



# Methylation, smoking, and reduced lung function

Stephanie J. London

**Affiliation:** Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Dept of Health and Human Services, Research Triangle Park, NC, USA.

**Correspondence:** Stephanie J. London, NIEHS, PO Box 12233, MD A3-05, RTP, NC 27709, USA.  
E-mail: london2@niehs.nih.gov

 @ERSpublications

**Epigenome-wide methylation studies have identified biomarkers of lifetime smoking; this study shows these are also related to lung function. While mechanisms are not clear, these studies could identify gene targets for prevention, screening and treatment.** <http://bit.ly/30F63i2>

**Cite this article as:** London SJ. Methylation, smoking, and reduced lung function. *Eur Respir J* 2019; 54: 1900920 [<https://doi.org/10.1183/13993003.00920-2019>].

IMBODEN *et al.* [1] have produced a well-conducted epigenome-wide association study examining DNA methylation in blood across the genome in relation to lung function. The major finding is that various sites in the DNA from blood, previously identified in many studies to be strongly differentially methylated by smoking, are also reproducibly differentially methylated in relation to lung function in smokers.

The authors suggest that these smoking-related methylation signals might be on the causal pathway for the well-documented and strong effects of smoking on lung function measures. They based this suggestion on some additional analyses. They replicated the C-phosphate-G sites (CpGs) in DNA that they found in their discovery cohorts to be differentially methylated in relation to lung function, in additional study populations. They found that all of the replicated CpGs had previously been identified as differentially methylated in relation to smoking in the literature. They then adjusted the analysis of methylation in relation to lung function for smoking status and pack-years, based on questionnaire self-report, and thus identified the small proportion of their replicated smoking-related CpGs that remained significantly associated with lung function after taking smoking into account. To investigate whether these replicated methylation signals, related to both lung function and smoking, might be on the causal pathway between exposure and lung function impairment, they used two statistical techniques commonly employed to interpret epigenome-wide association study findings: Mendelian randomisation and mediation analysis [2]. Below I discuss some issues that make determining whether smoking-related methylation signals are causal for smoking-related disease challenging.

The authors interpret the finding that the smoking-related methylation signals remain significantly associated with lung function after adjustment for questionnaire-based smoking metrics as strengthening the likelihood that CpGs differentially methylated by smoking might be causal for effects of this exposure on these outcomes. However, they acknowledge that this might occur because these CpGs capture smoking history better than questionnaire-based metrics. The fact that such a small proportion of the CpGs that replicated in their additional studies remained associated with lung function after adjustment for questionnaire-based smoking metrics confirms that unadjusted analyses of the association between methylation and lung function are substantially confounded by smoking, which could raise the possibility of incomplete exposure adjustment. Notably, the top CpG among those that both replicated and survived smoking adjustment was cg05575921 in the aryl hydrocarbon receptor repressor (*AHRR*) gene. This CpG is among the top

---

Received: May 08 2019 | Accepted after revision: May 10 2019

The content of this work is not subject to copyright. Design and branding are copyright ©ERS 2019

genome-wide significant sites differentially methylated in relation to smoking in nearly all studies, both in adults from their own smoking [3] and in newborns from maternal smoking during pregnancy [4]. *AHRR* cg05575921 is such a remarkable biomarker of lifetime smoking that it has been patented as a commercial test for the insurance industry [5]. Not only this, but many other smoking-related CpGs, quantitatively capture lifetime smoking history, and for past smokers, the time since quitting [3, 6, 7].

The discovery of reliable biomarkers of lifetime smoking history is a major contribution of epigenome-wide methylation studies. Previously, only biomarkers of recent smoking, mainly cotinine, were available, and so lifetime smoking history could only be assessed using self-reported questionnaire variables. The quantitative measure of lifetime smoking history estimated from questionnaires is pack-years, generally calculated by multiplying the usual number of cigarettes per day by the number of years of smoking. Pack-years can underestimate lifetime smoking history for several reasons. First, some smokers report being nonsmokers: about 5% in a representative sample of the US population [8]. Second, in developed countries, public health efforts to discourage smoking over recent decades have been successful in reducing both smoking prevalence and daily cigarette consumption among smokers [9]. Thus, the number of cigarettes per day reported at entry into a study is likely to be lower than amounts smoked earlier. In addition, some surveys only classify individuals as smokers if they smoked daily for some minimum period of time, perhaps 6 months. However, a substantial proportion of smokers do not smoke daily [10] and thus would not be analysed as smokers in some studies. Notably, *AHRR* cg05575921, the top CpG related to both smoking and lung function in this study, appears to indicate smoking history of as little as a half pack-year [11]. This and other smoking-related CpGs also reflect smoking of other tobacco products, including pipes, cigars and marijuana cigarettes [6], which impact lung function [12]. Maternal smoking during pregnancy also has an influence on lung function [12]. Smoking-related CpGs also are reliable biomarkers of maternal smoking during pregnancy in newborns [13], in children [14] and even adults [15]. Finally, even if the self-reported amount and duration of smoking were 100% accurately measured by smoking status and pack-years, differences in exposure to tobacco combustion products across cigarette brands and over time are typically not captured. In contrast, smoking-related differential methylation CpG biomarkers capture all of the additional relevant exposure sources mentioned above. Importantly, they also reflect the biologically relevant internal dose that is impacted by all of these exposures combined with individual variability in metabolism. Hence, it is not surprising that a small proportion of methylation signals that are most strongly related to lifetime smoking history remain significantly related to lung function, even after adjustment for questionnaire-based smoking metrics.

