ORIGINAL ARTICLE

Lung Cancer Risk in Never-Smokers of European Descent is Associated With Genetic Variation in the $5_p15.33$ *TERT-CLPTM1Ll* Region

Rayjean J. Hung, PhD,[a] Margaret R. Spitz, MD, MPH,[b] Richard S. Houlston, MD, PhD,[c] Ann G. Schwartz, PhD, MPH,[d] John K. Field, PhD, FRCPath,[e] Jun Ying, MS,[f] Yafang Li, PhD,[b] Younghun Han, PhD,[b] Xuemei Ji, PhD,[g] Wei Chen, MS,[h] Xifeng Wu, PhD,[h] Ivan P. Gorlov, PhD,[g] Jie Na, MS,[i] Mariza de Andrade, PhD,[i] Geoffrey Liu, MD, PhD,[j] Yonathan Brhane, PhD,[a] Nancy Diao, ScD,[k] Angela Wenzlaff, PhD,[d] Michael P. A. Davies, PhD,[e] Triantafillos Liloglou, PhD,[e] Maria Timofeeva, PhD,[l,m] Thomas Muley, PhD,[n,o] Hedy Rennert, PhD,[p] Walid Saliba, PhD,[p] Bríd M. Ryan, PhD, MPH,[q] Elise Bowman, MS,[q] Juan-Miguel Barros-Dios, PhD,[r] Mónica Pérez-Ríos, PhD,[r] Hal Morgenstern, PhD,[s] Shanbeh Zienolddiny, PhD,[t] Vidar Skaug, PhD,[t] Donatella Ugolini, PhD,[u] Stefano Bonassi, PhD,[v,w] Erik H. F. M. van der Heijden, MD, PhD,[x] Adonina Tardon, MD, PhD,[y] Stig E. Bojesen, MD,[z] Maria Teresa Landi, MD, PhD,[aa] Mattias Johansson, PhD,[bb] Heike Bickeböller, PhD,[cc] Susanne Arnold, MD,[dd] Loic Le Marchand, MD, PhD,[ee] Olle Melander, MD, PhD,[ff] Angeline Andrew, PhD,[gg] Kjell Grankvist, MD,[hh] Neil Caporaso, PhD,[aa] M. Dawn Teare, PhD,[ii] Matthew B. Schabath, PhD,[jj] Melinda C. Aldrich, PhD, MPH,[kk] Lambertus A. Kiemeney, PhD,[x] H-Erich Wichmann, MD, PhD,[ll] Philip Lazarus, PhD,[mm] Jose Mayordomo, MD, PhD,[nn] Monica Neri, PhD,[w] Aage Haugen, PhD,[t] Zuo-Feng Zhang, MD, PhD,[oo] Alberto Ruano-Raviña, PhD,[r] Hermann Brenner, PhD,[l] Curtis C. Harris, MD,[q] Irene Orlow, PhD,[pp] Gadi Rennert, PhD,[p] Angela Risch, PhD,[l,qq,rr] Paul Brennan, PhD,[bb] David C. Christiani, MD,[k] Christopher I. Amos, PhD, MS,[b] Ping Yang, MD, PhD,[ss] Olga Y. Gorlova, PhD[g,*]

[a]*Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada*
[b]*Baylor College of Medicine, Houston, Texas*
[c]*Institute of Cancer Research, London, United Kingdom*
[d]*Wayne State University, Detroit, Michigan*
[e]*University of Liverpool, Liverpool, United Kingdom*
[f]*University of Texas McGovern Medical School, Houston, Texas*
[g]*Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire*
[h]*The University of Texas, MD Anderson Cancer Center, Houston, Texas*
[i]*Mayo Clinic, Rochester, Minnesota*

*Corresponding author.

Disclosure: Dr. Field has received grants from HTA funding for the UKLS trial and Liverpool CCG; and has received personal fees from Epigenomics, Vision Gate, Astra Zeneca, and Janssen. Dr. Liu has received personal fees from Pfizer, Novartis, Astra Zeneca, Roche, Bayer, Abbvie, Takeda, Merck, and Bristol-Myers Squibb. Dr. Muley has received grants and personal fees from Roche Diagnostics. Dr. van der Heijden has received grants from AstraZeneca Oncology, Pentax Medical, and Philips Medical Systems; personal fees from Pentax Medical and Medtronic; and has received nonfinancial support from Pentax Medical, Philips Medical Systems, and Medtronic. Dr. Aldrich has received grants from the National Institutes of Health/National Cancer Institute. Dr. Risch has received grants from the National Institutes of Health/National Cancer Institute and Deutsche Krebshilfe. Dr. Gorlova has received grants from the National Institutes of Health/National Cancer Institute. The remaining authors declare no conflict of interest.

Address for correspondence: Olga Y. Gorlova, PhD, Geisel School of Medicine at Dartmouth, 1 Medical Center Dr., Lebanon, New Hampshire 03756. E-mail: olga.y.gorlova@dartmouth.edu

