# Smoking Methylation Marks for Prediction of Urothelial Cancer Risk

Chenglong Yu[1], Kristina M. Jordahl[2], Julie K. Bassett[3], Jihoon Eric Joo[4], Ee Ming Wong[1], Maree T. Brinkman[3], Daniel F. Schmidt[5], Damien M. Bolton[6], Enes Makalic[5], Theodore M. Brasky[7], Aladdin H. Shadyab[8], Lesley F. Tinker[2], Anthony Longano[9], John L. Hopper[5], Dallas R. English[3,5], Roger L. Milne[1,3,5], Parveen Bhatti[10], Melissa C. Southey[1,3,4], Graham G. Giles[1,3,5], and Pierre-Antoine Dugué[1,3,5]

## ABSTRACT

**Background:** Self-reported information may not accurately capture smoking exposure. We aimed to evaluate whether smoking-associated DNA methylation markers improve urothelial cell carcinoma (UCC) risk prediction.

**Methods:** Conditional logistic regression was used to assess associations between blood-based methylation and UCC risk using two matched case–control samples: 404 pairs from the Melbourne Collaborative Cohort Study (MCCS) and 440 pairs from the Women's Health Initiative (WHI) cohort, respectively. Results were pooled using fixed-effects meta-analysis. We developed methylation-based predictors of UCC and evaluated their prediction accuracy on two replication data sets using the area under the curve (AUC).

**Results:** The meta-analysis identified associations ($P < 4.7 \times 10^{-5}$) for 29 of 1,061 smoking-associated methylation sites, but these were substantially attenuated after adjustment for self-reported smoking. Nominally significant associations ($P < 0.05$) were found for 387 (36%) and 86 (8%) of smoking-associated markers without/with adjustment for self-reported smoking, respectively, with same direction of association as with smoking for 387 (100%) and 79 (92%) markers. A Lasso-based predictor was associated with UCC risk in one replication data set in MCCS [$N = 134$; odds ratio per SD (OR) = 1.37; 95% CI, 1.00–1.90) after confounder adjustment; AUC = 0.66, compared with AUC = 0.64 without methylation information. Limited evidence of replication was found in the second testing data set in WHI ($N = 440$; OR = 1.09; 95% CI, 0.91–1.30).

**Conclusions:** Combination of smoking-associated methylation marks may provide some improvement to UCC risk prediction. Our findings need further evaluation using larger data sets.

**Impact:** DNA methylation may be associated with UCC risk beyond traditional smoking assessment and could contribute to some improvements in stratification of UCC risk in the general population.

## Introduction

Urothelial cell carcinoma (UCC) is a type of malignancy arising from the urothelium. Although UCC accounts for more than 90% of urinary bladder cancers (1), some can also be found in the proximal urethra, the transitional epithelium of the renal pelvis, and the ureter (2). According to Global Cancer Statistics, bladder cancer was in 2020 the 12th most common cancer worldwide, with an estimated 573,000 new cases and 212,000 new deaths (3). Cigarette smoking has been established as a strong risk factor for UCC with approximately half of newly diagnosed patients reporting a history of smoking (4, 5). Many studies (6–9) have investigated the association between smoking and risk of UCC, and a meta-analysis of 89 observational studies reported an increased risk of bladder cancer for current smokers [odds ratio (OR) = 3.1; 95% confidence interval (CI) = 2.5–3.7] and former smokers (OR = 1.8; 95% CI, 1.5–2.1), compared with never smokers (10). However, information on smoking history used in most epidemiologic studies, such as smoking status (never, former, or current smoker) or pack-years, is typically collected via self-report and may be prone to substantial measurement error. The accuracy of self-reported information has also been questioned because of declining response rates and the increasing social stigmatization of smoking (11). Furthermore, such information cannot reflect secondhand smoke exposure during childhood or adulthood. Therefore, such less accurate information would have potential impact on studies of disease association and risk prediction.

Serum or urinary cotinine (12) and blood DNA methylation (13–16) have been established as valid biomarkers of cigarette smoking exposure. Although cotinine and methylation markers showed similar accuracy in distinguishing current from never smokers, only methylation markers can distinguish former from never smokers with high accuracy (17). Therefore, DNA methylation markers measured in

[1]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia. [2]Division of Public Health Sciences, Fred Hutchison Cancer Research Center, Seattle, Washington. [3]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia. [4]Department of Clinical Pathology, Melbourne Medical School, The University of Melbourne, Parkville, VIC, Australia. [5]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC, Australia. [6]Department of Surgery, University of Melbourne and Olivia Newton-John Cancer Centre, Austin Hospital, Melbourne, Australia. [7]Division of Medical Oncology, The Ohio State University College of Medicine, Columbus, Ohio. [8]Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, California. [9]Department of Anatomical Pathology, Eastern Health, Box Hill Hospital, Box Hill, VIC, Australia. [10]Cancer Control Research, BC Cancer Research Centre, Vancouver, Canada.

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

**Corresponding Author:** Pierre-Antoine Dugué, Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, 246 Clayton Road, Clayton, VIC 3168, Australia. Phone: 61-3-8572-2097; E-mail: pierre-antoine.dugue@monash.edu

Cancer Epidemiol Biomarkers Prev 2021;XX:XX–XX

**doi:** 10.1158/1055-9965.EPI-21-0313

©2021 American Association for Cancer Research

blood, which may also reflect different individuals' responses to lifetime exposure, can be used to augment self-reported smoking data to help refine individual risk profiling of smoking-induced diseases (18–20).

Authors of several studies (21–23) have evaluated the association of genome-wide cytosine-guanine (CpG) methylation in blood DNA with risk of UCC. Jordahl and colleagues (23), for example, identified potential methylation-based markers of susceptibility to urothelial carcinoma of the bladder, using the Illumina Infinium HumanMethylation450 Bead Array (~450,000 probes) on prediagnostic blood collected in the Women's Health Initiative (WHI). They subsequently found that two previously identified smoking-associated CpG sites mediated the effect of smoking on bladder cancer risk (24). With the current study, we aimed to expand on previous research by identifying associations between smoking-associated DNA methylation and bladder cancer risk and by developing a predictor of UCC risk using smoking-associated DNA methylation measures.

## Materials and Methods

### Study participants

The Melbourne Collaborative Cohort Study (MCCS) is an Australian prospective cohort study of 41,513 people recruited between 1990 and 1994 in the Melbourne metropolitan area. All participants were of white European origin. DNA was extracted from prediagnostic peripheral blood taken at recruitment (1990–1994) or at a subsequent follow-up visit (2003–2007) in participants free of UCC. More details about the cohort, blood collection, DNA extraction, and cancer ascertainment can be found elsewhere (22, 25). Information on tobacco use was self-reported by participants using questionnaires (24, 25). In this study, we utilized a case–control data set of urothelial cancer nested within the MCCS. Controls were matched to incident cases on age at blood draw, year of birth, sex, country of birth (Australia/New Zealand/UK/other, Italy, or Greece), sample collection period (baseline at recruitment or the follow-up visit), and sample type (peripheral blood mononuclear cells, dried blood spots, or buffy coats) using incidence density sampling. To minimize batch effects, samples from each matched case–control pair were plated to adjacent wells on the same BeadChip, with plate, chip, and position assigned randomly. We excluded from the analysis sex-discrepant and failed samples for DNA methylation measures. Case–control pairs with any missing values for the confounders measured were also excluded. Overall, 404 case–control pairs were included in the present study.

For replication and meta-analysis, we included the study sample previously used by Jordahl and colleagues (23, 24), which comprises 440 cases diagnosed with urothelial carcinoma of the bladder and 440 cancer-free controls matched on year of enrollment, age at enrollment (±2 years), follow-up time greater than or equal to their matched case, trial component and DNA extraction method (5-Prime, phenol, Bioserve, or PurGene). This case–control study was nested within the WHI, which includes 161,808 postmenopausal women recruited from 1993 to 1998 across the United States (26).

