Greater Than the Sum of Its Parts: Integrating Multiple Data Sets to Decipher Cigarette Smoking Effects on Airway Epithelium

Despite the fact that we have long understood that environmental exposures influence health, policies have been slow to reduce or eliminate those exposures. In an attempt to mitigate ongoing influences on health by environments, researchers are seeking molecular mechanisms that link exposure to outcome, with the goal of identifying targetable pathways and providing compelling evidence to drive policy change. Cigarette smoking in particular has a long list of adverse health effects, but the success of policies aimed at reducing exposure are complicated by tobacco dependence. Identification of specific genes or cellular functions that are dysregulated by cigarette smoke in physiologically relevant tissues both provides increased evidence for health impact and could provide the basis for novel treatments for those who struggle to stop smoking.

DNA methylation is a modification frequently studied because of its importance as an interface between environmental exposure and cellular effect (1). In this issue of the *Journal*, Reddy and colleagues (pp. 366–377) identify gene expression and DNA methylation differences between current smokers and ex-smokers in nasal epithelium. They also assessed the relationship between differential gene expression and DNA methylation, implicated potential pathways of interest, and finally assessed reversion of DNA methylation after smoking cessation (2). Studies examining associations between environmental exposures and either gene expression or DNA methylation typically suffer from a number of common limitations. This study mitigates some of these limitations, primarily through the creative use of external data sets.

Cell type specificity of gene expression and DNA methylation is a common challenge in two major ways. First, tissue specificity means that peripheral tissue patterns do not always reflect the central tissue of interest. For human studies it means that many diseases are difficult to study in anything other than postmortem tissue. A recent study showed that DNA methylation profiles from airway and nasal epithelium were better predictors of asthma than peripheral blood, which is a common sampling site in DNA methylation studies (3). Here, however, the authors compared their data on accessible nasal epithelium to previously collected bronchial epithelial data (4). Many of the differentially expressed genes had previously been identified as associated with cigarette smoking, either in epithelium or other tissues, and were replicated in external data of cultured bronchial epithelia exposed to cigarette smoke (4). Unfortunately, such data sets for DNA methylation were not available to validate the findings and would be a valid avenue for further study. Second, heterogeneity in cell composition within a tissue can confound analyses if it differs between groups (5, 6). In this manuscript, the authors used Cibersort to deconvolute the cell types in their gene expression data and show that there were no differences in cell composition between the groups (2). For DNA methylation, they also demonstrate that including cell composition in their analysis retained over 95% of the results. It is perhaps surprising that current versus ex-smokers do not show differences in cell type of the nasal epithelium, although it is possible that not all such changes were detectable using the Cibersort method.

When studying environmental exposure associations with gene expression and DNA methylation, the two are often approached in isolation, and when they are collected simultaneously, the integration of the two data sets can be challenging. Expression quantitative trait methylation (eQTM) analysis allows correlation of gene expression differences to CpG site methylation differences. Of the 809 genes and 18,814 CpGs that were associated with smoking status, only 171 gene:CpG pairs were identified, highlighting a high proportion of genes and CpG sites that associate with smoking status independently of each other. This analysis does not include possible genetic effects on both expression and DNA methylation (7). Here, it is important to consider that gene expression primarily represents a snapshot in time of the expression status when the sample was collected, whereas DNA methylation can also represent potential for changes in expression in the future. The potentially transient nature of these associations was further highlighted in this study by comparing their results to existing gene expression and DNA methylation data sets from smoking cessation studies, which showed that both gene expression and DNA methylation are reversed upon smoking cessation. Future analyses could examine relative representation of expression alone, DNA methylation alone, or eQTMs among genes or CpGs that showed reversal upon smoking cessation.

A final limitation of environmental studies is understanding the mechanism by which the exposure has an effect on the measured output. Here, the authors used gene expression pathway analysis to highlight an enrichment for aryl hydrocarbon receptor (AHR) and NRF2 pathways. Unsurprisingly, gene expression from publicly available data sets in cells where these factors are inhibited showed a high degree of overlap with differentially expressed genes from this study. Interestingly, however, they then showed that some of the smoking-associated eQTM DNA methylation sites were in potential AHR and/or NRF2 binding sites, as identified from available chromatin immunoprecipitation data sets, preliminarily linking smoking-related DNA methylation changes to transcription factor binding and subsequent changes in gene expression (2). Although preliminary, this hints at methodology and techniques that could be drawn on within the community to identify potential mechanisms and pathways that could then be studied *in vitro*.

In summary, this study is a very good example of creative use of external data to extend the reach of a small initial sample size. It has identified or confirmed a number of specific genes and

Originally Published in Press as DOI: 10.1165/rcmb.2021-0207ED on June 17, 2021

pathways that are promising candidates for further investigation. An attempt at a smoking score, which could have clinical or treatment implications, was hampered by a lack of external validation. Nonetheless, it has highlighted approaches to increase the validity (cell type composition) and extend the scope (mechanism, cessation, longitudinal data) of the original sample set. Finally, this study strengthens the argument for the benefits of smoking cessation at a molecular level. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Rachel L. Clifford, Ph.D.
Centre for Respiratory Medicine
University of Nottingham and Nottingham NIHR Biomedical Research Centre
Nottingham, United Kingdom

Meaghan J. Jones, Ph.D.
Department of Biochemistry and Medical Genetics
University of Manitoba
Manitoba, Canada

and

Children's Hospital Research Institute of Manitoba
Manitoba, Canada

ORCID ID: 0000-0003-1593-997X (M.J.J.).

References

1. Clifford RL, Jones MJ, MacIsaac JL, McEwen LM, Goodman SJ, Mostafavi S, *et al.* Inhalation of diesel exhaust and allergen alters human bronchial epithelium DNA methylation. *J Allergy Clin Immunol* 2017;139:112–121.
2. Reddy KD, Lan A, Boudewijn IM, Rathnayake SNH, Koppelman GH, Aliee H, *et al*. Current smoking alters gene expression and DNA methylation in the nasal epithelium of patients with asthma. *Am J Respir Cell Mol Biol* 2021;65:366–377.
3. Lin PI, Shu H, Mersha TB. Comparing DNA methylation profiles across different tissues associated with the diagnosis of pediatric asthma. *Sci Rep* 2020;10:151.
4. Moses E, Wang T, Corbett S, Jackson GR, Drizik E, Perdomo C, *et al.* Molecular impact of electronic cigarette aerosol exposure in human bronchial epithelium. *Toxicol Sci* 2017;155:248–257.
5. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* 2013; 31:142–147.
6. Esposito EA, Jones MJ, Doom JR, MacIsaac JL, Gunnar MR, Kobor MS. Differential DNA methylation in peripheral blood mononuclear cells in adolescents exposed to significant early but not later childhood adversity. *Dev Psychopathol* 2016;28:1385–1399.
7. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, *et al.* Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* 2013;2: e00523.