ISSN: 1556-0864

https://doi.org/10.1016/j.jtho.2019.04.008

[j]Princess Margaret Cancer Center, Toronto, Ontario, Canada
[k]Harvard T.H. Chan School of Public Health, Boston, Massachusetts
[l]German Cancer Research Center (DKFZ), Heidelberg, Germany
[m]University of Edinburgh, Edinburgh, United Kingdom
[n]German Center for Lung Research, Heidelberg, Germany
[o]University Hospital Heidelberg, Heidelberg, Germany
[p]Technion-Israel Institute of Technology, Haifa, Israel
[q]Center for Cancer Research, National Cancer Institute, Bethesda, Maryland
[r]University of Santiago de Compostela, Praza do Obradoiro, Coruña, Spain
[s]Medical School, University of Michigan, Ann Arbor, Michigan
[t]National Institute of Occupational Health (STAMI), Oslo, Norway
[u]University of Genoa, Genoa, Italy
[v]San Raffaele University, Rome, Italy
[w]San Raffaele Pisana - Scientific Hospitalization and Care Insitution, Rome, Italy
[x]Radboud University Medical Center, Nijmegen, Netherlands
[y]University of Oviedo and CIBERESP, Oviedo, Spain
[z]Copenhagen University Hospital, Copenhagen, Denmark
[aa]National Cancer Institute, Bethesda, Maryland
[bb]International Agency for Research on Cancer, Lyon, France
[cc]University Medical Center Goettingen, Goettingen, Germany
[dd]University of Kentucky, Lexington, Kentucky
[ee]University of Hawaii Cancer Center, Honolulu, Hawaii
[ff]Lund University, Lund, Sweden
[gg]Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire
[hh]Umeå University, Umeå, Sweden
[ii]University of Sheffield, Sheffield, United Kingdom
[jj]H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida
[kk]Vanderbilt University Medical Center, Nashville, Tennessee
[ll]Helmholtz Zentrum Munchen, German Research Center for Environmental Health (GmbH), Bavaria, Germany
[mm]Washington State University, Spokane, Washington
[nn]University of Colorado, Aurora, Colorado
[oo]University of California – Los Angeles, Los Angeles, California
[pp]Memorial Sloan Kettering Cancer Center, New York, New York
[qq]University of Salzburg, Salzburg, Austria
[rr]Cancer Cluster Salzburg, Salzburg, Austria
[ss]Mayo Clinic, Scottsdale, Arizona

Received 6 February 2019; revised 30 March 2019; accepted 11 April 2019
Available online - 19 April 2019

ABSTRACT

Introduction: Inherited susceptibility to lung cancer risk in never-smokers is poorly understood. The major reason for this gap in knowledge is that this disease is relatively uncommon (except in Asians), making it difficult to assemble an adequate study sample. In this study we conducted a genome-wide association study on the largest, to date, set of European-descent never-smokers with lung cancer.

Methods: We conducted a two-phase (discovery and replication) genome-wide association study in never-smokers of European descent. We further augmented the sample by performing a meta-analysis with never-smokers from the recent OncoArray study, which resulted in a total of 3636 cases and 6295 controls. We also compare our findings with those in smokers with lung cancer.

Results: We detected three genome-wide statistically significant single nucleotide polymorphisms rs31490 (odds ratio [OR]: 0.769, 95% confidence interval [CI]: 0.722–0.820; p value 5.31×10^{-16}), rs380286 (OR: 0.770, 95% CI: 0.723–0.820; p value 4.32×10^{-16}), and rs4975616 (OR: 0.778, 95% CI: 0.730–0.829; p value 1.04×10^{-14}). All three mapped to Chromosome 5 *CLPTM1L-TERT* region, previously shown to be associated with lung cancer risk in smokers and in never-smoker Asian women, and risk of other cancers including breast, ovarian, colorectal, and prostate.

Conclusions: We found that genetic susceptibility to lung cancer in never-smokers is associated to genetic variants with pan-cancer risk effects. The comparison with smokers shows that top variants previously shown to be associated with lung cancer risk only confer risk in the presence of tobacco exposure, underscoring the importance of gene-environment interactions in the etiology of this disease.

Keywords: Lung cancer; Never smokers; Genome-wide association study; Genetic susceptibility

Introduction

Lung cancer is the leading cause of cancer mortality worldwide, accounting for more than 1 million deaths each year.[1] Although most lung cancer is preventable, because the majority of cases occur in tobacco smokers, approximately 10% of cases are seen in lifetime never-smokers.[2] Although lung cancer is diagnosed in a minority of never-smokers, it still ranks as the seventh to ninth most common cause of cancer death worldwide.[2]

In never-smokers, lung cancer has characteristics distinct from those associated with smoking, including different histology and mutation spectrum.[3] The only well-established risk factors for lung cancer in never-smokers are exposure to radon,[4] secondhand smoke, dust, asbestos, and, notably, family history of cancer, which has provided evidence for inherited susceptibility.[5-7]

To date, genome-wide association studies (GWAS) on lung cancer have largely been focused on ever-smokers, and have identified 18 independent loci influencing risk.[7,8] Whereas several GWAS studies in never-smokers have been conducted, these have primarily been based on Asian women.[9-12] Several environmental risk factors for lung cancer, including cooking fumes and air pollution, are highly prevalent in Asian populations, raising the possibility of effect modification.[13] Identifying lung cancer susceptibility alleles among never-smoking European populations has been limited to candidate gene analyses and small GWAS.[14-18] Reported here are the results of a large GWAS of lung cancer in never-smokers of European descent, based on 3636 cases and 6295 controls.

Materials and Methods

Study Design and Samples

Never-smokers were defined as individuals who had smoked less than 100 cigarettes during their lifetime. The study had a discovery and a replication series, both from studies participating in the International Lung Cancer Consortium (ILCCO; http://ilcco.iarc.fr). The discovery series, after quality control (See Supplementary material), comprised 1287 cases and 1655 controls with European ancestry from seven centers (Supplementary Table 1). The replication series comprised 960 cases and 940 controls from 16 study centers, of which some centers (but not study subjects) also participated in the discovery phase (Supplementary Table 2). Comprehensive details of each series have been previously reported.[17,19-23] To increase statistical power, data on never-smokers recently generated by the OncoArray lung cancer study from ILCCO were also leveraged.[20] After excluding samples overlapping between the OncoArray and the discovery set and between the OncoArray and the replication set, 1149 cases and 1144 controls from the discovery, 1527 cases and 4211 controls from the OncoArray, and 960 cases and 940 controls from the replication sets were included in the final analyses. Most of the lung cancer cases (76.7% in the discovery, 69.2% in the replication, and 63.1% in the OncoArray sets) had histologically confirmed adenocarcinoma, followed by squamous and small cell carcinoma (Supplementary Tables 1–3). Given that subtype-specific associations are likely to exist, adenocarcinomas were also analyzed separately. Table 1 presents the demographic characteristics of the final dataset.