The authors included Mendelian randomisation analysis. In the current paper, this involved using single nucleotide polymorphisms, significantly related to methylation levels of the smoking-related CpGs, as genetic instrumental variables to infer causality. The authors were able to identify genetic instrumental variables for eight of the 57 CpGs that replicated in additional cohorts; these analyses included CpGs that no longer remained related to lung function after smoking adjustment. Unfortunately, the top CpG *AHRR* cg05575921 had no genetic instrument. The authors report that the results support causal effects. However, additional detail on strength of the associations of the genetic instrumental variables with methylation levels of the CpGs, would be useful to properly evaluate the Mendelian randomisation conclusions. Limitations of Mendelian randomisation of this method for drawing causal conclusions regarding methylation results, including weakness of the genetic instrumental variables used, power issues, and potential biases, have been recently discussed [2, 16].

The authors also used mediation analysis, another statistical technique [17] to assess the likelihood that an association in observational data is causal. They formed an index of several smoking-related CpGs recently reported to mediate the effects of smoking on lung function in one of their replication cohorts [18] and they again found evidence of mediation. However, in mediation analyses of smoking-related CpGs in relation to a smoking-related health outcome, when the smoking-related CpGs better reflect the relevant smoking exposure than self-reported smoking, false-positive evidence of mediation can result [19]. As noted above, substantial evidence indicates that smoking-related methylation signals are superior biomarkers of lifetime smoking history compared with questionnaire-based metrics. While correction for exposure measurement error can mitigate, to some degree, the resulting false-positive mediation results [19], it was not used here. Even when measurement error correction methods are applied, extreme caution is required in interpreting mediation analyses as evidence that smoking-related CpGs cause smoking-related outcomes [19]; in the current study, reduced lung function. The authors note that the top CpG for both smoking and lung function, *AHRR* cg05575921, is also differentially methylated in relation to smoking in lung macrophages [20], strengthening the plausibility of a causal role in lung pathogenesis. However, the lung is directly exposed to smoking combustion products and finding smoking-related *AHRR* differential methylation may be interpreted as supporting the validity of some blood-based methylation biomarkers to detect exposure in disease-relevant target tissues [21].

Although many studies document widespread gene-specific methylation differences according to smoking, the mechanism for these specific exposure effects is not well understood. A leading hypothesis is the transcription factor occupancy theory [22], whereby transcription factors activated or repressed in response to exposure either deny or allow access to the DNA methylation machinery, resulting in gene-specific reduced or increased methylation. Thus, smoking-related DNA methylation changes may themselves result from alterations of transcription factor binding. Alternatively, or in addition, exposure-related methylation changes may be proxies for histone modifications that lead to altered gene function and exposure-related disease pathogenesis [2]. Under these alternative biological mechanisms, or ones yet to be discovered, attempting to determine whether exposure-related methylation differences are themselves the causes of exposure-induced disease, might be expecting too much of available statistical techniques, such as Mendelian randomisation or mediation [2].

Given these complicating issues, can epigenome-wide methylation studies help us understand mechanisms of exposure-related disease processes? Can they inform risk prediction for lung function decline? When considered in light of the various cautions discussed above, analyses such as those presented in this paper can help to highlight the specific smoking–methylation signals worthy of follow-up in mechanistic studies for their role in smoking-related lung function impairment. Further, identifying robust circulating biomarkers of smoking, that reflect processes in target tissues, can facilitate disease risk prediction, even if they are not the underlying causal events. Valid quantitative biomarkers can also help identify previously undetected health effects of exposure to smoking combustion products. In addition, many environmental exposures, such as ambient air pollutants, have effects that are weaker or harder to detect than those of smoking. In this setting having robust smoking biomarkers to remove residual confounding by smoking has important public health relevance. Indeed, many CpGs differentially methylated in relation to smoking are correlated with gene expression, highlighting their potential biological relevance [3]. In addition, epigenome-wide association studies of smoking have identified genes that had not been previously shown to play a role in biological responses to this environmental cause of myriad adverse health outcomes. Surprisingly, despite many years of research, the specific mechanisms underlying many health effects of smoking remain incompletely understood. Studies such as the current one can identify new gene targets for prevention, screening, and treatment of lung disease. Thus, integrating results from epigenome-wide association studies of smoking and smoking-related health outcomes, such as lung function in this study, has great potential clinical relevance.