The study was approved by the Cancer Council Victoria's Human Research Ethics Committee, Melbourne, VIC, Australia, and the Institutional Review Board and Publications and Presentations Committee of WHI–Clinical Coordinating Center in the Fred Hutchinson Cancer Research Center, Seattle, Washington. All participants provided informed consent in accordance with the Declaration of Helsinki.

### Quality control and normalization of methylation data

Quality control (QC) details for measures of genome-wide DNA methylation in the MCCS have been reported previously (22). Briefly, we removed probes with missing rate > 20% and probes on Y-chromosome, and ultimately retained 484,966 CpG sites with their beta values for each sample. Methylation M-values, calculated as $\log_2$[beta/(1-beta)], were used for analysis as these are thought to be more statistically valid for detection of differential methylation (27). In the replication data of WHI, similar data processing on DNA methylation was performed, e.g., QC on CpGs sites using probe missing rate (> 10%) and beadcount (<3) in at least 10% of samples, and M-value transformation, as described previously (23, 24).

### Association analysis of genome-wide DNA methylation

An epigenome-wide association study (EWAS) based on the 404 case–control pairs in MCCS was conducted, using conditional logistic regression to estimate OR and 95% CI of UCC risk per SD at each of the 484,966 CpG sites. A first model (model 1) was adjusted for white blood cell composition (percentage of $CD4^+$ T cells, $CD8^+$ T cells, B cells, NK cells, monocytes, and granulocytes, estimated using the Houseman algorithm; ref. 28), and a second model (model 2) was additionally adjusted for smoking status (current/former/never) and pack-years (log-transformed). As a sensitivity analysis, we evaluated a third model (model 3) with additional adjustment for alcohol consumed in the previous week (in grams/day), body mass index (in $kg/m^2$), height (in meters), educational level (pseudo-continuous score ranging from 1 for "primary school only" to 8 for "tertiary or higher university degree"), physical activity (categorized score based on time spent doing vigorous/less vigorous activities), socioeconomic status (deciles of the relative socioeconomic disadvantage of area of residence index), and diet quality (Alternative Healthy Eating Index 2010). We also stratified analyses by sex and clinical subtype (muscle invasive or non–muscle invasive) and tested heterogeneity of the associations using the likelihood ratio test, by comparing models with and without interaction terms for these variables. The Bonferroni correction was applied to account for multiple comparisons ($P < 0.05/484{,}966 = 1.03 \times 10^{-7}$).

### Association analysis of smoking-associated DNA methylation

Among the 484,966 probes, we focused on 1,061 sites that were found to be strongly associated with a comprehensive smoking index in the MCCS ($P < 10^{-7}$) and also reported to be associated with smoking at this threshold $P < 10^{-7}$ in any of six large studies, as described in our previous publication (see Supplementary Table S1; ref. 29). For the replication study, we also used conditional logistic regression (models 1 and 2) to estimate associations of the 1,061 smoking-associated DNA methylation measures with risk of UCC in the WHI. For the WHI study, models 1 and 2 were additionally adjusted for race/ethnicity (Asian/Pacific Islander, Black/African American, Hispanic/Latino, non-Hispanic white, or other). The Bonferroni correction was applied to account for multiple comparisons ($P < 0.05/1{,}061 = 4.7 \times 10^{-5}$).

### Meta-analysis of MCCS and WHI studies

A fixed-effects meta-analysis with inverse-variance weights was conducted to combine associations with UCC risk at the 1,061 smoking-associated CpGs from the analyses of MCCS and WHI, using the *metagen* function in the R package *meta* (30). The I-square statistic was used to assess heterogeneity across the two studies.

### Predictive models

A predictor of UCC risk was developed using the data of 270 case–control pairs from the MCCS cohort for which blood was collected at baseline (1990–1994) as the training set (discovery phase), and 134 case–control pairs for which blood was drawn at follow-up (2003–2007) as an independent testing set in the testing phase. We used penalized logistic regression models with UCC risk as the outcome and the M-values at the 1,061 smoking-associated CpGs as the independent variables, applied to the training set using the R package *glmnet* (31). Five-fold cross-validation was used, and the mixing parameter (alpha) was set to 1 to apply a Lasso (least absolute shrinkage and selection operator) penalty. The covariates used in model 3 were forced in the penalized logistic models. Coefficients of the logistic Lasso model with the lambda value corresponding to the minimum mean cross-validated error were extracted and used as weights of the selected CpGs to construct a smoking methylation score (MS) for each participant. The smoking MS was then evaluated as a predictor of UCC risk in conditional logistic regression models (adjusted for covariates in model 3 for MCCS data and in model 2 for WHI data, respectively) in the testing sets.

Alternative ways to build methylation-based predictors of UCC risk were explored. We conducted univariate analyses using conditional logistic regression models to the training set to estimate ORs for the individual associations between DNA methylation and UCC risk at each of the 1,061 CpGs. The same covariates as those forced in the Lasso models were included as covariates. We considered three *P*-value cutoffs (0.05, 0.01, and 0.001) of individual associations at the 1,061 sites, and for each of them we calculated a smoking MS as a weighted average using as weights the logarithm of the OR for each selected CpG.

As a sensitivity analysis, we also used the logistic Lasso method (as described above) to develop a DNA methylation-based smoking predictor of UCC risk using all 404 MCCS case–control pairs. The external 440 case–control pairs from the WHI study were then used as an independent testing set to assess the proposed DNA methylation-based smoking predictors by using conditional logistic regression models (adjusted for covariates in model 2).

The accuracy of the predictive models with the smoking MS as UCC risk predictor was assessed using area under the receiver operating characteristic curve (AUC) estimates with unconditional logistic regression models (models A, B, and C), using the R package pROC (32). Model A used white blood cell composition as independent variables. Model B used white blood cell composition, smoking status, and pack-years (log-transformed) as independent variables. Race/ethnicity was also included in the two models for the WHI sample. Model C used white blood cell composition, smoking status, pack-years, and other covariates (age, sex, country of birth, sample type, alcohol, BMI, height, educational level, physical activity, socioeconomic status, and diet quality) as independent variables. The proposed MSs were then used as additional independent variables in the models to assess the prediction performance by AUC. The DeLong test (33) was used for comparing AUCs.

All methylation scores were rescaled to Z-scores for better comparability of their association with UCC risk. The flowchart of the statistical analysis pipelines and method details are shown in **Fig. 1**.

## Results

The distribution of sociodemographic, lifestyle, anthropometric, and clinical characteristics of the participants in the MCCS is presented

