776, rs667282, and rs2036534 (denoted as Bin 2). The minor alleles for variants in Bin 2 [which are associated with reduced lung cancer risk (15)] were associated with higher *IREB2* expression in the UHN study, consistent with the observed direction of effect for rs2036534. Minor alleles for Bin 1 variants [associated with increased lung cancer risk (5)] were associated with lower expression in the UHN study. Because Bin 1 and Bin 2 variants are in weak LD, we conducted further analyses to explore whether the association of Bin 1 variants with *IREB2* expression in the UHN study might be explained by Bin 2 variants, as represented by rs2036534. Regression models that included rs2036534 with either of the Bin 1 variants resulted in loss of significance and marked reduction in strength of effect at Bin 1 [e.g., rs16969968: the regression coefficient changed from −0.25 (Table 3) to −0.12, *P* value increased from 0.018 (Table 3) to 0.31], while change in the effect of rs2036534 was minimal [the regression coefficient decreased from 0.40 (Table 3) to 0.33; *P* value retained significance ($P = 0.017$)]. This result indicated that Bin 2 variants best accounted for the association between genetic variation at 15q25 and *IREB2* expression in this region.

Results of analyses examining the association of 15q25 variants with *IREB2* expression in lung adenocarcinoma patients are shown in Table 4. We found that all 15q25 Bin 2 variants were associated with *IREB2* expression in both the JBR.10 and UHN studies in this subgroup. In addition, the association between rs2036534 and *IREB2* expression became stronger with a higher regression coefficient and lower *P* value. Although Bin 1 variants also showed a statistically significant association with *IREB2* expression in the UHN study, significance was again lost after inclusion of the Bin 2 variant rs2036534 into the regression model (data not shown). Among nonadenocarcinoma patients, none of the 15q25 variants showed a significant association with *IREB2* expression in either JBR.10 or UHN patients (Supplementary Table S1). However, reduced power due to smaller sample size limited our chances of seeing significant associations in this subgroup.

The relationship between 15q25 Bin 2 variants and *IREB2* gene expression was explored further in analyses stratified by sex and smoking status. The trend for an association of the minor allele with higher *IREB2* expression was apparent in both sexes, with strength of

Published OnlineFirst April 26, 2012; DOI: 10.1158/1055-9965.EPI-11-1123-T

Table 3. Association of variants with gene expression in JBR.10 and UHN NSCLC patients

Region/gene	15q25 Variant (allele designation)[a]		JBR.10 β	JBR.10 *P*	JBR.10 *N*	UHN β	UHN *P*	UHN *N*
15q25								
CHNRB4	rs16969968	(G,A)	−0.084	0.533	129	0.021	0.843	173
	rs6495309	(C,T)	−0.081	0.631	129	−0.314	0.011[b]	174
	rs1051730	(C,T)	−0.084	0.533	129	0.019	0.860	174
	rs12910984	(A,G)	−0.107	0.545	129	−0.339	0.005[b]	176
	rs578776	(C,T)	−0.126	0.411	124	−0.186	0.057	179
	rs667282	(T,C)	−0.033	0.852	129	−0.311	0.009[b]	177
	rs2036534	(T,C)	−0.165	0.313	129	−0.237	0.057	172
CHRNA3	rs16969968	(G,A)	0.068	0.621	129	−0.059	0.457	173
	rs6495309	(C,T)	0.046	0.788	129	−0.117	0.219	174
	rs1051730	(C,T)	0.068	0.621	129	−0.040	0.624	174
	rs12910984	(A,G)	0.022	0.903	129	−0.148	0.112	176
	rs578776	(C,T)	−0.113	0.476	124	−0.074	0.324	179
	rs667282	(T,C)	0.045	0.801	129	−0.185	0.041[b]	177
	rs2036534	(T,C)	−0.052	0.755	129	−0.140	0.146	172
CHRNA5	rs16969968	(G,A)	−0.089	0.506	129	0.199	0.059	173
	rs6495309	(C,T)	0.183	0.273	129	−0.114	0.366	174
	rs1051730	(C,T)	−0.089	0.506	129	0.193	0.070	174
	rs12910984	(A,G)	0.370	0.031[b]	129	−0.128	0.299	176
	rs578776	(C,T)	0.246	0.107	124	−0.122	0.213	179
	rs667282	(T,C)	0.327	0.053	129	−0.144	0.236	177
	rs2036534	(T,C)	0.275	0.089	129	−0.151	0.232	172
PSMA4	rs16969968	(G,A)	−0.241	0.075	129	0.322	0.002[b]	173
	rs6495309	(C,T)	0.056	0.744	129	−0.047	0.710	174
	rs1051730	(C,T)	−0.241	0.075	129	0.306	0.004[b]	174
	rs12910984	(A,G)	0.123	0.490	129	−0.083	0.495	176
	rs578776	(C,T)	0.314	0.049[b]	124	−0.127	0.192	179
	rs667282	(T,C)	0.157	0.375	129	−0.053	0.655	177
	rs2036534	(T,C)	0.170	0.305	129	−0.075	0.551	172
IREB2	rs16969968	(G,A)	−0.117	0.391	129	−0.249	0.018[b]	173
	rs6495309	(C,T)	0.173	0.311	129	0.366	0.004[b]	174
	rs1051730	(C,T)	−0.117	0.391	129	−0.241	0.024[b]	174
	rs12910984	(A,G)	0.254	0.150	129	0.361	0.003[b]	176
	rs578776	(C,T)	0.226	0.154	124	0.244	0.013[b]	179
	rs667282	(T,C)	0.243	0.169	129	0.401	0.001[b]	177
	rs2036534	(T,C)	0.335	0.042[b]	129	0.399	0.002[b]	172
5p15								
TERT	rs401681	(C,T)	0.069	0.601	129	0.043	0.658	174
	rs4635969	(C,T	0.197	0.271	125	0.113	0.437	166
	rs402710	(C,T)	−0.103	0.455	131	0.089	0.383	180
	rs2736098	(G,A)	0.081	0.541	128	0.063	0.592	174
	rs2736100	(G,T)	−0.335	0.008[b]	131	0.013	0.909	175