Table 1. Characteristics of Never-Smoking Lung Cancer Cases and Controls Included in the Final Dataset

Characteristic	Cases (n = 3636)	Controls (n = 6296)
Age, y, mean (SD)	63.6 (12.4)	61.9 (11.9)
Sex, n (%)		
Male	1156 (31.8)	2595 (41.2)
Female	2480 (68.2)	3701 (58.8)
Histology, n (%)		
Adenocarcinoma	2509 (69.0)	6296
Squamous cell carcinoma	310 (8.5)	6296

Genotyping and Quality Control

Both cases and controls from the discovery set were genotyped using Illumina Infinium OmniExpress-24 v1.2 BeadChips, with the exception of cases and controls from Harvard School of Public Health (HSPH), genotyped on Illumina Human660W-Quad BeadChip. Genotyping of the replication series for 384 selected single nucleotide polymorphisms (SNPs) was performed using Illumina GoldenGate technology. Genotyping quality control and SNP selection procedures are detailed in the Supplementary material. The OncoArray genotyping platform, the never-smoker samples to which it was applied, and genotyping and quality control procedures are described in the Supplementary material and have been previously characterized in detail.[20,24]

Data Analysis

To harmonize data and address population stratification in the discovery set, the studies were grouped according to the genotyping array they used and the geographic origin of the subjects they enrolled. This resulted in two groups: United Kingdom studies and North American studies. Further, since the HSPH samples were genotyped on a different platform, these were analyzed separately. Thus, the following clusters were used: (1) HSPH, (2) United Kingdom, and (3) North America (see

Supplementary Table 4 for more detail). Three separate GWAS analyses were ran based on the three groups. We applied logistic regression analyses with case-control status as the outcome and the SNP genotype as a predictor to identify risk-associated SNPs in these three groups. Additive models, with 0 for the reference allele homozygotes, 1 for heterozygotes, and 2 for variant allele homozygotes, were used. Reference alleles were defined as in the hg19 reference genome. Age (continuous variable), sex, secondhand smoke exposure (SHS; from any venue at any period in a lifetime), education level, and study site within the group (if more than one site) were used as covariates. The definition of the education variables and more information on the SHS assessment are given in the Supplementary material. Missing values for SHS and education status were treated as a separate category. To offset potential effects of population stratification within clusters, SNP-based principal components analyses (PCAs) were performed and the corresponding first five principal components were included as covariates, even though the PCA of these three GWAS clusters do not suggest population stratification (Supplementary Fig. 1).[25] An inverse variance fixed-effects meta-analysis was used to combine the results for the three group-based GWAS.[26]

A brief description of the OncoArray never-smoker dataset is provided in the Supplementary material. To perform the joint analysis of the discovery and the OncoArray sets, inverse variance meta-analysis was used, whereby studies were grouped into five clusters (Discovery-North America, Discovery-United Kingdom, OncoArray-North America, OncoArray-United Kingdom, and OncoArray-Continental Europe), as detailed in Supplementary Table 5. This joint analysis was adjusted for age, sex, study site within the group, and the first five principal components, but not SHS or education level, as they were not available in the OncoArray set.

Criteria for SNP selection and the quality control procedures in the replication phase are described in the Supplementary material.

Results

We focus on the joint analysis of the discovery and OncoArray sets as having the largest sample size (the results for the discovery set separately are presented in Supplementary Figure 2 showing the Q-Q plot that shows no indication of an inflation of type I error ($\lambda = 1.005$), and Supplementary Table 6 presenting the list of the top SNPs derived from the discovery set ($p < 1 \times 10^{-4}$)).

Figure 1 presents the scatter plot of the $-\log_{10}$P values against the chromosome position (the so-called Manhattan plot) for the meta-analysis of the discovery and the OncoArray samples. The analysis identified 71 genome-wide statistically significant SNPs ($p < 5 \times 10^{-8}$, the accepted genome-wide level of statistical significance), all of them mapping to the 5p15.33 *CLPTM1* like (*CLPTM1L*)-telomerase reverse transcriptase (*TERT*) region.[27] Supplementary Table 7 presents the 229 top SNPs at p value less than 10^{-5}. There is also a peak on chromosome 9 in the cyclin dependent kinase inhibitor 2A (*CDKN2A*) region, but none of the SNPs in this regions attained statistical significance at the GWAS level.

The PCA of the replication samples showed no differences by the case-control status for the first five principal components (Supplementary Fig. 3).

Supplementary Table 8 presents the list of nominally statistically significant ($p < 0.05$) SNPs from the replication analysis. The most significant SNPs, rs380286 ($p = 3.88 \times 10^{-7}$), rs31490 ($p = 4.68 \times 10^{-7}$), and rs4975616 ($p = 2.50 \times 10^{-6}$) were located in the 5p15.33 (*CLPTM1L-TERT*) region (Table 2). These three

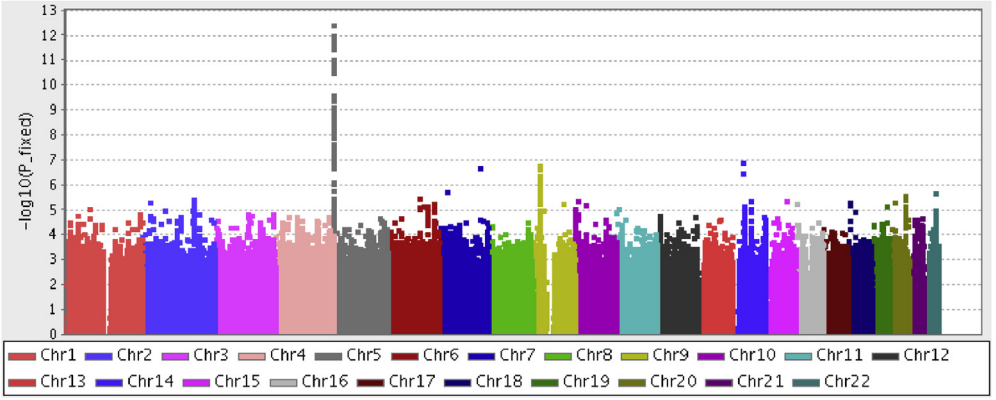

Figure 1. Manhattan plot of the association analysis of lung cancer in European ancestry never-smokers performed jointly in the discovery set and the OncoArray samples. The *x* axis is chromosomal position, and the *y* axis is the statistical significance on a $-\log_{10}$ scale.