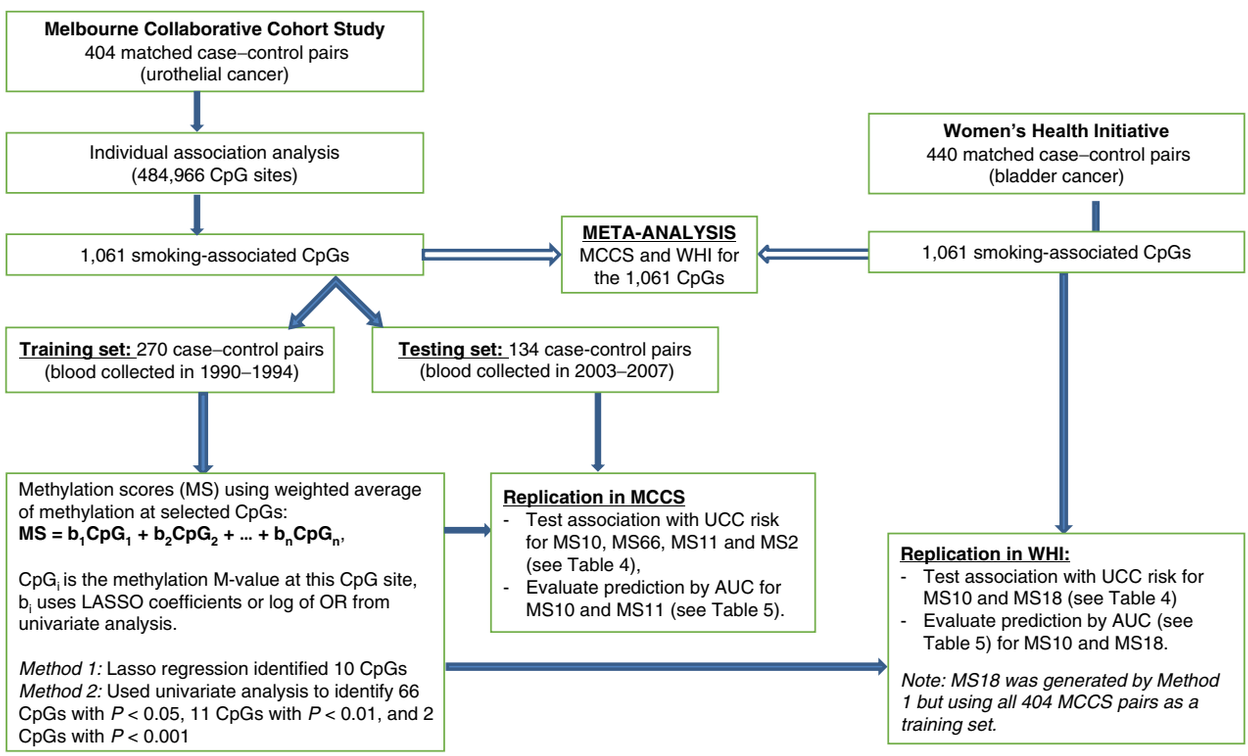


**Figure 1.**

Flowchart of the study. Description of the data and methods used for the analysis

in **Table 1**. Controls were matched to cases on age at blood draw, sex, country of birth (Australia/New Zealand/UK/other, Italy, or Greece) and sample type (peripheral blood mononuclear cells, dried blood spots, and buffy coats). The participants in the MCCS testing set were an average eight years older than in the training set. Compared with controls, cases were more frequently past and current smokers, and had greater smoking pack-years.

For the genome-wide probes tested on the 404 MCCS case–control pairs using models 1–3, there was no significant association between DNA methylation and risk of UCC after Bonferroni correction ($P < 1.03 \times 10^{-7}$). Nominally significant associations ($P < 0.05$) were observed for 40,664 (∼8%), 32,137 (∼7%), and 31,319 (∼6%) of the 484,966 CpGs using models 1–3, respectively.

Focusing on the 1,061 smoking-associated CpG sites that we previously identified (29), there was no significant association between DNA methylation and UCC risk in the MCCS after Bonferroni correction ($P < 4.7 \times 10^{-5}$). Comparing to genome-wide results, there were more methylation markers associated with risk of UCC for the smoking-associated loci, e.g., 19 of the 25 CpGs most strongly with smoking had $P < 0.05$ in model 1 (Supplementary Table S1). Nominally significant associations ($P < 0.05$) were observed for 206 (∼19%) and 93 (∼9%) of the 1,061 CpGs in models 1 and 2, respectively (Supplementary Table S1), and the direction of the association was the same as for smoking for 205/206 (100%) and 88/93 (95%) CpG sites. Adjustment for a more comprehensive set of variables (model 3) did not substantially change the associations (**Table 2** and Supplementary Fig. S1). Furthermore, the direction of association at 883 (83%, 662 negative and 221 positive, model 1) and 766 (72%, 586 negative and 180 positive, model 2) of the 1,061 CpGs was the same as for their association with smoking (Supplementary Table S1). The results for the 20 most significant associations are presented in **Table 2**; for all of these associations, the direction of association was the same as with smoking. The stratified results by UCC subtype and sex are shown in Supplementary Tables S2 and S3; we observed no evidence of significant UCC subtype or sex heterogeneity after Bonferroni correction for multiple testing ($P < 4.7 \times 10^{-5}$).

The replication study using WHI data identified nominally significant associations ($P < 0.05$) for 229 (∼22%) and 47 (∼4%) of the 1,061 smoking-based CpGs in models 1 and 2, respectively (Supplementary Tables S4 and S5). Among these associations, 51 CpGs (model 1) and 3 CpGs (model 2) were also nominally significant and in the same direction as in the MCCS data.

The meta-analysis of MCCS and WHI results identified nominally significant associations for 387 (∼36%) and 86 (∼8%) CpG sites in models 1 and 2, respectively (Supplementary Tables S4 and S5), and the direction of the association was the same as the association with smoking for 387/387 (100%) and 79/86 (92%) of the CpGs. There were 29 significant associations in model 1 after Bonferroni correction ($P < 4.7 \times 10^{-5}$), and among these associations, 9 CpGs overlapping the *AHRR, GPR15, F2RL3, PRSS23*, and *GFI1* genes were genome-wide significant ($P < 1.03 \times 10^{-7}$). The associations were nevertheless substantially attenuated (all $P > 4.7 \times 10^{-5}$) after adjusting for self-reported smoking variables (model 2). For the majority of the 1,061 CpGs, there was little heterogeneity between MCCS and WHI results (81% and 83% of the CpGs had $I^2 < 0.5$ in models 1 and 2, respectively; see Supplementary Tables S4 and S5). The 20 strongest associations in the meta-analyses of models 1 and 2 are shown in **Table 3**.

**Table 1.** Characteristics of the MCCS participants included in the analyses.

| | Training set (1990–1994) | | Testing set (2003–2007) | |
|---|---|---|---|---|
| **Participant characteristics** | **UCC cases (*N* = 270)** | **Controls (*N* = 270)** | **UCC cases (*N* = 134)** | **Controls (*N* = 134)** |
| Age at blood draw, median [IQR] | 63 [58–67] | 64 [58–67] | 72 [67–77] | 72 [67–77] |
| Sex | | | | |
| Male, *N* (%) | 207 (77%) | 207 (77%) | 101 (75%) | 101 (75%) |
| Female, *N* (%) | 63 (23%) | 63 (23%) | 33 (25%) | 33 (25%) |
| Country of birth | | | | |
| Australia/NZ/UK/other, *N* (%) | 168 (62%) | 166 (61%) | 104 (78%) | 104 (78%) |
| Italy, *N* (%) | 56 (21%) | 58 (21%) | 20 (15%) | 20 (15%) |
| Greece, *N* (%) | 46 (17%) | 46 (17%) | 10 (7%) | 10 (7%) |
| Blood sample type | | | | |
| Dried blood spots, *N* (%) | 170 (63%) | 170 (63%) | 1 (1%) | 1 (1%) |
| Peripheral blood mononuclear cells, *N* (%) | 93 (34%) | 93 (34%) | 0 (0%) | 0 (0%) |
| Buffy coats, *N* (%) | 7 (3%) | 7 (3%) | 133 (99%) | 133 (99%) |
| Smoking | | | | |
| Current, *N* (%) | 51 (19%) | 41 (15%) | 22 (16%) | 13 (10%) |
| Former, *N* (%) | 146 (54%) | 111 (41%) | 68 (51%) | 63 (47%) |
| Never, *N* (%) | 73 (27%) | 118 (44%) | 44 (33%) | 58 (43%) |
| Smoking pack-years, median [IQR] | 18 [0–40.7] | 4.2 [0–29.6] | 11.4 [0–35.1] | 5.2 [0–19.8] |
| Height (cm), median [IQR] | 168 [162–173] | 168 [163–173] | 169 [162–176] | 170 [164–175] |
| Body mass index (kg/m$^2$), median [IQR] | 27.5 [25.4–29.8] | 27.1 [24.8–29.5] | 27.3 [24.7–29.8] | 27.2 [24.5–29.5] |
| Alcohol (ethanol) consumption (g/day), median [IQR] | 4.5 [0–20.5] | 6.8 [0–17.7] | 9.2 [1.3–23.6] | 8.7 [0.6–23.4] |
| Diet quality: AHEI-2010, median [IQR] | 63.0 [55.0–70.9] | 64.5 [57.0–72.0] | 64.5 [55.0–70.5] | 63.0 [57.5–72.4] |
| Physical activity score, median [IQR] | 2 [1.3–2] | 2 [2–2] | 2 [2–3] | 2 [2–2.8] |
| Education score, median [IQR] | 4 [3–5] | 4 [3–6] | 4 [4–7] | 4 [4–8] |
| Socioeconomic status, SEIFA-10, median [IQR] | 5 [3–8] | 5 [3–8] | 6 [4–9] | 6 [3–9] |

Note: Physical activity score is a categorized score based on time spent doing vigorous/less vigorous activities. Educational score is a pseudo-continuous score ranging from 1 for "primary school only" to 8 for "tertiary or higher university degree."