[a]Major allele, minor allele.
[b]Statistically significant at $P \leq 0.05$.
β, regression coefficient.

association for Bin 2 variants somewhat stronger in women in both the JBR.10 and UHN studies [(e.g., for rs2036534 males: JBR.10 ($\beta = 0.26$, $P = 0.18$, $n = 87$), UHN ($\beta = 0.34$, $P = 0.05$, $n = 92$); females: JBR.10 ($\beta = 0.59$, $P = 0.07$, $n = 42$), UHN ($\beta = 0.46$, $P = 0.01$, $n = 80$)]. There was no compelling evidence indicating that smoking status modified the association between genotype and gene expression [e.g., for rs2036534 in the UHN data set—current smokers ($\beta = 0.45$, $P = 0.06$, $n = 56$), former smokers ($\beta = 0.26$, $P = 0.18$, $n = 76$), never-smokers ($\beta = 0.57$, $P = 0.12$, $n = 23$)].

Published OnlineFirst April 26, 2012; DOI: 10.1158/1055-9965.EPI-11-1123-T

Table 4. Association of 15q25 variants with *IREB2* gene expression in JBR.10 and UHN adenocarcinoma patients

Gene	15q25 Variant (allele designation)[a]		JBR.10 β	JBR.10 *P*	JBR.10 *N*	UHN β	UHN *P*	UHN *N*
IREB2	rs16969968	(G,A)	−0.097	0.656	70	−0.298	0.019[b]	125
	rs6495309	(C,T)	0.572	0.033[b]	70	0.416	0.004[b]	126
	rs1051730	(C,T)	−0.097	0.656	70	−0.287	0.025[b]	126
	rs12910984	(A,G)	0.607	0.021[b]	69	0.406	0.004[b]	126
	rs578776	(C,T)	0.520	0.026[b]	66	0.321	0.005[b]	128
	rs667282	(T,C)	0.602	0.025[b]	69	0.443	0.001[b]	127
	rs2036534	(T,C)	0.639	0.009[b]	70	0.471	0.001[b]	123

[a]Referent allele, minor allele.

[b]Statistically significant at $P \leq 0.05$.

β, regression coefficient.