Table 2. The Three GWAS-Significant ($p < 5 \times 10^{-8}$) Variants for Lung Cancer in European Ancestry Never-Smokers, Found in the Joint Analysis of the Original Discovery Set, the Never-Smoker Subset of the OncoArray Set, and the Replication Set (6 Clusters, 3636 Cases, 6295 Controls), Adjusted for Age, Sex, and the First Five Principal Components

SNP ID	CHR	Position	Odds Ratio[a]	95% CI	*p* Value[a]	Reference Allele	Effect Allele	EAF	Gene Symbol
rs380286[b]	5	1320247	0.770	0.723-0.820	4.32×10^{-16}	A	G	0.4169	*CLPTM1L*
rs31490[c]	5	1344458	0.769	0.722-0.820	5.31×10^{-16}	G	A	0.4142	*CLPTM1L*
rs4975616[d]	5	1315660	0.778	0.730-0.829	1.04×10^{-14}	G	A	0.4005	*CLPTM1L*

[a]Adjusted for age, sex, and the first five principal components.
[b]Intronic variant.
[c]Splice variant.
[d]Downstream gene variant.
GWAS, genome-wide association study; SNP, single nucleotide polymorphism; ID, identification; CHR, chromosome; CI, confidence interval; EAF, effect allele frequency

SNPs were significant after the Bonferroni correction for 370 tests resulting in the *p* value of 1.35×10^{-4} to declare significance (the false discovery rate [FDR] approach identified the same three SNPs as statistically significant) (Supplementary Table 8).

The 370 candidate SNPs selected for replication (see Supplementary material for the selection criteria) were analyzed using all three study population sets: the discovery, the replication, and the OncoArray (total 3636 cases and 6295 controls). The analysis identified three SNPs statistically significant at the genome-wide level: rs380286 ($p = 1.6 \times 10^{-14}$), rs31490 ($p = 5.1 \times 10^{-14}$), and rs4975616 ($p = 5.8 \times 10^{-14}$) (Table 2). These three SNPs are from the *CLPTM1L-TERT* region and the association with the variant alleles was consistently negative (odds ratio < 1). These SNPs belong to a wide linkage disequilibrium (LD) block corresponding to the LD region 2 marked by rs451360 as described in Wang et al.[28] The very high LD between the pairs of SNPs (0.925 for rs380286 and rs31490; 0.915 for rs380286 and rs4975616; and 0.955 for rs31490 and rs4975616) did not allow identifying the leading SNP among the three as there was very little variation within an SNP when the genotypes of the other two were fixed. The results of the joint analysis of the discovery and replication sets without the OncoArray samples are shown in Supplementary Table 9. In brief, the same three SNPs from the *CLPTM1L-TERT* region were identified to be genome-wide statistically significant.

Analysis of only adenocarcinoma cases produced nearly identical results, with only the *CLPTM1L-TERT* region SNPs showing statistical significance (Supplementary Tables 10 and 11).

Table 3 summarizes the comparisons between our study results and previous published findings reported in never-smokers from genome-wide and candidate gene/SNP association studies in both individuals of European descent and Asians. Our study confirmed SNPs located in the 5p15.33 (*CLPTM1L-TERT*) region. The direction of the association is highly concordant among the studies for the SNPs in this region. The results for 3q28 (tumor protein p63 [*TP63*]) and 6q22.2 (*ROS1*-discoidin, CUB and LCCL domain containing 1 [*DCBLD1*]) regions are suggestive in our analysis (*p* values of $\sim 10^{-4}$ for both these regions). The results from our study for the loci identified in the recently published largest-to-date lung cancer study that involved mostly smokers are shown in Supplementary Table 12.[20]

A comparison of the regional association plots for the *CLPTM1L-TERT* region and 15q25 (cholinergic receptor nicotinic alpha 3 subunit [*CHRNA3*]) region in never-smokers and smokers was also performed (whereby the smokers' data were obtained from the lung OncoArray project) (Figs. 2*A* and *B*). We found that the risk association profile plotted as the $-\log_{10}$P for the SNPs in the *CLPTM1L-TERT* region in never-smokers tightly followed that in smokers (Fig. 2*A*). By contrast, the association profiles in the *CHRNA3* region (implicated in nicotine dependence) are strikingly different in never- and ever-smokers, with very high $-\log_{10}$P values in smokers and a flat profile in never-smokers (Fig. 2*B*). Analogous comparisons for two other regions, *TP63* and *CDKN2A*, are presented in Supplementary Figure 4.

The analyses of associations for the three most statistically significant SNPs from the *CLPTM1L-TERT* region stratified by SHS exposure status are shown in Supplementary Table 13. There was no indication of SNP-SHS interaction effects or a SNP effect modification by the SHS exposure as the interaction term was not significant for any of the SNPs.