**Table 2.** EWAS results of UCC risk in the 20 most significant associations (model 1) of the 1,061 smoking-associated CpGs based on the 404 matched case–control pairs in the MCCS.

| | | | | Association with smoking[29] | | Association with UCC risk (Model 1) | | Association with UCC risk (Model 2) | | Association with UCC risk (Model 3) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CpG | Chr. | Position | Gene | Effect | *P* | OR (95% CI) | *P* | OR (95% CI) | *P* | OR (95% CI) | *P* |
| cg21566642 | 2 | 233284661 | | −0.27 | <5E−308 | 0.72 (0.61–0.84) | 5.68E−05 | 0.80 (0.66–0.96) | 1.99E−02 | 0.80 (0.66–0.98) | 2.89E−02 |
| cg19089201 | 7 | 45002287 | *MYO1G* | 0.08 | 1.22E−21 | 1.39 (1.18–1.64) | 6.13E−05 | 1.35 (1.15–1.60) | 3.21E−04 | 1.36 (1.15–1.60) | 3.70E−04 |
| cg12803068 | 7 | 45002919 | *MYO1G* | 0.20 | 2.07E−71 | 1.37 (1.18–1.61) | 6.91E−05 | 1.31 (1.11–1.54) | 1.32E−03 | 1.31 (1.11–1.55) | 1.28E−03 |
| cg17924476 | 5 | 323794 | *AHRR* | 0.07 | 3.52E−29 | 1.36 (1.17–1.60) | 1.14E−04 | 1.31 (1.11–1.54) | 1.24E−03 | 1.31 (1.11–1.54) | 1.62E−03 |
| cg05575921 | 5 | 373378 | *AHRR* | −0.39 | <5E−308 | 0.74 (0.63–0.86) | 1.16E−04 | 0.78 (0.63–0.97) | 2.26E−02 | 0.79 (0.63–0.98) | 3.25E−02 |
| cg10399789 | 1 | 92945668 | *GFI1* | −0.06 | 2.42E−16 | 0.70 (0.59–0.84) | 1.41E−04 | 0.69 (0.57–0.84) | 1.24E−04 | 0.69 (0.57–0.83) | 1.43E−04 |
| cg12876356 | 1 | 92946825 | *GFI1* | −0.13 | 1.07E−66 | 0.72 (0.61–0.85) | 1.49E−04 | 0.75 (0.63–0.90) | 1.48E−03 | 0.75 (0.63–0.90) | 1.75E−03 |
| cg27457191 | 7 | 77429766 | *PHTF2* | −0.03 | 2.05E−08 | 0.57 (0.42–0.76) | 1.51E−04 | 0.59 (0.44–0.80) | 6.77E−04 | 0.58 (0.43–0.79) | 4.91E−04 |
| cg09935388 | 1 | 92947588 | *GFI1* | −0.19 | 1.94E−119 | 0.72 (0.61–0.86) | 1.93E−04 | 0.78 (0.64–0.93) | 6.68E−03 | 0.79 (0.66–0.96) | 1.49E−02 |
| cg01940273 | 2 | 233284934 | | −0.19 | 1.69E−304 | 0.75 (0.64–0.87) | 2.77E−04 | 0.82 (0.68–0.99) | 3.43E−02 | 0.82 (0.68–0.98) | 3.28E−02 |
| cg05951221 | 2 | 233284402 | | −0.21 | <5E−308 | 0.75 (0.65–0.88) | 2.95E−04 | 0.85 (0.71–1.01) | 6.66E−02 | 0.86 (0.72–1.03) | 1.10E−01 |
| cg08884752 | 1 | 2162001 | *SKI* | −0.04 | 6.97E−14 | 0.67 (0.54–0.84) | 4.56E−04 | 0.70 (0.56–0.88) | 2.16E−03 | 0.69 (0.55–0.87) | 1.84E−03 |
| cg19859270 | 3 | 98251294 | *GPR15* | −0.12 | 1.71E−104 | 0.75 (0.63–0.88) | 4.74E−04 | 0.81 (0.68–0.97) | 2.10E−02 | 0.81 (0.67–0.97) | 1.99E−02 |
| cg06126421 | 6 | 30720080 | | −0.24 | 2.10E−259 | 0.73 (0.61–0.88) | 6.47E−04 | 0.82 (0.67–1.00) | 5.14E−02 | 0.81 (0.66–1.00) | 4.71E−02 |
| cg23576855 | 5 | 373299 | *AHRR* | −0.33 | 4.63E−96 | 0.76 (0.65–0.89) | 6.73E−04 | 0.80 (0.68–0.94) | 5.54E−03 | 0.79 (0.67–0.93) | 4.72E−03 |
| cg16151960 | 5 | 133890280 | *PHF15* | −0.02 | 4.45E−12 | 0.70 (0.57–0.86) | 6.80E−04 | 0.74 (0.60–0.91) | 4.53E−03 | 0.73 (0.59–0.91) | 4.99E−03 |
| cg09662411 | 1 | 92946132 | *GFI1* | −0.06 | 8.65E−33 | 0.72 (0.60–0.87) | 7.86E−04 | 0.76 (0.63–0.93) | 7.98E−03 | 0.76 (0.63–0.93) | 8.88E−03 |
| cg03636183 | 19 | 17000585 | *F2RL3* | −0.21 | <5E−308 | 0.75 (0.64–0.89) | 8.61E−04 | 0.84 (0.69–1.02) | 7.09E−02 | 0.83 (0.68–1.02) | 7.68E−02 |
| cg03707168 | 19 | 49379127 | *PPP1R15A* | −0.09 | 1.96E−48 | 0.68 (0.54–0.86) | 1.00E−03 | 0.74 (0.58–0.94) | 1.40E−02 | 0.72 (0.57–0.93) | 1.03E−02 |
| cg04011474 | 2 | 28904455 | | −0.05 | 3.79E−17 | 0.69 (0.55–0.86) | 1.02E−03 | 0.71 (0.56–0.89) | 2.95E−03 | 0.70 (0.56–0.89) | 3.19E−03 |

Note: Association of methylation with smoking was estimated by linear mixed-effects regression on a comprehensive smoking index (29) with a parameter tau = 1.5. Association of methylation with UCC risk was estimated by conditional logistic regression model. Model 1 was adjusted for white blood cell composition. Model 2 was adjusted for white blood cell composition, smoking status, and pack-years. Model 3 was adjusted for white blood cell composition, smoking status, pack-years, and other covariates (alcohol, BMI, height, educational level, physical activity, socioeconomic status, and diet quality).

**Table 3.** Meta-analysis results for MCCS and WHI in the 20 most significant associations between DNA methylation at 1,061 smoking-associated CpGs and risk of UCC by models 1 and 2, respectively.