None of the genotyped SNPs were in *IREB2*. In fact, rs2036534 was the closest variant lying between *LOCI123688* and *PSMA4*, 35,481bp 3′ to *IREB2* (Fig. 1). To capture potential unknown variants near *IREB2*, we constructed haplotypes from rs2036534-rs578776 and rs1051730-rs2036534 and tested their association with *IREB2* gene expression. On the basis of HapMap data, these haplotypes appeared to improve tagging of common SNPs at *IREB2* relative to the 7 SNPs genotyped at 15q25 [15 of 31 SNPs tagged at $R^2 = 0.65$ by haplotypes (including 5′ and 3′ *IREB2* SNPs) vs. 11 of 31 using SNPs only]. After analysis using haplotypes, we found significant associations for specific haplotypes in the UHN study (Supplementary Table S2), but not the JBR.10 study. An omnibus test for haplotype association (in which all haplotypes are represented by a single variable) provided similar results, with significant associations found for UHN patients only (JBR.10: rs2036534-rs578776 $P = 0.270$, rs1051730-rs2036534 $P = 0.121$; UHN: rs2036534-rs578776 $P = 0.008$, rs1051730-rs2036534 $P = 0.004$). These results indicate that the association between the 15q25 region and *IREB2* expression is best captured by rs2036534, as opposed to the haplotypes we constructed to capture genetic variation at *IREB2*.

Discussion

In this study, we examined the association of heritable genetic variants with gene expression in NSCLC tumor tissue for genes at 15q25 and 5p15 that were shown to be associated with lung cancer risk in GWAS (3–6). We used 2 independent studies in our analyses: the JBR.10 clinical trial and a sample of patients treated at UHN (Toronto). Our main finding indicates that Bin 2 variants, but not Bin 1 variants, in the 15q25 lung cancer susceptibility region are associated with *IREB2* expression. We consider the evidence for this conclusion to be robust as we found a significant association (at $P \leq 0.05$) between the minor allele (C allele) of Bin 2 variant rs2036534 with greater lung tumor *IREB2* expression in each of our 2 independent study samples, and our adjustment for multiple comparisons using the Benjamini-Hochberg-Yekutieli procedure resulted in a significant association in the UHN sample while controlling the FDR at a stringent level (FDR ≤ 0.05). We found stronger associations in adenocarcinoma patients but cannot rule out associations among nonadenocarcinoma patients due to small sample size.

Our findings suggest that there are at least 2 carcinogenic mechanisms that explain how 15q25 variants influence lung cancer risk. So far, the nicotinic acetylcholine receptor genes (*CHRNA3*, *CHRNA5*, and *CHRNB4)* have appeared to be the best candidates to explain the association as previous studies reported Bin 1 and Bin 2 variants to be independently associated with nicotine dependence (7–10). Indeed, it has been argued that smoking behavior accounts for all of the association between genetic variation at this region and lung cancer risk (25). Other investigators, however, report residual association between 15q25 variants and lung cancer risk remains after adjustment for smoking (7, 26), which could be explained by an additional direct effect of variants in this region on lung carcinogenesis (26). Our results suggest that the residual variation not accounted for by smoking may be explained by the influence of genetic variants at 15q25 on *IREB2* expression.

A relationship between *IREB2* expression and NSCLC is biologically plausible. *IREB2* is an RNA binding protein that plays a key role in regulating iron homeostasis (27). High levels of iron in the lungs may produce oxidative stress leading to inflammation and increased lung cancer risk (28). We hypothesize that the increased expression of *IREB2* associated with the minor allele of rs2036534 produces a favorable response to iron accumulation in the lung, reducing oxidative stress that precipitates inflammation, thus reducing lung cancer risk.

Most of the variants chosen for this study have no known functional significance, meaning functional

Published OnlineFirst April 26, 2012; DOI: 10.1158/1055-9965.EPI-11-1123-T

variants that are in LD with Bin 2 markers are most likely to explain the observed association with *IREB2* expression. At the outset of the study, it was not known whether expression of specific genes at the 15q25 or 5p15 regions might explain associations of genetic variants and lung cancer risk. We therefore decided to first attempt to establish associations between markers strongly associated with risk in GWAS and gene expression, which would enable further exploration of the region focusing on functional variants that may explain these associations.

