Translating genomics into risk prediction

Emily S Wan, Dawn L DeMeo

Cigarette smoking is a leading risk factor in the development of cardiovascular, pulmonary and malignant diseases worldwide.¹ Yet, it remains one of the most challenging environmental exposures to quantify. Rudimentary categorisations of 'never', 'former' and 'current' smoker capture only a fraction of the complexity associated with the exposure. Attempts to quantify cumulative cigarette smoke exposure, for example by calculating 'pack-years', are notoriously imprecise, partially because of a failure to account for variations in behaviour such as (1) depth of inhalations, retention time before exhalation and timing between puffs, (2) intermittent periods of abstinence or relapse, and (3) a reliance on self-reported recall of smoking habits over extended periods of time (often years to decades). Additionally, biological differences related to age, gender, body size, and variations in absorption and metabolism among smokers further complicate our attempts to quantify the 'effective exposure' to cigarette smoke for any given individual.

Given these complexities, the use of biomarkers, both qualitative and quantitative, as an objective measurement of 'effective exposure' is conceptually appealing. To date, the most widely used smoking biomarkers are cotinine and carbon monoxide (CO). Measurement of CO, either in exhaled breath or in blood as carboxyhaemoglobin, has been shown to correlate with smoking intensity, however the relatively short half-life of CO ($t_{1/2}$ 1–6 hours) limits its utility to assessments of *recent* smoking behaviours.² Cotinine, a direct metabolite of nicotine which is measurable in the blood, urine and tissues is another well validated biomarker of cigarette smoke exposure. The slightly longer half-life ($t_{1/2}$ 16–20 hours), as well as the assessment of cotinine in slower growing tissues such as hair and nail clippings, extends the potential period of monitoring for this assay to the order of weeks to months.³ However, neither cotinine nor CO are suitable for

assessments of long-term smoke exposure and the associations between these exposures and epidemiological outcomes such as mortality and cancer incidence.

DNA methylation, a process where methyl groups are covalently bonded to cytosine residues in genomic DNA, can reflect environmental exposures over long periods of time and has drawn interest as a possible mediator between the extended latency between cigarette smoke exposure and the development of chronic diseases such as atherosclerosis, COPD and lung cancer. Prior to the genomics revolution, evaluation of DNA methylation was resource intensive and largely focused on a small number of candidate genes such as imprinted loci and tumour suppressors. By modifying the technology employed in genome-wide genetic studies, *quantitative* assessments of DNA methylation at tens to hundreds of thousands of sites throughout the genome became possible⁴ and led to the identification of numerous, previously unsuspected smoking-associated loci.^{5–7}

A site located in the third intron of the aryl hydrocarbon receptor repressor (AHRR) gene (OMIM 606517), known by its industry-assigned name as 'cg05575921', has been consistently identified in epigenome-wide association studies of blood and lung-derived tissues as one of the most strongly associated sites with smoking.⁶ Active smoking causes progressive demethylation at cg05575921 in blood-derived leucocytes.^{6,8,9} In addition to reflecting current smoking status, a strong inverse correlation between methylation levels and self-reported measures of smoking *intensity* exists and can be corroborated through established biomarkers such as serum cotinine.¹⁰ In former smokers, methylation levels at cg05575921 in blood are correlated with time since quitting,^{8–10} with evidence of active remethylation detectable as early as a few months following cessation.⁹ Notably, remethylation at this locus is often incomplete, even after years of abstinence from smoking; thus methylation at cg05575921 is unique among smoking biomarkers in the ability to distinguish former smokers from never smokers.¹⁰

Dynamic changes in methylation at cg05575921 in blood may provide

additional information on the *biological impact* of chronic cigarette smoke exposure on the risk of the future development of disease. Associations between methylation at cg05575921 and subclinical carotid atherosclerosis¹¹ and cardiovascular mortality¹² have been reported in observational cohorts. Case-control studies have found that methylation levels at AHRR, both alone and in conjunction with data from other smoking-related methylation sites, was strongly associated with the future development of lung cancer¹³ and lung cancer-specific mortality.¹⁴ Importantly, in many of these studies, AHRR methylation levels remained predictive *even after adjustment* for self-reported smoking parameters.

