



# Disentangling tobacco-related lung cancer—genome-wide interaction study of smoking behavior and non-small cell lung cancer risk

María Lorenzo-González<sup>1,2</sup>, Alberto Fernández-Villar<sup>3,4</sup>, Alberto Ruano-Ravina<sup>2,5,6</sup>

<sup>1</sup>Service of Preventive Medicine, University Hospital Complex of Ourense, Ourense, Spain; <sup>2</sup>Department of Preventive Medicine and Public Health, University of Santiago de Compostela, A Coruña, Spain; <sup>3</sup>Service of Neumology, University Hospital Complex of Vigo, Vigo, Spain; <sup>4</sup>Instituto de Investigación Sanitaria Galicia Sur (IISGS), Vigo, Spain; <sup>5</sup>CIBER de Epidemiología y Salud Pública CIBERESP, CIBERESP, Spain; <sup>6</sup>Department of Epidemiology, Brown School of Public Health, Brown University, Providence, Rhode Island, USA

*Correspondence to:* Prof. Alberto Ruano-Ravina. Departamento de Medicina Preventiva y Salud Pública, Facultad de Medicina, C/San Francisco s/n, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, A Coruña, Spain. Email: alberto.ruano@usc.es.

*Provenance:* This is an invited Editorial commissioned by Executive Editor-in-Chief Jianxing He (Director of the Thoracic Surgery Department, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China).

*Comment on:* Li Y, Xiao X, Han Y, *et al.* Genome-wide interaction study of smoking behavior and non-small cell lung cancer risk in Caucasian population. *Carcinogenesis* 2018;39:336-46.

Submitted Sep 04, 2018. Accepted for publication Nov 05, 2018.

doi: 10.21037/jtd.2018.11.29

**View this article at:** <http://dx.doi.org/10.21037/jtd.2018.11.29>

Lung cancer poses a serious health problem worldwide being the first cancer in incidence and the leading cause of cancer death in developed countries (1). According to its histological characteristics lung cancer is subdivided into two types; the main one is non-small cell lung cancer (NSCLC) which originates from bronchial epithelial-cell precursors and small cell lung cancer (SCLC) that originates in neuroendocrine cells. NSCLC is the most abundant form of lung cancer comprising more than 80% of all lung cancers (2). It includes several subtypes such as adenocarcinomas, squamous cell carcinomas, and large cell carcinomas. Nowadays, squamous cell carcinoma incidence is decreasing while adenocarcinomas are increasing. Adenocarcinoma has become the predominant histological type worldwide, and this is more evident in women and never-smokers (3).

Despite much investigation, the causes for this change in histological characteristics are not fully understood. It is known that lung cancer is a multifactorial disease being tobacco the main risk factor followed by residential radon. Other factors have also shown to be associated with lung cancer risk, as professional occupation (especially those occupations involving exposure to asbestos) or leisure time activities (4-6). However, it is not clear why some

individuals, exposed to the same burden of carcinogens, develop lung cancer and others do not. It has been reported that only 10–20% of smokers develop lung cancer in their lifetime, which suggests the essential role that individual genetic susceptibility may play in the occurrence and development of lung cancer (7). The main explanation proposed for the decrease of squamous cell carcinomas and the increase of adenocarcinomas resides in changes in tobacco ingredients. The progressive decrease in nicotine content in cigarettes has led to more depth inhalation and tobacco smoke reaching the whole lung. Therefore, more central-borne tumors such as squamous carcinomas have changed to more peripheric tumors as adenocarcinoma (8).

Different human genomic regions located at chromosomes 15q25 (CHRNA5, CHRNA3 and CHRNA4), 5p15 (TERT), and 6p21 (HLA) have been identified in recent years to pose an association with individual susceptibility to lung cancer in genome-wide association studies (GWAS) (9-11). Despite the known importance of gene-environment interaction for lung cancer, few studies until now have assessed the gene-environment interactions on a genome-wide scale. Most studies have only focused on detecting the main effect of genetic variants. As shown in recent publications, the

interactions between the genetic polymorphisms and the fact of being or not being a smoker play a fundamental role in the development of this disease (12).

In a recently published study, Li *et al.* (13) investigated the effect of significant single nucleotide polymorphisms (SNPs) stratified by smoking status on NSCLC lung cancer risk. To do this, a genome-wide interaction analysis between SNPs and smoking status (never *vs.* ever smokers) was performed in a Caucasian population. They identified three novel SNPs with significant interactions with tobacco smoking: rs6441286, rs17723637 and rs4751674. The first two were identified for overall lung cancer risk whereas interaction of smoking with rs4751674 was detected only in the case of squamous cell lung carcinoma. They obtained an interaction odds ratio and meta-analysis P value of 1.24 with  $6.96 \times 10^{-7}$ , 1.37 with  $3.49 \times 10^{-7}$  and 0.58 with  $8.12 \times 10^{-7}$ , respectively.