Previous work by Falvella and colleagues found an inverse association between the minor allele of Bin 1 variant rs16969968 and *CHRNA5* expression based on 69 normal lung tissue samples from adenocarcinoma patients (29), while Wu and colleagues found an association of the common allele of rs6495309 and higher *CHRNA3* expression in 55 normal lung tissue samples from lung cancer patients (16). Similar associations were not found in this study, although it is possible that this is due to the use of expression profiles obtained from normal tissue in the studies discussed above, whereas we obtained expression profiles from tumor tissues. Investigation of the association of 15q25 Bin 2 variants and *IREB2* expression in paired normal and tumor tissue can provide insight on whether genetic variation modulates expression in lung tissue before transformation. To date, there has been no previous study examining the association between Bin 2 variants and *IREB2* gene expression in lung tissue. Future work could use a fine mapping approach to determine whether variants closer to *IREB2* are more strongly associated with *IREB2* expression. This approach would be particularly useful in African-Americans who exhibit lower levels of LD, thus permitting better localization of association signals. Fine mapping studies that examine the association of SNPs in this region with risk in African-Americans have already been undertaken (30, 31). In addition, resequencing of this region would further reveal its genetic architecture and may lead to detection of *de novo* variants at or near *IREB2* that are causally related to expression of this gene.

A potential limitation of this study is that undetected confounding due to population stratification may have influenced our results, as ethnicity information was not available for JBR.10 and UHN patients. However, given that the observed allele frequencies are consistent with the HapMap CEU population (data not shown), we do not consider population stratification to be a major concern (32). Another potential limitation is that measurement of gene expression using the Affymetrix HG-U133A GeneChip may be biased due to the presence of G-quadruplex structures which form in the presence of sequential runs (4 or more) of guanines in probes of some probe sets (33). The *IREB2* probe set does not have these sequential runs of guanine so measurement of gene expression will not be directly affected. Still, the presence of these runs of guanine in other probe sets could bias *IREB2* gene expression measurement following correction of background noise and normalization. However, this bias is likely to be nondifferential (i.e., bias toward the null) which does not alter our conclusions.

In summary, we have shown an association between *IREB2* expression and 15q25 Bin 2 variants (best represented by rs2036534) in patients from 2 independent studies. Our finding suggests that in addition to smoking behavior, there may be a second mechanism that explains the association between the 15q25 susceptibility region and lung cancer risk, operating through modulation of *IREB2* gene expression. This finding may have implications on future research of lung cancer chemoprevention strategies such as using iron chelators to compensate for exposure of lung tissue to iron.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: G. Liu, M.-S. Tsao, R. Hung
Development of methodology: G. Liu, D. Cheng, N. Liu, R. Hung
Acquisition of data: D. Cheng, Z. Chen, L. Seymour, S.D. Der, M.-S. Tsao,
Analysis and interpretation of data: G. Fehringer, G. Liu, M. Pintilie, J. Sykes, L. Seymour, F.A. Shepherd, R. Hung
Writing, review, and/or revision of the manuscript: G. Fehringer, G. Liu, M. Pintilie, J. Sykes, L. Seymour, F.A. Shepherd, M.-S. Tsao, R. Hung
Administrative, technical, or material support: G. Fehringer, N. Liu,
Study supervision: G. Fehringer, G. Liu, R. Hung

Grant Support

This work was supported by Samuel Lunenfeld New Opportunity grants (Dr. Hung) and grants from the Canadian Cancer Society Research Institute (#020214 to Dr. Hung and #020527 to Dr. Tsao), and the Allan Brown Chair in Molecular Genomics (Dr. Liu). Partial funding for the UHN study was provided by Med Biogene Inc. Dr. Tsao is the M. Qasim Choksi Chair in Lung Cancer Translational Research and Dr. Shepherd is the Scott Taylor Chair in Lung Cancer Research. Works carried out at the UHN and Ontario Cancer Institute/Princess Margaret Hospital was supported in part by the Ontario Ministry of Health and Long Term Care.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 30, 2011; revised April 11, 2012; accepted April 17, 2012; published OnlineFirst April 26, 2012.

References

1. IARC. Globocan 2008, 2010.
2. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. World Cancer Research Fund, American Institute for Cancer Research; 2007.
3. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 2008;40:616–22.
4. Broderick P, Wang Y, Vijayakrishnan J, Matakidou A, Spitz MR, Eisen T, et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. Cancer Res 2009;69:6633–41.
5. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 2008;452:633–7.