In *Thorax*, Bojesen *et al*¹⁵ examine the associations between methylation at the AHRR and morbidity and mortality associated with smoking-related diseases in a large cohort with extended longitudinal follow-up (median 19 years). The authors begin by assessing the performance of a targeted, PCR-based assay for methylation at cg05575921 in blood. The validity of the assay is supported by its ability to detect differential methylation associated with self-reported smoking variables as well as a biologically plausible interaction between methylation levels and a genetic variant (rs1051730) within the nicotinic acetylcholine receptor gene (*CHRNA3*) which has been associated with nicotine addiction and smoking intensity.¹⁶ The authors then identify associations between methylation at cg05575921 in blood and all-cause mortality, COPD hospitalisations and, among ever-smokers, lung cancer incidence. Among these, the association with COPD hospitalisations represents the most novel finding, though a potential limitation in this analysis is the lack of spirometric data and the possible misdiagnosis of COPD exacerbations in patients presenting clinically with acute dyspnoea.

Lastly, the authors present an intriguing analysis in a subset of high-risk ever smokers eligible for CT screening which demonstrates that the incorporation of cg05575921 methylation data allows for additional risk stratification in subjects with similar risk estimates based on the modified Prostate, Lung, Colorectal and Ovarian (PLCO_{M2012}) Score.¹⁷ Although too cumbersome for clinical applications, the PLCO_{M2012} Score, which is calculated using detailed information on smoking behaviour, demographic, socioeconomic, and personal and family history, has been shown to have increased sensitivity and comparable specificity relative to the less selective criteria typically employed for

Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA

Correspondence to Dr Emily S Wan, Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA; emily.wan@channing.harvard.edu

CT screening based upon the National Lung Screening Trial.¹⁸ Thus, the potential utility of a simple biological assay to provide meaningful data which could potentially obviate unnecessary screening in low-risk individuals as well as target more intensive screening in higher-risk individuals is obvious. However, several steps remain to be taken before we can truly translate these findings into clinical practice. First, validation and generalisability of the findings reported by Bojesen *et al*¹⁵ in independent cohorts should be demonstrated. The population studied in this manuscript, though large, is relatively isolated and genetically homogeneous. Several studies have demonstrated that differential methylation by ancestry exists, with at least one study demonstrating reduced sensitivity of *AHRR* methylation in identifying current smokers among populations not of European ancestry.¹⁹ Second, the values reported in this manuscript are relative (quintiles) and do not take into account the possibility of tissue-type and/or cellular heterogeneity. DNA methylation patterns differ by tissue type and, because whole blood is comprised of a collection of distinct cell types in variable proportions, assessing the impact of differences in cell count will be critical in the development of a standardised assay for which defined absolute cut-offs and performance attributes can be obtained. Finally, randomised clinical trials to test the added value of an *AHRR* methylation assay relative to current standards of care will need to be conducted.

A significant lag between variant discovery and translation into clinical practice exists for most genomics-based discoveries due to the modest effect sizes and the relative paucity of knowledge on how such variants are mechanistically related to disease development; most discoveries require years of additional work in functional laboratories before potential therapeutic strategies are developed. In contrast, the effect sizes from methylation studies at the *AHRR* locus are larger than those observed from most genetic studies and also demonstrate a direct association with public health outcomes. These attributes, which are applied towards several clinically relevant end points in the

manuscript by Bojesen *et al*, may allow for future application of *AHRR* methylation data towards clinical care guidelines—more research is needed.