Conflict of interest: S.J. London has nothing to disclose.

Support statement: S.J. London is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences, ZO1 ES43012. Funding information for this article has been deposited with the Crossref Funder Registry.

## References

- 1 Imboden M, Wielscher M, Rezwani FI, *et al.* Epigenome-wide association study of lung function level and its change. *Eur Respir J* 2019; 54: 1900457.
- 2 Tobi EW, van Zwet EW, Lumey LH, *et al.* Why mediation analysis trumps Mendelian randomization in population epigenomics studies of the Dutch Famine [pre-print]. *bioRxiv* 2018: 362392.
- 3 Joehanes R, Just AC, Marioni RE, *et al.* Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet* 2016; 9: 436–447.
- 4 Joubert BR, Felix JF, Yousefi P, *et al.* DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. *Am J Hum Genet* 2016; 98: 680–696.
- 5 Andersen AM, Ryan PT, Gibbons FX, *et al.* A droplet digital PCR assay for smoking predicts all-cause mortality. *J Insur Med* 2018; 47: 220–229.
- 6 Philibert R, Dogan M, Noel A, *et al.* Dose response and prediction characteristics of a methylation sensitive digital PCR assay for cigarette consumption in adults. *Front Genet* 2018; 9: 137.
- 7 Zhang Y, Schottker B, Florath I, *et al.* Smoking-associated DNA methylation biomarkers and their predictive value for all-cause and cardiovascular mortality. *Environ Health Perspect* 2016; 124: 67–74.
- 8 Caraballo RS, Giovino GA, Pechacek TF, *et al.* Racial and ethnic differences in serum cotinine levels of cigarette smokers: Third National Health and Nutrition Examination Survey, 1988–1991. *JAMA* 1998; 280: 135–139.
- 9 Ng M, Freeman MK, Fleming TD, *et al.* Smoking prevalence and cigarette consumption in 187 countries, 1980–2012. *JAMA* 2014; 311: 183–192.
- 10 Schane RE, Glantz SA, Ling PM. Nondaily and social smoking: an increasingly prevalent pattern. *Arch Intern Med* 2009; 169: 1742–1744.
- 11 Andersen AM, Philibert RA, Gibbons FX, *et al.* Accuracy and utility of an epigenetic biomarker for smoking in populations with varying rates of false self-report. *Am J Med Genet B Neuropsychiatr Genet* 2017; 174: 641–650.
- 12 National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. *In: US Department of Health and Human Services, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014.*
- 13 Reese SE, Zhao S, Wu MC, *et al.* DNA methylation score as a biomarker in newborns for sustained maternal smoking during pregnancy. *Environ Health Perspect* 2017; 125: 760–766.

- 14 Ladd-Acosta C, Shu C, Lee BK, *et al.* Presence of an epigenetic signature of prenatal cigarette smoke exposure in childhood. *Environ Res* 2016; 144: 139–148.
- 15 Richmond RC, Suderman M, Langdon R, *et al.* DNA methylation as a marker for prenatal smoke exposure in adults. *Int J Epidemiol* 2018; 47: 1120–1130.
- 16 Pingault J-B, O'Reilly PF, Schoeler T, *et al.* Using genetic data to strengthen causal inference in observational research. *Nat Rev Genet* 2018; 19: 566–580.
- 17 Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986; 51: 1173–1182.
- 18 de Vries M, van der Plaat DA, Nedeljkovic I, *et al.* From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function. *Respir Res* 2018; 19: 212.
- 19 Valeri L, Reese SL, Zhao S, *et al.* Misclassified exposure in epigenetic mediation analyses. Does DNA methylation mediate effects of smoking on birthweight? *Epigenomics* 2017; 9: 253–265.
- 20 Monick MM, Beach SR, Plume J, *et al.* Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet B Neuropsychiatr Genet* 2012; 159B: 141–151.
- 21 Bakulski KM, Dou J, Lin N, *et al.* DNA methylation signature of smoking in lung cancer is enriched for exposure signatures in newborn and adult blood. *Sci Rep* 2019; 9: 4576.
- 22 Martin EM, Fry RC. Environmental influences on the epigenome: exposure-associated DNA methylation in human populations. *Annu Rev Public Health* 2018; 39: 309–333.