Methylation scores for smoking, alcohol consumption and body mass index and risk of seven types of cancer

Pierre Antoine Dugué, Chenglong Yu, Allison M. Hodge, Eeming Wong, Ji
Hoon E. Joo, Cholhee Jung, Daniel Francis Schmidt, Enes Makalic, Daniel D.
Buchanan, Gianluca Severi, et al.

Methylation scores for smoking, alcohol consumption, and body mass index and risk of seven types of cancer

Pierre-Antoine Dugué,[1,2,3,*] Chenglong Yu,[1] Allison M Hodge,[2,3] Ee Ming Wong,[1,4] JiHoon E Joo,[5] Chol-Hee Jung,[6] Daniel Schmidt,[3] Enes Makalic,[3] Daniel D Buchanan,[5,6,7] Gianluca Severi,[8] Dallas R English,[2,3] John L Hopper,[3] Roger L Milne,[1,2,3] Graham G Giles,[1,2,3] Melissa C Southey[1,2,4]

[1] Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia
[2] Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia
[3] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC, Australia
[4] Department of Clinical Pathology, Melbourne Medical School, The University of Melbourne, Parkville, VIC, Australia
[5] Colorectal Oncogenomics Group, Department of Clinical Pathology, Melbourne Medical School, The University of Melbourne, Parkville, Victoria, Australia
[6] University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Parkville, Victoria Australia
[7] Genomic Medicine and Family Cancer Clinic, Royal Melbourne Hospital, Parkville, Victoria, Australia
[8] Centre de Recherche en Epidémiologie et Santé des Populations (CESP, Inserm U1018), Facultés de Médecine Universités Paris-Saclay, UVSQ, Gustave Roussy, Villejuif, France

* Correspondence and requests for reprints: Pierre-Antoine Dugué, Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, 246 Clayton Rd, Clayton, VIC 3168, Australia, Email: pierre-antoine.dugue@monash.edu

Novelty & Impact Statements: Methylation marks of exposure to health risk factors may better capture current and past exposures than questionnaires, and reflect different individual responses to exposure. We calculated several methylation scores for smoking, body mass index and alcohol consumption, using summary statistics from published epigenome-wide association studies, and found that these were associated with cancer risk, independently of health-related confounders. Methylation scores for lifestyle may provide some improvements in the prediction of cancer risk.

Abbreviations: AHRR: aryl hydrocarbon receptor repressor; AUC: area under the receiver operating characteristic curve; BMI: body mass index; EWAS: epigenome-wide association study; MCCS: Melbourne Collaborative Cohort Study; MS: methylation score; SD: standard deviation,

Submitted: 18/12/2020 | **Words:** 3,026 | **Tables**: 5 | **Supplementary** Methods and 5 Tables

ABSTRACT

Methylation marks of exposure to health risk factors may be useful markers of cancer risk as they might better capture current and past exposures than questionnaires, and reflect different individual responses to exposure. We used data from seven case-control studies nested within the Melbourne Collaborative Cohort Study of blood DNA methylation and risk of colorectal, gastric, kidney, lung, prostate and urothelial cancer, and B-cell lymphoma (N cases=3,123). Methylation scores (MS) for smoking, body mass index (BMI), and alcohol consumption were calculated based on published data as weighted averages of methylation values. Rate ratios (RR) and 95% confidence intervals for association with cancer risk were estimated using conditional logistic regression and expressed per standard deviation increase of the MS, with and without adjustment for health-related confounders. The contribution of MS to discriminate cases from controls was evaluated using the area under the curve (AUC). After confounder adjustment, we observed: large associations (RR~1.5-1.7) with lung cancer risk for smoking MS; moderate associations (RR~1.2-1.3) with urothelial cancer risk for smoking MS and with mature B-cell neoplasm risk for BMI and alcohol MS; moderate to small associations (RR~1.1-1.2) for BMI and alcohol MS with several cancer types and cancer overall. Generally small AUC increases were observed after inclusion of several MS in the same model (colorectal, gastric, kidney, urothelial cancers: +3%; lung cancer: +7%; B-cell neoplasms: +8%). Methylation scores for smoking, BMI, and alcohol consumption show independent associations with cancer risk, and may provide some improvements in risk prediction.