| Meta-analysis (Model 1) | | | | | | Meta-analysis (Model 2) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **CpG** | **Chr.** | **Position** | **Gene** | **OR (95% CI)** | ***P*** | **CpG** | **Chr** | **Position** | **Gene** | **OR (95% CI)** | ***P*** |
| cg21566642 | 2 | 233284661 | | 0.67 (0.60–0.75) | 2.35E−12 | cg26203136 | 7 | 739057 | *PRKAR1B* | 0.81 (0.72–0.93) | 1.67E−03 |
| cg05575921 | 5 | 373378 | *AHRR* | 0.64 (0.56–0.72) | 2.59E−12 | cg05575921 | 5 | 373378 | *AHRR* | 0.76 (0.63–0.91) | 2.72E−03 |
| cg05951221 | 2 | 233284402 | | 0.69 (0.62–0.77) | 4.72E−11 | cg23110422 | 21 | 40182073 | *ETS2* | 0.84 (0.74–0.94) | 2.85E−03 |
| cg06126421 | 6 | 30720080 | | 0.68 (0.61–0.77) | 2.88E−10 | cg17924476 | 5 | 323794 | *AHRR* | 1.19 (1.06–1.33) | 3.31E−03 |
| cg01940273 | 2 | 233284934 | | 0.71 (0.64–0.79) | 1.15E−09 | cg04332373 | 4 | 15779642 | *CD38* | 1.28 (1.08–1.51) | 3.74E−03 |
| cg19859270 | 3 | 98251294 | *GPR15* | 0.71 (0.64–0.80) | 2.65E−09 | cg19089201 | 7 | 45002287 | *MYO1G* | 1.19 (1.06–1.34) | 3.79E−03 |
| cg03636183 | 19 | 17000585 | *F2RL3* | 0.69 (0.61–0.78) | 4.88E−09 | cg11660018 | 11 | 86510915 | *PRSS23* | 0.80 (0.68–0.93) | 4.15E−03 |
| cg11660018 | 11 | 86510915 | *PRSS23* | 0.68 (0.59–0.77) | 1.37E−08 | cg07123182 | 11 | 2722391 | *KCNQ1OT1* | 0.84 (0.75–0.95) | 4.81E−03 |
| cg09935388 | 1 | 92947588 | *GFI1* | 0.73 (0.65–0.82) | 5.72E−08 | cg15013801 | 10 | 73976790 | *ASCC1* | 0.82 (0.71–0.94) | 5.64E−03 |
| cg19798735 | 7 | 110730805 | *IMMP2L* | 0.64 (0.54–0.75) | 1.22E−07 | cg25560398 | 2 | 233252170 | *ECEL1P2* | 0.84 (0.74–0.95) | 6.15E−03 |
| cg17924476 | 5 | 323794 | *AHRR* | 1.31 (1.18–1.45) | 1.37E−07 | cg19798735 | 7 | 110730805 | *IMMP2L* | 0.77 (0.64–0.93) | 7.29E−03 |
| cg06644428 | 2 | 233284112 | | 0.73 (0.64–0.82) | 4.48E−07 | cg10399789 | 1 | 92945668 | *GFI1* | 0.83 (0.72–0.95) | 7.59E−03 |
| cg25560398 | 2 | 233252170 | *ECEL1P2* | 0.75 (0.66–0.84) | 7.67E−07 | cg26337070 | 2 | 85999873 | *ATOH8* | 0.80 (0.68–0.94) | 8.15E−03 |
| cg12803068 | 7 | 45002919 | *MYO1G* | 1.31 (1.18–1.45) | 7.90E−07 | cg04086928 | 9 | 134612644 | *RAPGEF1* | 0.80 (0.68–0.94) | 8.62E−03 |
| cg23110422 | 21 | 40182073 | *ETS2* | 0.78 (0.70–0.86) | 1.36E−06 | cg21566642 | 2 | 233284661 | | 0.82 (0.71–0.95) | 9.07E−03 |
| cg12876356 | 1 | 92946825 | *GFI1* | 0.77 (0.69–0.86) | 2.09E−06 | cg05677062 | 12 | 123874707 | *SETD8* | 0.82 (0.70–0.95) | 9.13E−03 |
| cg03991871 | 5 | 368447 | *AHRR* | 0.77 (0.69–0.86) | 2.30E−06 | cg09935388 | 1 | 92947588 | *GFI1* | 0.84 (0.73–0.96) | 1.01E−02 |
| cg03707168 | 19 | 49379127 | *PPP1R15A* | 0.66 (0.55–0.79) | 6.48E−06 | cg22052143 | 5 | 78067856 | | 0.83 (0.72–0.96) | 1.05E−02 |
| cg12806681 | 5 | 368394 | *AHRR* | 0.78 (0.70–0.87) | 7.18E−06 | cg26361535 | 8 | 144576604 | *ZC3H3* | 0.86 (0.77–0.97) | 1.16E−02 |
| cg25189904 | 1 | 68299493 | *GNG12* | 0.78 (0.70–0.87) | 8.80E−06 | cg26529655 | 5 | 424371 | *AHRR* | 0.77 (0.63–0.94) | 1.17E−02 |

Note: Association of methylation with UCC risk was estimated by conditional logistic regression model. Model 1 was adjusted for white blood cell composition. Model 2 was adjusted for white blood cell composition, smoking status, and pack-years.

**Table 4.** OR (per 1 SD increase), 95% CI, and *P* value for the association between methylation-based predictors and risk of UCC.

| | Replication data sets | | | |
|---|---|---|---|---|
| | MCCS (*N* = 134 pairs) | | WHI (*N* = 440 pairs) | |
| Predictor | OR (95% CI) | *P* | OR (95% CI) | *P* |
| MS10 | 1.37 (1.00–1.90) | 0.05 | 1.09 (0.91–1.30) | 0.37 |
| MS66 | 1.35 (0.95–1.91) | 0.09 | | |
| MS11 | 1.42 (1.01–1.99) | 0.04 | | |
| MS2 | 1.05 (0.78–1.40) | 0.76 | | |
| MS18 | | | 1.09 (0.92–1.30) | 0.33 |

Note: The predictor was built by weighted average on methylation at selected CpGs: $MS = b_1CpG_1 + b_2CpG_2 + \ldots + b_nCpG_n$, where $CpG_i$ is M-value at this CpG site, $b_i$ use Lasso coefficients (for MS10, MS18) or log of OR from univariate analyses (for MS66, MS11, and MS2). The association was estimated by conditional logistic regression model 3 for MCCS data and model 2 for WHI data, respectively.

The logistic Lasso regression of UCC risk on the 1,061 smoking-based CpGs using the 270 MCCS baseline case–control pairs selected ten CpGs (MS10): cg01324550 (*LOC404266*), cg02743070 (*ZMIZ1*), cg07058086 (*KIF13B*), cg10399789 (*GFI1*), cg16622061 (chr16: 86888736), cg17924476 (*AHRR*), cg18979623 (*ZBTB46*), cg19089201 (*MYO1G*), cg23110422 (*ETS2*), and cg24139443 (chr17: 74131549; Supplementary Table S6). The associations with risk of UCC for the 1,061 smoking-associated methylation sites on the training data are shown in Supplementary Table S6. The derived methylation scores based on associations at $P < 0.05$, $P < 0.01$, and $P < 0.001$ included 66 (MS66), 11 (MS11), and 2 (MS2) CpGs, respectively. The associations of these four predictors with UCC risk in the MCCS testing data set ($N = 134$ cases, model 3) are presented in **Table 4**. MS10 and MS11 had five overlapping CpGs (cg07058086, cg10399789, cg17924476, cg19089201, and cg23110422) and were associated with risk of UCC in the testing data set (OR = 1.37; 95% CI, 1.00–1.90) and (OR = 1.42; 95% CI, 1.01–1.99), respectively. The distribution of MS10 by smoking status is presented in Supplementary Fig. S2, showing it was elevated in current compared with never smokers. The association of MS10 with UCC risk in the WHI data (model 2) was weaker (OR = 1.09; 95% CI, 0.91–1.30).

Using all 404 case–control pairs of MCCS as the training set, as a sensitivity analysis, the logistic Lasso models selected 18 CpGs (MS18) from the 1,061 smoking-associated CpGs (Supplementary Table S7). MS18 and MS10 had eight overlapping CpGs (cg02743070, cg07058086, cg10399789, cg16622061, cg17924476, cg19089201, cg23110422, and cg24139443). We assessed the resulting predictor MS18 by examining its association with UCC risk in the WHI data, and the result was very similar as for MS10 (OR = 1.09; 95% CI, 0.92–1.30; **Table 4**). The fixed-effects meta-analysis for MS10 of the two replication sets in MCCS ($N = 134$) and WHI ($N = 440$) gave an estimated OR of 1.15; 95% CI, 0.98–1.34, $P = 0.08$.