Given the accelerating pace of translation of genomics-based discoveries into clinical care, practitioners will likely be called upon to administer, apply and interpret these discoveries in the care of their patients in the near future. However, while the majority of healthcare providers believe that genomics-based discoveries and developments will directly impact future clinical practice and general interest in personalised medicine is high, genomics knowledge in both healthcare providers and the lay public is low.²⁰ Increasing the number of opportunities and types of forums for education to elevate the level of genomic literacy in both patients and healthcare providers will be an essential component in realising the promise of the precision medicine movement.

Competing interests None declared.

Provenance and peer review Commissioned; internally peer reviewed.

CrossMark

To cite Wan ES, DeMeo DL. *Thorax* 2017;**72**:598–599.

Published Online First 7 March 2017

► <http://dx.doi.org/10.1136/thoraxjnl-2016-208789>

Thorax 2017;**72**:598–599.

doi:10.1136/thoraxjnl-2016-209934

REFERENCES

- Ng M, Freeman MK, Fleming TD, *et al*. Smoking prevalence and cigarette consumption in 187 countries, 1980–2012. *JAMA* 2014;311:183–92.
- Scherer G. Carboxyhemoglobin and thiocyanate as biomarkers of exposure to carbon monoxide and hydrogen cyanide in tobacco smoke. *Exp Toxicol Pathol* 2006;58:101–24.
- Benowitz NL, Hukkanen J, Jacob P, III. Nicotine chemistry, metabolism, kinetics and biomarkers. *Handb Exp Pharmacol* 2009;29–60.
- Dedeurwaerder S, Defrance M, Calonne E, *et al*. Evaluation of the Infinium Methylation 450K technology. *Epigenomics* 2011;3:771–84.
- Breitling LP, Yang R, Korn B, *et al*. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 2011;88:450–7.
- 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:141–51.
- Wan ES, Qiu W, Baccarelli A, *et al*. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet* 2012;21:3073–82.
- Flanagan JM, Brook MN, Orr N, *et al*. Temporal stability and determinants of White blood cell DNA methylation in the breakthrough generations study. *Cancer Epidemiol Biomarkers Prev* 2015;24:221–9.
- Philibert R, Hollenbeck N, Andersen E, *et al*. Reversion of *AHRR* demethylation is a quantitative biomarker of smoking cessation. *Front Psychiatry* 2016;7:55.
- Zhang Y, Florath I, Saum KU, *et al*. Self-reported smoking, serum cotinine, and blood DNA methylation. *Environ Res* 2016;146:395–403.
- Reynolds LM, Wan M, Ding J, *et al*. DNA methylation of the aryl hydrocarbon receptor repressor associations with cigarette smoking and subclinical atherosclerosis. *Circ Cardiovasc Genet* 2015;8:707–16.
- Zhang Y, Schöttker 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.
- Fasanelli F, Baglietto L, Ponzi E, *et al*. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nat Commun* 2015;6:10192.
- Zhang Y, Breitling LP, Balavarca Y, *et al*. Comparison and combination of blood DNA methylation at smoking-associated genes and at lung cancer-related genes in prediction of lung cancer mortality. *Int J Cancer* 2016;139:2482–92.
- Bojesen SE, Timpson N, Relton C, *et al*. *AHRR* (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax* 2017;72:646–53.
- Thorgeirsson TE, Gudbjartsson DF, Surakka I, *et al*. Sequence variants at *CHRNA3-CHRNA6* and *CYP2A6* affect smoking behavior. *Nat Genet* 2010;42:448–53.
- Tammemägi MC, Katki HA, Hocking WG, *et al*. Selection criteria for lung-cancer screening. *N Engl J Med* 2013;368:728–36.
- Aberle DR, Adams AM, Berg CD, *et al*. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
- Elliott HR, Tillin T, McArdle WL, *et al*. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics* 2014;6:4.
- Scheuner MT, Sieverding P, Shekelle PG. Delivery of genomic medicine for common chronic adult diseases: a systematic review. *JAMA* 2008;299:1320–34.