INTRODUCTION

Population-based estimates indicate that smoking, alcohol consumption and obesity combined explain approximately 20% of cancers diagnosed in Australia.[1] These estimates are based on relatively strong assumptions that health risk factors are measured without error in cohort studies, and that population-average estimates reflect individual-level risk. In recent years, a growing body of research has reported associations of genetic and epigenetic measures with these lifestyle factors, raising the potential to increase our understanding of the role they play in the aetiology of cancer at the individual level. Recent research has made successful use of DNA methylation to study, for example, biological ageing [2,3] and its association with cancer risk and mortality.[4-8]

Tobacco smoking, adiposity and alcohol consumption are the health risk factors for which most convincing evidence exist regarding their widespread impact on the methylome; studies have identified thousands of CpG sites at which blood DNA methylation is associated with these exposures.[9-14] This raises potential for increasing the precision with which we measure these health risk factors at the individual level, as it is plausible that i) methylation changes are more abundant and marked in individuals who are more sensitive to these exposures, ii) the levels of exposure collected via cohort questionnaires may be prone to substantial measurement error, and iii) harmful levels of exposure accumulate over the life course and cannot be tracked by only one, two, or more time-point assessments. These aspects likely result in underestimates of association of lifestyle factors with cancer risk. The usefulness of integrated genomic measures to enhance epidemiological exposure assessment is increasingly recognised.[15-18] Recent studies have developed and validated epigenetic predictors for several lifestyle factors.[11,19]

3

In this study, we aimed to assess the association between several methylation-based scores for smoking, alcohol consumption and body mass index (BMI) and the risk of cancer in adulthood. We used data from seven prospective case-control studies nested in the Melbourne Collaborative Cohort Study (MCCS), including 3,123 incident cases of colorectal, gastric, kidney, lung, prostate and urothelial cancers and B-cell lymphoma.

MATERIALS AND METHODS

Study samples and blood collection

The MCCS is a prospective study of 41,513 adult volunteers (24,469 women) aged between 27 and 76 years (99% aged 40-69) when recruited between 1990 and 1994. More details on the cohort are available elsewhere.[20, 21]

We used data from case-control studies of DNA methylation and risk of colorectal, gastric, kidney, lung, mature B-cell lymphoma, prostate and urothelial cancer, nested within the MCCS.[22-27] These cancers account for about 50% of new cancer cases in Victoria in 2018.[28] DNA was extracted from peripheral blood taken at the time of recruitment (1990-1994) except for 151 case-control pairs of the urothelial cancer study, for which blood samples were taken at a follow-up visit in 2003-2007. The DNA source was dried blood spots, peripheral blood mononuclear cells or buffy coats for 70%, 28% and 2% of participants, respectively. More details can be found elsewhere.[21, 29] Incident cancer cases were identified by linkage with the Victorian Cancer Registry and the Australian Cancer Database (Australian Institute of Health and Welfare), which are considered to be virtually complete. Cases were followed from inclusion until 2012-2013, so the follow-up time was up to 23 years. The median from recruitment to diagnosis was 9.7 years (interquartile range: 5.4 to

4

13.6); these figures for individual cancer types are shown in **Supplementary Table 1**. For each nested case-control study, controls were individually matched to incident cases (diagnosed after blood sample collection) on age using incidence density sampling (i.e. they had to be free of the cancer of interest up to the age at diagnosis of the corresponding case), sex, country of birth (Australia/New-Zealand, Southern Europe, Northern Europe), blood DNA source (dried blood spots, peripheral blood mononuclear cells or buffy coat) and blood collection period (baseline or wave 2 for the urothelial cancer study). Controls were also matched to cases on year of birth, except for the colorectal cancer study where controls were matched on year of baseline attendance. For the lung cancer study, controls were also matched on smoking status at the time of blood collection.