Discussion

This is the largest lung cancer GWAS so far conducted in never-smokers of European descent. However, only one region (*CLPTM1L-TERT*) strongly associated with lung cancer risk in this patient population was found. Our results for this region corroborate findings by

Table 3. Previous Findings From the Association Analyses of Lung Cancer in Never-Smokers, With a Comparison to This Study

Previously Published Studies											This Study[a]	
Region	Gene	RefSeq	Study Type	Pubmed ID	Histology	Ethnicity	Discovery Cases | Controls	Replication Cases | Controls	OR	*p* Value	OR	*p* Value
13q31.3	*GPC5*	rs2352028	GWAS	Li et al.[17]	NSCLC	Mostly Eur. descent	377 | 377	328 | 407	1.46	5.90E-06	0.99	0.95
5p15.33	*CLPTM1L*	rs4975616	Candidate	Wang et al.[15]	NSCLC	Eur. descent	239 | 553	–	0.69	7.90E-04	0.78	1.04E-14[b]
5p15.33	*CLPTM1L-TERT*	rs2736100	GWAS	Hsiung et al.[9]	Adeno	Asian women	584 | 585	2184 |2515	1.5	5.40E-11	1.3	2.66E-09[b]
10q25.2	*VTI1A*	rs7086803	GWAS	Lan et al.[10]	NSCLC	Asian women	5547 | 4492	1085 | 2877	1.3	5.10E-17	1.3	0.011[b]
6q22.2	*ROS1-DCBLD1*	rs9387478							0.85	7.80E-08	0.86	1.50E-04[b]
6p21.32	*HLA II*	rs2395185							1.16	2.60E-06	1.04	0.34
5p15.33	*CLPTM1L-TERT*	rs2736100							1.38	4.20E-27	1.27	2.66E-09[b]
5p15.33	*CLPTM1L-TERT*	rs2853677	GWAS	Shiraishi et al.[12]	Adeno	Asians (Japanese)	1695 | 5333	3328 | 8168	1.44	3.90E-23	1.28	1.12E-09[b]
5p15.33	*CLPTM1L-TERT*	rs2736100							1.37	9.90E-19	1.27	2.66E-09[b]
3q28	*TP63*	rs10937405							1.28	2.00E-10	1.16	1.50E-04[b]
17q24.3	*BPTF*	rs7216064							1.21	1.50E-06	1.1	0.054
6p21.3	*BTNL2*	rs3817963							1.21	1.50E-07	1.06	0.2
1q25.1	*ACVR1B*	rs10127728	Candidate	Spitz et al.[14]	NSCLC	Mostly Eur. descent	451 | 508	–	1.68	3.00E-04	1.06	0.34
3q28	*TP63*	rs4488809	Replication of GWAS findings	Seow et al.[11]	Adeno	Asian women		7448 | 7007	0.8	4.30E-17	0.82	8.52E-07[b]
5p15.33	*TERT*	rs2736100						7505 | 7070	1.43	6.12E-43	0.79	2.66E-09[b]
6p21.1	*FOXP4*	rs7741164						10531 | 10648	1.17	3.96E-13	0.97	8.28E-01
6p21.3	*BTNL2*	rs3817963						7255 | 6745	1.16	1.63E-07	1.06	1.97E-01
6p21.32	*HLA-DPB1*	rs2179920						7457 | 7020	1.17	1.69E-05	1.08	9.42E-02
6p21.32	*HLA class II*	rs2395185						7757 | 9637	1.16	2.04E-09	1.04	3.91E-01
6q22.2	*ROS1/DCBLD1*	rs9387478						8022 | 9970	0.86	5.25E-11	0.86	1.53E-04[b]
9p21.3		rs72658409						10780 | 10938	0.76	2.37E-10	0.89	1.43E-01
10q25.2	*VTI1A*	rs7086803						7964 | 9914	1.25	9.22E-17	1.31	1.12E-02[b]
12q13.13		rs11610143						10267 | 10634	0.85	3.55E-13	0.97	4.88E-01
17q24.3	*BPTF*	rs7216064						7720 | 8630	0.86	6.19E-09	1.10	5.43E-02

[a]"This study" pertains to the results of the meta-analysis of the discovery and OncoArray sets, except for rs4975616, for which the result from the meta-analysis of the discovery, OncoArray, and replication sets is shown.
[b]Nominally significant *p* values.
RefSeq, Reference sequence or single nucleotide polymorphism identification; GWAS, genome-wide association study; OR, odds ratio; Eur., European; Adeno, adenocarcinoma.