Moreover, when the risk effect of these three SNPs was investigated in smokers and never-smokers individually, both the minor's alleles at SNP rs6441286 and at SNP rs17723637 had protective effect on NSCLC in the never smoking group with an overall OR of 0.83 (95% CI, 0.77–0.90) and 0.76 (95% CI, 0.68–0.85), respectively. These protective effects are not existing in smokers. However, contrary to the results obtained previously, a risk effect was found for squamous cell lung cancer in case of SNP rs4751674 with an overall OR of 1.66 (95% CI, 1.35–2.05) in individuals who have never smoked. These results suggest that this SNP is involved in squamous cell carcinoma in never-smokers. This may be interesting since squamous cell lung cancer is very related to tobacco smoking but a minor part of these cases occurs in never-smokers which makes us to consider a possible genetic basis.

Although few studies to date have researched the SNPs—tobacco smoke interactions on a genome-wide scale, in 2014 two SNPs, rs1316298 and rs4589502, were described by Zhang *et al.* (12) in a genome-wide gene-smoking interaction scanning with a sample size of 3,865 cases and 4,566 controls carried out in a Han Chinese population. They obtained a negative interaction between rs1316298 and the fact of smoking showing an OR of 0.71 and a positive interaction between rs4589502 and smoking with and OR of 1.55. These results should be treated with caution since it was carried out in a very specific population (Han Chinese) which complicates, as we have already mentioned, the extrapolation to the general population. A negative interaction between rs1316298 and smoking behavior was identified, similar to the one described for

SNP rs4751674 published by Li *et al.* (13) suggesting that smoking exposure decreases the genetic risk of lung cancer disease in these two concrete genetic locus.

A two-step test method was proposed by Murcay *et al.* (14) for gene-environment interaction analysis. This method, in contrast to the traditional test, is based on incorporating a preliminary selection step. For this reason, in the current study, a genome-wide interaction analysis was initially carried out encompassing a discovery stage that consists in identifying candidate SNPs testing the association between SNPs and smoking behavior performing only on cases and next, a case and control replication study was performed to validate the candidate SNPs elected in the previous stage.

The studies investigating the genetic risk factors predisposing to lung cancer have been initially performed with a candidate-gene approach. In this way, multiple genes involved in carcinogen metabolism and in DNA-repair have been identified as probable risk factors of lung cancer (15–17). An important limitation is that it is necessary to know a priori which genes are the ones implicated in carcinogenesis. Currently, with the development of the new technologies, these types of genetic studies have evolved towards a larger scale GWAS approach, allowing the sequencing of multiple genetic variants simultaneously without having the need to understand previously the behavior of these genes. Furthermore, tobacco contains several chemical carcinogens which are metabolized through different pathways implying many genes at the same time. Some of them activate carcinogenic compounds to join covalently to DNA, others repair these DNA-adducts and other genes convert these carcinogens in more easily excretable substances. To have the overall picture of tobacco effect on lung cancer risk is therefore necessary to analyze dozens on genes with its different polymorphisms, and this is complicated with the candidate gene approach embedded in a case-control study with a few participating hospitals.

GWAS have also limitations. Many of these studies have made an important selection bias selecting a specific fraction of the population which does not necessarily represent correctly the worldwide population. For example, a very common failure is to select only a single-ethnic population or individuals with the same single smoking status. In fact, an important limitation of the commented study is precisely that it was carried out exclusively in a Caucasian population, which clearly reduces its external validity and extrapolation to the general population. We find this same limitation in

studies published by other authors. For instance, Dong *et al.* (18) described in the Chinese population novel lung cancer susceptibility genes but there are other studies more restrictive performed only in never-smokers (19) and even in a population of never-smoking Asian women (20). Therefore, it would be interesting to conduct research on multiethnic populations.

Other important aspect to consider is to take into account the diversity of smoking statuses (never *vs.* ever-smokers) and the different histologic types of NSCLC, which allows to reflect in a more reliable way the general population. These last two aspects have been considered by Li *et al.* (13), in which analysis was stratified by smoking status (never *vs.* ever smokers) and by histological subtypes (NSCLC, adenocarcinoma and squamous cell lung cancer).

It is also important to highlight the importance of this study due to its large sample size using genotype from 35,737 individuals including both discovery and validation datasets, becoming distantly in the largest genome-wide SNP-smoking interaction analysis reported for lung cancer. The discovery genotype data in this research was obtained from OncoArray consortium (21), whose objective is, through the effort made by worldwide researchers, to better understand the genetic susceptibility and its relationship with carcinogenesis, trying to detect risk genetic biomarkers for the different types of common cancer, for example lung cancer. Sharing data of the consortium helps to increase the statistical power of the study, especially when analysis by subgroups are performed.

But these three novel SNPs described in this article have not only been investigated for their relation to lung cancer. For example, in the case of SNP rs6441286 apart from the lung cancer association it has also been assessed in its association with other diseases (22,23). Previous GWAS conducted on both European and Han Chinese populations have revealed that SNP rs6441286 is strongly associated with primary biliary cirrhosis.