Published OnlineFirst April 26, 2012; DOI: 10.1158/1055-9965.EPI-11-1123-T

6. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 2008;452:638–42.
7. Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, et al. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. PLoS Genet 2010;6. pii: e1001053
8. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 2010;42:441–7.
9. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet 2010;42:436–40.
10. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet 2010;42:448–53.
11. Duarte RL, Paschoal ME. Molecular markers in lung cancer: prognostic role and relationship to smoking. J Bras Pneumol 2006;32:56–65.
12. Lantuejoul S, Salon C, Soria JC, Brambilla E. Telomerase expression in lung preneoplasia and neoplasia. Int J Cancer 2007;120:1835–41.
13. Yamamoto K, Okamoto A, Isonishi S, Ochiai K, Ohtake Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. Biochem Biophys Res Commun 2001;280:1148–54.
14. Zienolddiny S, Skaug V, Landvik NE, Ryberg D, Phillips DH, Houlston R, et al. The TERT-CLPTM1L lung cancer susceptibility variant associates with higher DNA adduct formation in the lung. Carcinogenesis 2009;30:1368–71.
15. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, 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;85:679–91.
16. Wu C, Hu Z, Yu D, Huang L, Jin G, Liang J, et al. Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations. Cancer Res 2009;69:5065–72.
17. Winton T, Livingston R, Johnson D, Rigas J, Johnston M, Butts C, et al. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. N Engl J Med 2005;352:2589–97.
18. Zhu CQ, Ding K, Strumpf D, Weir BA, Meyerson M, Pennell N, et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. J Clin Oncol 2010;28:4417–24.
19. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185–93.
20. Affymetrix: Transcript assignment for NetAffxTM annotation, Affymetrix GeneChip IVT array whitepaper collection. 2006. Santa Clara, CA: Affymetrix Inc; 2006. Available from: http://media.affymetrix.com/support/technical/whitepapers/netaffxannot_whitepaper.pdf
21. UNC Microrray database. Chapel Hill, NC: The University of North Carolina at Chapel Hill; 2012. Available from: https://genome.unc.edu/
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
23. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217–23.
24. Benjamini Y, Yekutiele D. On the control of the false discovery rate in multiple testing under dependency. Ann Statist 2001;29:1165–88.
25. Wang Y, Broderick P, Matakidou A, Eisen T, Houlston RS. Chromosome 15q25 (CHRNA3-CHRNA5) variation impacts indirectly on lung cancer risk. PLoS One 2011;6:e19085.
26. Lips EH, Gaborieau V, McKay JD, Chabrier A, Hung RJ, Boffetta P, et al. Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. Int J Epidemiol 2010;39:563–77.
27. Rouault TA. The role of iron regulatory proteins in mammalian iron homeostasis and disease. Nat Chem Biol 2006;2:406–14.
28. Ghio AJ, Hilborn ED, Stonehuerner JG, Dailey LA, Carter JD, Richards JH, et al. Particulate matter in cigarette smoke alters iron homeostasis to produce a biological effect. Am J Respir Crit Care Med 2008;178:1130–8.
29. Falvella FS, Galvan A, Frullanti E, Spinola M, Calabro E, Carbone A, et al. Transcription deregulation at the 15q25 locus in association with lung adenocarcinoma risk. Clin Cancer Res 2009;15:1837–42.
30. Amos CI, Gorlov IP, Dong Q, Wu X, Zhang H, Lu EY, et al. Nicotinic acetylcholine receptor region on chromosome 15q25 and lung cancer risk among African Americans: a case-control study. J Natl Cancer Inst 2010;102:1199–205.
31. Hansen HM, Xiao Y, Rice T, Bracci PM, Wrensch MR, Sison JD, et al. Fine mapping of chromosome 15q25.1 lung cancer susceptibility in African-Americans. Hum Mol Genet 2010;19:3652–61.
32. The International HapMap Project. Nature 2003;426:789–96.
33. Shanahan HP, Memon FN, Upton GJ, Harrison AP. Normalized Affymetrix expression data are biased by G-quadruplex formation. Nucleic Acids Res 2012;40:3307–15.

Published OnlineFirst April 26, 2012; DOI: 10.1158/1055-9965.EPI-11-1123-T

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

Association of the 15q25 and 5p15 Lung Cancer Susceptibility Regions with Gene Expression in Lung Tumor Tissue

Gord Fehringer, Geoffrey Liu, Melania Pintilie, et al.

Cancer Epidemiol Biomarkers Prev 2012;21:1097-1104. Published OnlineFirst April 26, 2012.