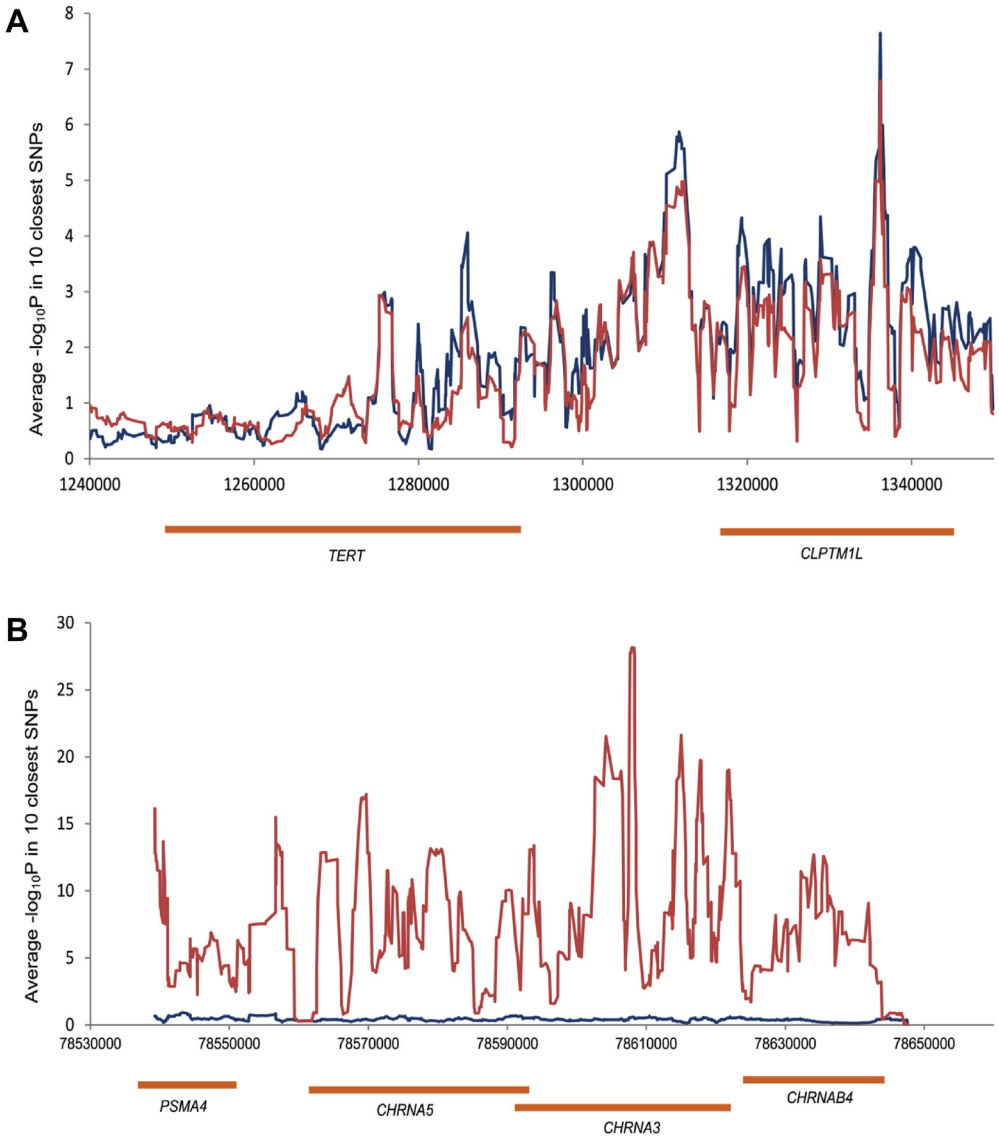

Figure 2. Regional association plots for smokers (*red line*) and never smokers (*blue line*) in *CLPTM1L-TERT* region (*A*) and *CHRNA3-5* region (*B*). The *y* axis corresponds to -$\log_{10}P$ for 650 SNPs in the *CLPTM1L-TERT* region and -$\log_{10}P$ for 535 SNPs in *CHRNA3-5* region. To aid visual representation we selected the 10 closest single nucleotide polymorphism (SNP) and computed average -$\log_{10}$P- values.

earlier studies of lung cancer in never-smokers (Table 3), showing consistent direction of effect. The 5p15.33 *CLPTM1L-TERT* region SNPs have also been reported to be associated with multiple cancers including lung cancer in smokers, breast cancer, glioma, nasopharyngeal cancer, and prostate cancer.[16,29-32] *TERT* encodes the catalytic subunit of the telomerase reverse transcriptase, which takes part in adding nucleotide repeats to chromosome ends.[33] Although active in early development and germ cells, this gene is not expressed in most adult tissues, resulting in a shortening of telomeres with each cell division. When telomeres become critically short, the cell can no longer divide. However, cancer cells can upregulate telomerase, which enables them to continue dividing.[34] The *CLPTM1L* gene is reported to be overexpressed in lung and pancreatic cancer where it promotes growth and survival.[35,36] Also, there is a locus within the *CLPTM1L* gene that serves as a binding site for ZNF148, which promotes expression of *TERT*.[37]

Functional annotation of the top-identified SNPs using Encyclopedia of DNA Elements (ENCODE) found that rs4975616 coincides with the binding site for three transcription factors: ELF1, ZEB1, and BCLAF1.[38] Both *TERT* and *CLPTM1L* are among the many target genes for ELF1 and ZEB1; *CLPTM1L* (but not TERT) is among the target genes for BCLAF1. According to Ensemble regulatory database, SNP rs31490 is located in the region

that acts as a promotor for *CLPTM1L* in the developing lung.[39] In the Genotype-Tissue Expression all three SNPs: rs31490, rs380286, and rs4975616 are reported as expression quantitative trait loci (eQTL) for *TERT* in esophagus and *CLPTM1L* in skin tissue.[40]

Previously, a fine-mapping study has been conducted on this locus to deeply investigate its association with lung cancer risk.[41] The study included a limited number of never-smokers and the novel loci identified did not show significant effect, specifically in never-smokers. However, the direction of the effect was largely consistent with that in smokers, in line with what our study reports (Fig. 2*A*).

For other SNPs, for example, those reported by Li et al.,[17] no association in our study was detected. However, the study by Li et al.[17] used additional covariates (e.g., chronic obstructive pulmonary disease and lung cancer family history) to adjust for in their analyses. This may have made a comparison of their results with our study less straightforward because the data on these covariates were not available from the majority of the sites participating in our study. The SNPs rs10937405 for 3q28 and rs9387478 for 6q22.2, previously reported to be significant in Asian never-smoking women (Table 3), showed at best a suggestive association (p values of $\sim 10^{-4}$ in both cases). These two regions have been shown also to be implicated in other cancer sites. SNPs in the *TP63* region have been shown to be associated with lung adenocarcinoma in the U.K. population, acute lymphoblastic leukemia, bladder cancer, and pancreatic cancer.[8,42-44] SNPs in the *ROS1-DCBLD1* region have been shown to be associated with colorectal cancer.[45] This further suggests that SNPs/regions associated with lung cancer risk in never-smokers are not specific for this type of cancer but rather have pleiotropic effects.