DNA extraction and bisulfite conversion, and DNA methylation data processing

Methods relating to DNA extraction and bisulfite conversion, and DNA methylation data processing have been described previously [4, 29] and are detailed in **Supplementary Methods**.

Methylation scores for smoking, alcohol consumption and BMI

We calculated several methylation scores, all as weighted averages: for smoking we calculated **Smk-233**: score comprising 233 CpGs resulting from the Lasso regression applied to predict smoking pack-years (log-transformed), using as weights the regression coefficients available from the original publication from McCartney et al.[19] and **Smk-1,061**: score comprising 1,061 CpGs at which methylation associated with a comprehensive smoking index in the MCCS ($P<10^{-7}$), and also reported to be associated with smoking ($P<10^{-7}$) in previous epigenome-wide association studies (EWAS), using as weights the individual association coefficients from the MCCS EWAS.[12] Similarly, for BMI and alcohol

consumption, we calculated, **Bmi-1109** and **Alc-450**, respectively, based on the predictors developed by McCartney et al.; and **Bmi-85** and **Alc-459** based on associations reported with $P<10^{-7}$ in the MCCS and external data.[13, 14] For alcohol consumption, we also calculated four scores using the weights available from the meta-analysis of Liu et al.[11] These scores included 5, 23, 78 and 144 CpGs (**Alc-5**, **Alc-23**, **Alc-78** and **Alc-144**) and were reported to explain from 5 to 14% of variance in alcohol consumption in testing datasets. The scores calculated from the studies by McCartney et al. and Liu et al. were obtained via *Lasso* and *Elastic net* regression, respectively, hence they aimed to predict these lifestyle factors accurately. The scores calculated from our previous EWAS conducted in the MCCS [12-14] may better represent the cumulative lifestyle-related changes to the methylome. Finally, we calculated the age-adjusted measures of biological ageing *PhenoAge* [30] and *GrimAge,* [31] based on 513 and 1,030 CpGs, respectively, as done previously.[32] All scores were winsorised at five standard deviations from the mean to minimise the potential influence of outliers, and were then rescaled to Z-scores for better comparability of relative risk estimates.

Statistical analysis

Correlations between methylation scores were calculated using Spearman correlations in participants selected as controls (to avoid collider bias). The variance explained in each of smoking, BMI and alcohol consumption by their respective methylation scores was also assessed in controls. We used conditional logistic regression to calculate odds ratios, which are estimates of the rate ratios [RRs] when incidence density sampling matching is used, [33] for the associations between methylation scores, per standard deviation (1SD), and risk of cancer. Two models were used: in Model 1, no covariates were included other than blood cell composition, estimated using the Houseman algorithm;[34, 35] Model 2, further adjusted for smoking status (never; former; current), smoking pack-years, age at starting (never smoked;

age 16 or less; between age 17 and 21; after age 21 years); years since quitting (never smoked; more than 10 years without smoking, between 5 and 10 years without smoking; less than 5 years without smoking), body-mass index (continuous, in kg/m^2), and alcohol consumption in the past week (continuous, in g/day), a score for physical activity,[36], the Alternate Healthy Eating Index 2010 to reflect overall diet quality,[37] education (score ranging from 1 to 8), socioeconomic status (score ranging from 1 to 10 [38]), and height (continuous, in meters). As a sensitivity analysis, the same models were performed without adjustment for white blood cell composition. These models were used to analyse each cancer separately, and all seven cancer types combined. For the combined analysis, where an individual was diagnosed with several cancers, we included the first diagnosis only (respecting the incidence density sampling procedure), so that participants did not contribute twice to the pooled estimates. Case-control pairs with any missing values for the confounders measured at baseline were excluded and missing values at follow-up (urothelial cancers) were replaced by baseline values; 3% of the initial sample was excluded due to missing values.