The ability of the methylation scores to predict risk of UCC with different models on the testing data sets is presented in **Table 5**. For the MCCS testing set, the predictions by model C + MS10 and model C + MS11 achieved the highest AUC estimate of 0.66, which was only slightly greater than the same model without methylation information (AUC = 0.64, $P = 0.43$ for MS10 and 0.39 for MS11). For the WHI testing set, the prediction by model B + MS10 or MS18 achieved an AUC estimate of 0.68, which was of the same as model B alone ($P = 0.11$ or 0.22).

## Discussion

Most previous studies that investigated the association of smoking with development of urothelial cancer used self-reported smoking history. We included two self-reported variables, smoking status and pack-years, in our analyses. There are other aspects of smoking history, such as age at starting or passive smoking that are typically not or inaccurately captured by questionnaires. As DNA methylation in blood can capture lifetime exposure or different individual responses to smoking, we evaluated the association between smoking-associated methylation and risk of UCC. Although potential associations with UCC were identified at 206 (∼19%) and 93 (∼9%) smoking-based

**Table 5.** AUC estimates and comparisons for predictions of UCC risk on the testing data sets using several models.

| MCCS (*N* = 134 pairs) | | | WHI (*N* = 440 pairs) | | |
|---|---|---|---|---|---|
| | AUC | *P* | | AUC | *P* |
| Model A | 0.61 | 0.18 (vs. model C) | Model A | 0.58 | 0.0002 (vs. model B) |
| Model B | 0.63 | 0.52 (vs. model C) | Model B | 0.68 | |
| Model C | 0.64 | | | | |
| Model A + MS10 | 0.63 | 0.27 (vs. model A) | Model A + MS10 | 0.61 | 0.05 (vs. model A) |
| Model A + MS11 | 0.64 | 0.19 (vs. model A) | Model A + MS18 | 0.61 | 0.07 (vs. model A) |
| Model B + MS10 | 0.65 | 0.36 (vs. model B) | Model B + MS10 | 0.68 | 0.11 (vs. model B) |
| Model B + MS11 | 0.65 | 0.30 (vs. model B) | Model B + MS18 | 0.68 | 0.22 (vs. model B) |
| Model C + MS10 | 0.66 | 0.44 (vs. model C) | | | |
| Model C + MS11 | 0.66 | 0.45 (vs. model C) | | | |

Note: The AUC was estimated based on unconditional logistic regression models. Model A used white blood cell composition as independent variables (for WHI, race/ethnicity was also used). Model B used white blood cell composition, smoking status, and pack-years as independent variables (for WHI, race/ethnicity was also used). Model C used white blood cell composition, smoking status, pack-years, and other covariates (age, sex, country of birth, sample type, alcohol, BMI, height, educational level, physical activity, socioeconomic status, and diet quality) as independent variables. MS10, MS11, and MS18 were additional independent variables. *P* value was obtained by the DeLong test versus other models.

CpG sites in the MCCS in models without and with adjustment for self-reported smoking, respectively, and most associations were in the expected direction, these associations were overall quite weak. In the meta-analysis, DNA methylation at genes, including *AHRR, GPR15, F2RL3, PRSS23*, and *GFI1* (major smoking-related genes), was strongly ($P < 10^{-7}$) associated with UCC risk; however, the associations were substantially attenuated after adjusting for self-reported smoking history, likely because these self-reported variables might have captured almost full information of smoking exposure. Thus, these methylation markers added relatively little to the prediction of urothelial cancer risk beyond their association with self-reported smoking. A methylation score combining measures at ten smoking-associated CpG sites developed in the MCCS cohort showed some evidence of association with risk of UCC (OR per SD $\sim$ 1.4) independently of self-reported smoking in an independent data set of MCCS participants (**Table 4**). Although these results suggest that the combination of smoking methylation markers may improve the prediction of urothelial cancer risk, limited evidence of replication was found in the WHI cohort (OR per SD $\sim$ 1.1).

The previous study by Jordahl and colleagues (24) using WHI data investigated three specific smoking-related probes (cg05575921 in the gene *AHRR*, cg03636183 in *F2RL3*, and cg19859270 in *GPR15*) in relation to risk of UCC and showed that methylation alterations at cg05575921 and cg19859270 might mediate the effects of smoking on UCC. Our MCCS data also detected nominally significant associations with UCC risk at these CpGs (cg05575921: OR = 0.78; 95% CI, 0.63–0.97; $P = 0.02$ and cg19859270: OR = 0.81; 95% CI, 0.68–0.97; $P = 0.02$) in the adjusted model, which indicate they may add information about risk, in addition to the potential mediation of effect.

DNA methylation at *AHRR* cg05575921 was previously reported to be strongly associated with lung cancer risk (19, 34–36), e.g., OR = 0.50 (95% CI, 0.43–0.59), $P = 4.3 \times 10^{-17}$ in a pooled analysis of five case–control studies (19). Six CpGs in the *AHRR* gene also showed nominally significant association ($P < 0.05$) with risk of UCC in our meta-analysis (model 2): cg05575921 (OR = 0.76, $P = 0.003$), cg17924476 (OR = 1.19, $P = 0.003$), cg26529655 (OR = 0.77, $P = 0.01$), cg12806681 (OR = 0.86, $P = 0.02$), cg01899089 (OR = 0.88, $P = 0.03$), and cg03991871 (OR = 0.88, $P = 0.04$; see Supplementary Table S5). Moreover, cg03636183 in the *F2RL3* gene, cg21566642 and cg05951221 in 2q37.1, and cg06126421 in 6p21.33 were also reported to be strongly associated ($P = 2 \times 10^{-15}$) with lung cancer risk (19). Among them, three CpGs also showed nominally significant association with UCC risk in our meta-analyses (model 2): cg21566642 (OR = 0.82, $P = 0.009$), cg05951221 (OR = 0.86, $P = 0.04$), and cg06126421 (OR = 0.85, $P = 0.03$; see Supplementary Table S5). These associations appeared to be weaker than in the lung cancer studies, likely because smoking is not as strong a risk factor for urothelial cancer as it is for lung cancer. In a recent study (37), we showed that *GrimAge*, a composite biomarker based on several DNA methylation surrogates for plasma proteins and a methylation-based estimator of smoking pack-years (38), is substantially more strongly associated with lung cancer risk (OR per SD = 2.03; 95% CI, 1.56–2.64) than with risk of UCC (OR = 1.22; 95% CI, 0.98–1.52).

The samples used in the WHI cohort were all postmenopausal women, and smoking accounts for approximately half of bladder cancer incidence among postmenopausal women (4, 23). Sex is associated with distinct DNA methylation patterns (39). However, we did not find that associations of DNA methylation smoking markers with UCC varied by sex in the MCCS data, nor did we find heterogeneity between MCCS and WHI results. In this study, we used two common methods to develop risk predictors: (i) Lasso and (ii) univariate analysis with weighted average based on individual CpG associations with UCC risk. For the latter, it is difficult to decide on an appropriate *P*-value cutoff, and our results showed that the Lasso performed well in this setting. Although there was a reasonably large association of the Lasso predictor in the testing set (OR per 1 SD $\sim$ 1.4), this translated into only moderately improved risk prediction (**Table 5**).

DNA methylation changes strongly with age (40, 41). In a recent study using methylation case–control studies nested in the MCCS, we have identified and replicated 32,659 age-associated CpGs (42). Among the 1,061 smoking-associated CpGs considered in the current study, methylation at 475 (45%), 328 (31%), and 118 (11%) CpGs was found to be associated with age in never, former, and current smokers, respectively ($P < 0.05/1{,}061 = 4.7 \times 10^{-5}$, based on the data set used in (42), results not shown). Specifically, cg01324550, cg16622061, and cg24139443, which were included in MS10, showed significant associations with age in the overall sample and in never smokers, but not other CpGs (42). This implies that aging (or other cancer risk factors; refs. 43 and 44) may affect DNA methylation at the same loci, which may contribute to explain why these methylation marks add information about cancer risk, in addition to unmeasured smoking exposure.