Our analysis was designed to control for demographic variables (age and sex, as controls were slightly but statistically significantly younger [$p < 0.001$] and had a higher proportion of men than cases [$p < 0.001$]) as well as for known and potential risk factors, specifically, where possible, for education status and self-reported SHS exposure.[46] To account for possible population stratification, the first five principal components and the study site were also adjusted. However, the information on radon exposure, asbestos, prior respiratory conditions, and diet was not available from most studies. As such, these established and putative risk factors were not accounted for in the analyses. A further limitation is the self-reported nature of the never-smoker status. Differential misreporting of the smoking status, for example, if a modest proportion of former or current smoker controls reported that they have never smoked, might lead to SNPs associated with smoking appearing as protective. Unfortunately, the great majority of the participating studies did not verify it by cotinine measurements. However, SNPs in *CHRNA3*-5 or cytochrome P450 family 2 subfamily A member 6 (*CYP2A6*) regions, known to be associated with smoking, did not show any effect in this study (Fig. 2*B*; Supplementary Table 11).[20]

Latest GWAS of lung cancer in smokers have generated many more findings than did this study, which is not surprising given that the former are much larger. Most SNPs reported as statistically significant in smokers showed the same direction of effect in never-smokers (Supplementary Table 12). Gene-smoking interaction may be another factor contributing to the higher number of positive findings among smokers than never-smokers: some of the sequence variations that are neutral in the absence of tobacco smoking confer risk when smoking and the associated tissue and DNA damage are present.

High body mass index and alcohol exposure are common and may also explain a proportion of the lung cancer risk in never-smokers.[47,48] It is possible that there are rare variants influencing risk that could not be detected by a GWAS that focuses on common variants. Additionally, gene-gene interactions that are beyond the scope of this study may in part explain variability in the incidence of lung cancer in never-smokers. Very rarely, individuals can carry inherited mutations in *TP53* increasing lung cancer risk.[49] The availability of results from our GWAS will allow additional exposures to be studied using Mendelian Randomization approaches (as exemplified in Wang et al.[50]). Developing models that can identify never-smokers at highest risk for lung cancer development could improve early detection.

Acknowledgments

This work was supported in part by National Institutes of Health (NIH) grants CA149462, CA209414, CA092824, ES00002, U01CA209414, U19CA203654, 1K07CA172294, P50CA119997, R01CA060691, R01CA87895, P30CA22453, P30CA008748, P30CA076292, and U01CA164973; Department of Health and Human Services grant HHSN261201300011; James & Esther King Biomedical Research Program Grant 09KN-15; Helmholtz-DAAD fellowship A/07/97379; the Society of Memorial Sloan Kettering Cancer Center through their annual appeal and Steps for Breath; Italian Ministry of Health grant for Institutional Research 2017-2018 and Associazione Italiana per la Ricerca sul Cancro grant IG2015/17564IO; and Instituto de Salud Carlos III. PI15/01211 grant and Xunta de Galicia grant 10CSA208057PR. The Toronto study was supported by The Canadian Cancer Society Research Institute (020214), and the Alan Brown Chair and Lusi Wong

Programs at the Princess Margaret Hospital Foundation. The LUCY study was funded in part by the Germany National Genome Research Network (NGFN), the DFG (BI576/2-1; BI 576/2–2, Bi 576/4-1; Bi 576/4-2; Wi 621/10-1; Wi 621/10-2), the Helmholtzgemeinschaft (HGF) and the Federal office for Radiation Protection (BfS:STSch4454). KORA Surveys were funded by the Helmholtz-Zentrum München (HMGU), which is funded by the German Federal Ministry of Education, Science, Research and Technology and the State of Bavaria. The Liverpool Lung Project is funded by the Roy Castle Lung Cancer Foundation. The Resource for the Study of Lung Cancer Epidemiology in North Trent (ReSoLuCENT) study was funded by the Sheffield Hospitals Charity, Sheffield Experimental Cancer Medicine Centre and Weston Park Hospital Cancer Charity. ILCCO data harmonization was supported by Canada Research Chair to Dr. Hung. Partial support for this research was provided by Cancer Prevention Research Institute of Texas grant RR170048 which supports Dr. Christopher Amos, a CPRIT Scholar in Cancer Research.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *Journal of Thoracic Oncology* at www.jto.org and at https://doi.org/10.1016/j.jtho.2019.04.008.