As we have commented previously, tobacco is the main risk factor for developing a lung cancer, however, in spite of all the efforts made from public health, smokers yet represent a considerable fraction of the population. To ensure that teenagers do not initiate in the smoking habit continues to be considered as the best preventive measure that can be carried out. As shown in the results published by Li *et al.* it is relevant to detect those genetic polymorphisms that, due to their interaction with tobacco exposure, could modify the individual risk of lung cancer. In this way, it could be predicted in a more detailed way which individuals

among smokers is more or less probably to suffer a lung cancer related to tobacco exposure, thus identifying high-risk subgroups in the population.

Further gene-smoking interactions studies should be developed in order to understand in a more reliable way the SNPs—environmental interaction and the genetic individual susceptibility. It is very likely that with the imminent technological development other approaches may emerge in the coming years to better investigate the lung and other types of cancer tumorigenesis.

## Acknowledgements

None

## Footnote

*Conflicts of Interest:* The authors have no conflicts of interest to declare.

## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;66:7-30.
2. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med* 2008;359:1367-80.
3. Devesa SS, Bray F, Vizcaino AP, et al. International lung cancer trends by histologic type: male:female differences diminishing and adenocarcinoma rates rising. *Int J Cancer* 2005;117:294-9.
4. Barros-Dios JM, Ruano-Ravina A, Pérez-Ríos M, et al. Residential radon exposure, histologic types, and lung cancer risk. A case-control study in Galicia, Spain. *Cancer Epidemiol Biomarkers Prev* 2012;21:951-8.
5. Nielsen LS, Bælum J, Rasmussen J, et al. Occupational asbestos exposure and lung cancer--a systematic review of the literature. *Arch Environ Occup Health* 2014;69:191-206.
6. Ruano-Ravina A, García-Lavandeira JA, Torres-Durán M, et al. Leisure time activities related to carcinogen exposure and lung cancer risk in never smokers. A case-control study. *Environ Res* 2014;132:33-7.
7. Hecht SS. Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. *Lancet Oncol* 2002;3:461-9.
8. Stellman SD, Muscat JE, Thompson S, et al. Risk of squamous cell carcinoma and adenocarcinoma of the lung in relation to lifetime filter cigarette smoking. *Cancer*

- 1997;80:382-8.
9. Sun Y, Li J, Zheng C, et al. Study on polymorphisms in CHRNA5/CHRNA3/CHRNA4 gene cluster and the associated with the risk of non-small cell lung cancer. *Oncotarget* 2017;9:2435-44.
  10. Yuan Y, Lu C, Xue L, et al. Association between TERT rs2736100 polymorphism and lung cancer susceptibility: evidence from 22 case-control studies. *Tumour Biol* 2014;35:4435-42.
  11. 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;102:959-71.
  12. Zhang R, Chu M, Zhao Y, et al. A genome-wide gene-environment interaction analysis for tobacco smoke and lung cancer susceptibility. *Carcinogenesis* 2014;35:1528-35.
  13. Li Y, Xiao X, Han Y, et al. Genome-wide interaction study of smoking behavior and non-small cell lung cancer risk in Caucasian population. *Carcinogenesis* 2018;39:336-46.
  14. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. *Am J Epidemiol* 2009;169:219-26.
  15. Ruano-Ravina A, Pereyra MF, Castro MT, et al. Genetic susceptibility, residential radon, and lung cancer in a radon prone area. *J Thorac Oncol* 2014;9:1073-80.
  16. Kiyohara C, Horiuchi T, Takayama K, et al. Genetic polymorphisms involved in carcinogen metabolism and DNA repair and lung cancer risk in a Japanese population. *J Thorac Oncol* 2012;7:954-62.
  17. López-Cima ME, González-Arriaga P, García-Castro L, et al. Polymorphisms in XPC, XPD, XRCC1, and XRCC3 DNA repair genes and lung cancer risk in a population of northern Spain. *BMC Cancer* 2007;7:162.
  18. Dong J, Hu Z, Wu C, et al. Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population. *Nat Genet* 2012;44:895-9.
  19. 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;11:321-30.
  20. 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;44:1330-5.
  21. 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;26:126-35.
  22. Li P, Lu G, Cui Y, et al. Association of IL12A Expression Quantitative Trait Loci (eQTL) With Primary Biliary Cirrhosis in a Chinese Han Population. *Medicine (Baltimore)* 2016;95:e3665.
  23. Hirschfield GM, Liu X, Xu C, et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N Engl J Med* 2009;360:2544-55.

**Cite this article as:** Lorenzo-González M, Fernández-Villar A, Ruano-Ravina A. Disentangling tobacco-related lung cancer—genome-wide interaction study of smoking behavior and non-small cell lung cancer risk. *J Thorac Dis* 2019;11(1):10-13. doi: 10.21037/jtd.2018.11.29