There are several limitations in this study. First, even with pre-diagnostic blood samples, we cannot rule out the possibility that DNA methylation measures in blood reflected early cancer or development of other smoking-associated diseases. Second, the participants included in the MCCS testing set were an average eight years older than in the training set. We noted that model 1, which included only white blood cell composition variables, achieved an AUC of 0.53 for the training set but an AUC of 0.61 for the replication set (older MCCS participants). It may be that age, a strong cancer risk factor, is associated with changes in white blood cell composition over time (45) that are also associated with cancer risk (46, 47). Third, we considered the two MCCS data sets as independent because there was no participant overlap, and participants with follow-up blood samples were substantially older; however, the samples were drawn from the same cohort and might have a shared environment; thus, the two data sets might not be completely independent, which may have an influence on results of validation and risk prediction. Fourth, the modest improvement of AUC may suggest that other factors, such as germline genetic variation, and incorporation of more environmental exposures, should be considered in the predictive models. Fifth, the biological mechanisms underlying our findings were not assessed because the aim of our study was to improve UCC risk prediction using smoking-associated methylation marks. For example, TET proteins may stimulate and regulate DNA methylation at genes that were included (48), but this requires further investigation using functional studies. Finally, compared with the MCCS cohort, the methylation measures in WHI were produced using different methods of sample collection and storage, DNA extraction, and DNA methylation processing, which may have influenced some findings, e.g., high heterogeneity for some CpGs across the two studies when performing meta-analysis.

In conclusion, our findings suggest that blood-based DNA methylation markers for smoking may be associated, albeit weakly, with risk of UCC independent of self-reported smoking history, and could provide some improvement to the prediction of urothelial cancer risk. The overall utility of our findings needs to be further assessed using additional external data sets.

## Authors' Disclosures

K.M. Jordahl reports grants from American Cancer Society and NCI during the conduct of the study. J.K. Bassett reports grants from Australian National Health and Medical Research Council (NHMRC) during the conduct of the study. D.R. English reports grants from Australian National Health and Medical Research Council during the conduct of the study. R.L. Milne reports grants from NHMRC during the conduct of the study. G.G. Giles reports grants from NHMRC (Australia) during the conduct of the study. P. Dugué reports grants from NHMRC during the conduct of the study. No disclosures were reported by the other authors.

## Authors' Contributions

**C. Yu:** Conceptualization, formal analysis, methodology, writing–original draft. **K.-M. Jordahl:** Data curation, formal analysis, writing–review and editing. **J.K. Bassett:** Data curation, writing–review and editing. **J.E. Joo:** Data curation, writing–review and editing. **E. Wong:** Data curation, writing–review and editing. **M.T. Brinkman:** Writing–review and editing. **D.F. Schmidt:** writing–review and editing. **D.M. Bolton:** Writing–review and editing. **E. Makalic:** Writing–review and editing. **T.M. Brasky:** Writing–review and editing. **A.H. Shadyab:** Writing–review and editing. **L.F. Tinker:** Funding acquisition, writing–review and editing. **A. Longano:** Writing–review and editing. **J.L. Hopper:** Methodology, writing–review and editing. **D.R. English:** Methodology, writing–review and editing. **R.L. Milne:** Resources, methodology, writing–review and editing. **P. Bhatti:** Resources, funding acquisition, project administration, writing–review and editing. **M.C. Southey:** Resources, funding acquisition, project administration, writing–review and editing. **G.G. Giles:** Conceptualization, funding acquisition, project administration, writing–review and editing. **P. Dugué:** Conceptualization, formal analysis, supervision, funding acquisition, methodology, writing–original draft, and project administration.

## Acknowledgments

This work was supported by the Australian NHMRC project grants 209057, 251553, 504711, 1043616, and 1164455. Cohort recruitment was funded by Cancer Council Victoria (http://www.cancervic.org.au/) and VicHealth (https://www.vichealth.vic.gov.au/). M.C. Southey is a recipient of a Senior Research Fellowship from the NHMRC (GTN1155163). The WHI program was funded by the National Heart, Lung, and Blood Institute (NIH, U.S. Department of Health and Human Services) through grants 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, and 75N92021D00005.

The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database. The MCCS component of the work was funded by the Australian NHMRC, including grants 209057, 251553, 504711, 1043616 (to G.G. Giles) and 1164455 (to P.-A. Dugué). The WHI program was funded by the National Heart, Lung, and Blood Institute (NIH, U.S. Department of Health and Human Services) through grants 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, and 75N92021D00005 (to the Fred Hutchinson Cancer Research Center).

We would like to acknowledge the WHI investigators: *Program Office*: (National Heart, Lung, and Blood Institute, Bethesda, MD) Jacques Rossouw, Shari Ludlam, Joan McGowan, Leslie Ford, and Nancy Geller; *Clinical Coordinating Center*: (Fred Hutchinson Cancer Research Center, Seattle, WA) Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg; *Investigators and Academic Centers*: (Brigham and Women's Hospital, Harvard Medical School, Boston, MA) JoAnn E. Manson; (MedStar Health Research Institute/Howard University, Washington, DC) Barbara V. Howard; (Stanford Prevention Research Center, Stanford, CA) Marcia L. Stefanick; (The Ohio State University, Columbus, OH) Rebecca Jackson; (University of Arizona, Tucson/Phoenix, AZ) Cynthia A. Thomson; (University at Buffalo, Buffalo, NY) Jean Wactawski-Wende; (University of Florida, Gainesville/Jacksonville, FL) Marian Limacher; (University of Iowa, Iowa City/Davenport, IA) Jennifer Robinson; (University of Pittsburgh, Pittsburgh, PA) Lewis Kuller; (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker; (University of Nevada, Reno, NV) Robert Brunner; *Women's Health Initiative Memory Study*: (Wake Forest University School of Medicine, Winston-Salem, NC) Mark Espeland. For a list of all the investigators who have contributed to WHI science, please visit: https://s3-us-west-2.amazonaws.com/www-whi-org/wp-content/uploads/WHI-Investigator-Long-List.pdf

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 8, 2021; revised April 22, 2021; accepted September 10, 2021; published first September 14, 2021.

## References

1. Griffiths TL. Action on bladder cancer, current perspectives in bladder cancer management. Int J Clin Pract 2013;67:435–48.
2. Prasad SM, DeCastro GJ, Steinberg GD. Urothelial carcinoma of the bladder: definition, treatment and future efforts. Nature Reviews Urology 2011;8:631–42.
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.
4. Freedman ND, Silverman DT, Hollenbeck AR, Schatzkin A, Abnet CC. Association between smoking and risk of bladder cancer among men and women. JAMA 2011;306:737–45.
5. Burger M, Catto JW, Dalbagni G, Grossman HB, Herr H, Karakiewicz P, et al. Epidemiology and risk factors of urothelial bladder cancer. Eur Urol 2013;63:234–41.
6. Crivelli JJ, Xylinas E, Kluth LA, Rieken M, Rink M, Shariat SF. Effect of smoking on outcomes of urothelial carcinoma: a systematic review of the literature. Eur Urol 2014;65:742–54.
7. Hou L, Hong X, Dai M, Chen P, Zhao H, Wei Q, et al. Association of smoking status with prognosis in bladder cancer: a meta-analysis. Oncotarget 2017;8:1278.
8. Teleka S, Häggström C, Nagel G, Bjørge T, Manjer J, Ulmer H, et al. Risk of bladder cancer by disease severity in relation to metabolic factors and smoking: a prospective pooled cohort study of 800,000 men and women. Int J Cancer 2018;143:3071–82.
9. Mori K, Mostafaei H, Abufaraj M, Yang L, Egawa S, Shariat SF. Smoking and bladder cancer: review of the recent literature. Curr Opin Urol 2020;30:720–5.
10. van Osch FH, Jochems SH, van Schooten FJ, Bryan RT, Zeegers MP. Quantified relations between exposure to tobacco smoking and bladder cancer risk: a meta-analysis of 89 observational studies. Int J Epidemiol 2016;45:857–70.
11. Liber AC, Warner KE. Has underreporting of cigarette consumption changed over time? Estimates derived from US national health surveillance systems between 1965 and 2015. Am J Epidemiol 2018;187:113–9.
12. Thomas CE, Wang R, Adams-Haduch J, Murphy SE, Ueland PM, Midttun Ø, et al. Urinary cotinine is as good a biomarker as serum cotinine for cigarette smoking exposure and lung cancer risk prediction. Cancer Epidemiol Biomarkers Prev 2020;29:127–32.
13. Benowitz NL. Biomarkers of environmental tobacco smoke exposure. Environ Health Perspect 1999;107(suppl2):349–55.
14. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet 2011;88:450–7.
15. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, et al. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology 2013;24:712–6.
16. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet 2016;9:436–47.
17. Zhang Y, Florath I, Saum KU, Brenner H. Self-reported smoking, serum cotinine, and blood DNA methylation. Environ Res 2016;146:395–403.
18. Besingi W, Johansson Å. Smoke-related DNA methylation changes in the etiology of human disease. Hum Mol Genet 2014;23:2290–7.
19. Baglietto L, Ponzi E, Haycock P, Hodge A, Bianca Assumma M, Jung CH, et al. DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. Int J Cancer 2017;140:50–61.
20. Sabogal C, Su S, Tingen M, Kapuku G, Wang X. Cigarette smoking related DNA methylation in peripheral leukocytes and cardiovascular risk in young adults. Int J Cardiol 2020;306:203–5.