References

1. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019;144:1941–1953.
2. Thun MJ, Hannan LM, Adams-Campbell LL, et al. Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. *PLoS Med*. 2008;5:e185.
3. Subramanian J, Govindan R. Lung cancer in 'never-smokers': a unique entity. *Oncology (Williston Park)*. 2010;24:29–35.
4. Torres-Duran M, Ruano-Ravina A, Parente-Lamelas I, et al. Residential radon and lung cancer characteristics in never smokers. *Int J Radiat Biol*. 2015;91:605–610.
5. Gorlova OY, Zhang Y, Schabath MB, et al. Never smokers and lung cancer risk: a case-control study of epidemiological factors. *Int J Cancer*. 2006;118:1798–1804.
6. Markowitz SB, Levin SM, Miller A, et al. Asbestos, asbestosis, smoking, and lung cancer. New findings from the North American insulator cohort. *Am J Respir Crit Care Med*. 2013;188:90–96.
7. Gorlova OY, Weng SF, Zhang Y, et al. Aggregation of cancer among relatives of never-smoking lung cancer patients. *Int J Cancer*. 2007;121:111–118.
8. Wang Y, Broderick P, Matakidou A, et al. Variation in TP63 is associated with lung adenocarcinoma in the UK population. *Cancer Epidemiol Biomarkers Prev*. 2011;20:1453–1462.
9. Hsiung CA, Lan Q, Hong YC, et al. The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS Genet*. 2010;6. pii: e1001051.
10. Lan Q, Hsiung CA, Matsuo K, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat Genet*. 2012;44:1330–1335.
11. Seow WJ, Matsuo K, Hsiung CA, et al. Association between GWAS-identified lung adenocarcinoma susceptibility loci and EGFR mutations in never-smoking Asian women, and comparison with findings from Western populations. *Hum Mol Genet*. 2017;26:454–465.
12. 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;44:900–903.
13. Liao Y, Xu L, Lin X, et al. Temporal trend in lung cancer burden attributed to ambient fine particulate matter in Guangzhou, China. *Biomed Environ Sci*. 2017;30:708–717.
14. Spitz MR, Gorlov IP, Amos CI, et al. Variants in inflammation genes are implicated in risk of lung cancer in never smokers exposed to second-hand smoke. *Cancer Discov*. 2011;1:420–429.
15. Wang Y, Broderick P, Matakidou A, et al. Role of 5p15.33 (TERT-CLPTM1L), 6p21.33 and 15q25.1 (CHRNA5-CHRNA3) variation and lung cancer risk in never-smokers. *Carcinogenesis*. 2010;31:234–238.
16. 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.
17. Li Y, Sheu CC, Ye Y, et al. Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncol*. 2010;11:321–330.
18. Landi MT, Chatterjee N, Caporaso NE, et al. GPC5 rs2352028 variant and risk of lung cancer in never smokers. *Lancet Oncol*. 2010;11:714–716. author reply 716.
19. Eisen T, Matakidou A, Houlston R, et al. Identification of low penetrance alleles for lung cancer: the GEnetic Lung CAncer Predisposition Study (GELCAPS). *BMC Cancer*. 2008;8:244.
20. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*. 2017;49:1126–1132.
21. Schwartz AG, Yang P, Swanson GM. Familial risk of lung cancer among nonsmokers and their relatives. *Am J Epidemiol*. 1996;144:554–562.
22. Wenzlaff AS, Cote ML, Bock CH, et al. GSTM1, GSTT1 and GSTP1 polymorphisms, environmental tobacco smoke exposure and risk of lung cancer among never smokers: a population-based study. *Carcinogenesis*. 2005;26:395–401.
23. Ugolini D, Neri M, Canessa PA, et al. The CREST biorepository: a tool for molecular epidemiology and translational studies on malignant mesothelioma, lung cancer, and other respiratory tract diseases. *Cancer Epidemiol Biomarkers Prev*. 2008;17:3013–3019.

24. Amos CI, Dennis J, Wang Z, et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomarkers Prev*. 2017;26:126–135.
25. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–909.
26. Viechtbauer. Conducting meta-analyses in R with the metafor Package. *J Stat Software*. 2010;36:1–48.
27. Amos CI. Successful design and conduct of genome-wide association studies. *Hum Mol Genet*. 2007;16(Spec No. 2):R220–R225.
28. Wang Z, Zhu B, Zhang M, et al. Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum Mol Genet*. 2014;23:6616–6633.
29. Haiman CA, Chen GK, Vachon CM, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet*. 2011;43:1210–1214.
30. Rajaraman P, Melin BS, Wang Z, et al. Genome-wide association study of glioma and meta-analysis. *Hum Genet*. 2012;131:1877–1888.
31. Bei JX, Su WH, Ng CC, et al. A GWAS meta-analysis and replication study identifies a novel locus within CLPTM1L/TERT associated with nasopharyngeal carcinoma in individuals of Chinese ancestry. *Cancer Epidemiol Biomarkers Prev*. 2016;25:188–192.
32. Kote-Jarai Z, Olama AA, Giles GG, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet*. 2011;43:785–791.
33. Cheung AL, Deng W. Telomere dysfunction, genome instability and cancer. *Front Biosci*. 2008;13:2075–2090.
34. Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. *Eur J Cancer*. 1997;33:787–791.
35. James MA, Wen W, Wang Y, et al. Functional characterization of CLPTM1L as a lung cancer risk candidate gene in the 5p15.33 locus. *PLoS One*. 2012;7:e36116.
36. Jia J, Bosley AD, Thompson A, et al. CLPTM1L promotes growth and enhances aneuploidy in pancreatic cancer cells. *Cancer Res*. 2014;74:2785–2795.
37. Fang J, Jia J, Makowski M, et al. Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. *Nat Commun*. 2017;8:15034.
38. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46:D794–D801.
39. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;46:D754–D761.
40. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580–585.
41. Kachuri L, Amos CI, McKay JD, et al. Fine mapping of chromosome 5p15.33 based on a targeted deep sequencing and high density genotyping identifies novel lung cancer susceptibility loci. *Carcinogenesis*. 2016;37:96–105.
42. Ellinghaus E, Stanulla M, Richter G, et al. Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia*. 2012;26:902–909.
43. Figueroa JD, Ye Y, Siddiq A, et al. Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum Mol Genet*. 2014;23:1387–1398.
44. Childs EJ, Mocci E, Campa D, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat Genet*. 2015;47:911–916.
45. Peters U, Jiao S, Schumacher FR, et al. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology*. 2013;144:799–807 e724.
46. Couraud S, Zalcman G, Milleron B, et al. Lung cancer in never smokers—a review. *Eur J Cancer*. 2012;48:1299–1311.
47. Gao C, Patel CJ, Michailidou K, et al. Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. *Int J Epidemiol*. 2016;45:896–908.
48. Fehringer G, Brenner DR, Zhang ZF, et al. Alcohol and lung cancer risk among never smokers: a pooled analysis from the international lung cancer consortium and the SYNERGY study. *Int J Cancer*. 2017;140:1976–1984.
49. Hwang SJ, Cheng LS, Lozano G, et al. Lung cancer risk in germline p53 mutation carriers: association between an inherited cancer predisposition, cigarette smoking, and cancer risk. *Hum Genet*. 2003;113:238–243.
50. Wang C, Qin N, Zhu M, et al. Metabolome-wide association study identified the association between a circulating polyunsaturated fatty acids variant rs174548 and lung cancer. *Carcinogenesis*. 2017;38:1147–1154.