21. Marsit CJ, Koestler DC, Christensen BC, Karagas MR, Houseman EA, Kelsey KT. DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. J Clin Oncol 2011;29:1133–9.
22. Dugué PA, Brinkman MT, Milne RL, Wong EM, FitzGerald LM, Bassett JK, et al. Genome-wide measures of DNA methylation in peripheral blood and the risk of urothelial cell carcinoma: a prospective nested case–control study. Br J Cancer 2016;115:664–73.
23. Jordahl KM, Randolph TW, Song X, Sather CL, Tinker LF, Phipps AI, et al. Genome-wide DNA methylation in prediagnostic blood and bladder cancer risk in the women's health initiative. Cancer Epidemiol Biomarkers Prev 2018;27: 689–95.
24. Jordahl KM, Phipps AI, Randolph TW, Tindle HA, Liu S, Tinker LF, et al. Differential DNA methylation in blood as a mediator of the association between cigarette smoking and bladder cancer risk among postmenopausal women. Epigenetics 2019;14:1065–73.
25. Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK, et al. Cohort profile: the Melbourne collaborative cohort study (Health 2020). Int J Epidemiol 2017;46:1757–1757i.
26. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. The women's health initiative recruitment methods and results. Ann Epidemiol 2003; 13:S18-S77.
27. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 2010;11:587.
28. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;13:86.
29. Dugué PA, Jung CH, Joo JE, Wang X, Wong EM, Makalic E, et al. Smoking and blood DNA methylation: an epigenome-wide association study and assessment of reversibility. Epigenetics 2020;15:358–68.
30. Schwarzer G., meta: An R package for meta-analysis. R news 2007;7:40–45.
31. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Statist Softw 2010;33:1.
32. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.
34. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun 2015;6:1–9.
35. Zhang Y, Breitling LP, Balavarca Y, Holleczek B, Schöttker B, Brenner H. 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.
36. Bojesen SE, Timpson N, Relton C, Smith GD, Nordestgaard BG. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax 2017;72:646–53.
37. Dugué PA, Bassett JK, Wong EM, Joo JE, Li S, Yu C, et al. Biological aging measures based on blood DNA methylation and risk of cancer: a prospective study. JNCI Cancer Spectrum 2021;5:pkaa109.
38. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY) 2019;11: 303.
39. Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, et al. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. PLoS One 2009;4:e6767.
40. Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging. Aging Cell 2015;14:924–32.
41. Pal S, Tyler JK. Epigenetics and aging. Sci Adv 2016;2:e1600584.
42. Geurts YM, Dugué PA, Joo JE, Makalic E, Jung CH, Guan W, et al. Novel associations between blood DNA methylation and body mass index in middle-aged and older adults. Int J Obes 2018;42:887–96.
43. Dugué PA, Wilson R, Lehne B, Jayasekara H, Wang X, Jung CH, et al. Alcohol consumption is associated with widespread changes in blood DNA methylation: analysis of cross-sectional and longitudinal data. Addict Biol 2021; 26:e12855.
44. Yu C, Wong EM, Joo JE, Hodge AM, Makalic E, Schmidt D, et al. Epigenetic drift association with cancer risk and survival, and modification by sex. Cancers 2021; 13:1881.
45. Ruggiero C, Metter EJ, Cherubini A, Maggio M, Sen R, Najjar SS, et al. White blood cell count and mortality in the Baltimore longitudinal study of aging. J Am Coll Cardiol 2007;49:1841–50.
46. Erlinger TP, Muntner P, Helzlsouer KJ. WBC count and the risk of cancer mortality in a national sample of US adults: results from the second national health and nutrition examination survey mortality study. Cancer Epidemiol Biomarkers Prev 2004;13:1052–6.
47. Margolis KL, Rodabough RJ, Thomso CA, Lopez AM, McTiernan A. Women's health initiative research group, prospective study of leukocyte count as a predictor of incident breast, colorectal, endometrial, and lung cancer and mortality in postmenopausal women. Arch Intern Med 2007;167:1837–44.
48. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev 2016;30:733–50.

AUTHOR QUERIES

## AUTHOR PLEASE ANSWER ALL QUERIES

Q1: Page: 1: Author: Per journal style, genes, alleles, loci, and oncogenes are italicized; proteins are roman. Please check throughout to see that the words are styled correctly. AACR journals have developed explicit instructions about reporting results from experiments involving the use of animal models as well as the use of approved gene and protein nomenclature at their first mention in the manuscript. Please review the instructions at http://aacrjournals.org/content/authors/editorial-policies#genenomen to ensure that your article is in compliance. If your article is not in compliance, please make the appropriate changes in your proof.

Q2: Page: 1: Author: Please verify the drug names and their dosages used in the article.

Q3: Page: 1: Author: Please verify the affiliations and their corresponding author links.

Q4: Page: 1: Author: Please verify the corresponding author details.

Q5: Page: 3: Author: Please confirm quality/labeling of all images included within this article. Thank you.

Q6: Page: 4: Author: Please verify the layout of Tables for correctness.

Q7: Page: 9: Author: The Authors' Disclosures statement that appears in the proof incorporates the information from forms completed and signed off on by each individual author. No factual changes can be made to disclosure information at the proof stage. However, typographical errors or misspelling of author names should be noted on the proof and will be corrected before publication. Please note if any such errors need to be corrected. Is the disclosure statement correct?

Q8: Page: 9: Author: The contribution(s) of each author are listed in the proof under the heading "Authors' Contributions." These contributions are derived from forms completed and signed off on by each individual author. As the corresponding author, you are permitted to make changes to your own contributions. However, because all authors submit their contributions individually, you are not permitted to make changes in the contributions listed for any other authors. If you feel strongly that an error is being made, then you may ask the author or authors in question to contact us about making the changes. Please note, however, that the manuscript would be held from further processing until this issue is resolved.

AU: Below is a summary of the name segmentation for the authors according to our records. The First Name and the Surname data will be provided to PubMed when the article is indexed for searching. Please check each name carefully and verify that the First Name and Surname are correct. If a name is not segmented correctly, please write the correct First Name and Surname on this page and return it with your proofs. If no changes are made to this list, we will assume that the names are segmented correctly, and the names will be indexed as is by PubMed and other indexing services.

| **First Name** | **Surname** |
|---|---|
| Chenglong | Yu |
| Kristina M. | Jordahl |
| Julie K. | Bassett |
| Jihoon Eric | Joo |
| Ee Ming | Wong |
| Maree T. | Brinkman |
| Daniel F. | Schmidt |
| Damien M. | Bolton |
| Enes | Makalic |
| Theodore M. | Brasky |
| Aladdin H. | Shadyab |
| Lesley F. | Tinker |
| Anthony | Longano |
| John L. | Hopper |
| Dallas R. | English |
| Roger L. | Milne |
| Parveen | Bhatti |
| Melissa C. | Southey |
| Graham G. | Giles |
| Pierre-Antoine | Dugué |