We assessed the contribution of methylation scores to the ability of the model to discriminate between cases and controls using the area under the receiver operating characteristic curve (AUC) statistics obtained from unconditional logistic regression models adjusted for the matching variables. We evaluated i) the prediction obtained from models including methylation scores for smoking, BMI, and alcohol consumption instead of the lifestyle variables collected in the cohort; ii) the additional contribution of methylation scores compared with a model with a large collection of traditionally collected health risk factors.

RESULTS

The characteristics of study participants were described previously [32] and are shown in **Supplementary Table 2**. The variance explained in the corresponding risk factor was 56% and 28% for Smk-233 and Smk-1061; 15% and 11% for Bmi-1109 and Bmi-85; and 11% and 5% for Alc-450 and Alc-459, respectively (**Table 1**). The correlations between scores were moderate to high within each lifestyle factor, r=0.62 between Smk-233 and Smk-1061; r=0.46 between Bmi-1109 and Bmi-85, and r=0.27 between Alc-450 and Alc-459. Moderate correlations were also observed across lifestyle factors, in particular between smoking and alcohol scores, between Smk-1,061 and Bmi-85 and between smoking scores and *GrimAge* (**Table 1**). There was little overlap of CpGs between methylation scores, e.g. 17 CpGs between Smk-233 and Smk-1061 (**Supplementary Table 3**).

The associations of methylation scores for smoking, BMI and alcohol consumption with risk of cancer are shown in **Table 2, 3 and 4**, respectively. The smoking scores showed strong associations with risk of lung and urothelial cancers in unadjusted models (Model 1, adjustment for white blood cell proportions only): lung cancer, Smk-233: RR per SD=1.84, 95%CI: 1.44-2.36, P=2×10^{-6}; Smk-1061: RR=1.59, 95%CI: 1.26-2.01, P=0.0001; urothelial cancers: Smk-233: RR=1.35, 95%CI: 1.15-1.58, P=0.0002; Smk-1061: RR=1.47, 95%CI: 1.24-1.75, P=8×10^{-6}. There was little attenuation after adjustment for lifestyle factors including several smoking-related variables collected by questionnaires (Model 2), except for Smk33 and urothelial cancer risk (point estimates: lung: Smk-233: 1.68, Smk-1061: 1.49, urothelial: Smk-233: 1.16, Smk-1061: 1.32). Positive but weak associations were observed with risk of cancer overall: unadjusted: Smk-233: 1.09, P=0.006, Smk-1061: 1.11, P=0.0009; adjusted: Smk-233: 1.06, P=0.20, Smk-1061: 1.09, P=0.02. Associations observed without

adjustment for cell composition were similar, albeit slightly weaker (**Supplementary Table 4**).

Methylation scores for BMI generally showed associations with cancer risk in the expected direction but these were weaker than for smoking (**Table 3**). For colorectal cancer, the associations were substantially attenuated in comprehensively adjusted models and were in Model 2: Bmi-1109: RR=1.11, 95%CI: 0.98-1.26, and Bmi-85: RR=1.11, 95%CI: 0.96-1.29. Similar findings were obtained for gastric and kidney cancer risk. Bmi-85 was associated with risk of mature B-cell neoplasms (Model 2: 1.32, 95%CI: 1.04-1.66) and this association was not substantially stronger in models not adjusted for cell proportions, which may substantially confound the associations for this cancer type (RR=1.45, **Supplementary Table 5**). A moderately strong association was observed with risk of urothelial cancer (RR=1.35 and 1.23 in Model 1 and 2, respectively). Bmi-85 showed a stronger association with overall cancer risk than Bmi-1109 (RR=1.10, 95%CI: 1.02-1.19 and RR=1.02, 95%CI: 0.96-1.08, respectively).

For alcohol consumption, we found that all scores were associated with risk of lung cancer (Model 2: RR=1.37, P=0.002 and RR=1.36, P=0.008 for Alc-450 and Alc-459, respectively), **Table 4**. Small associations were observed with risk of colorectal cancer (RR=1.11 and RR=1.09) and a relatively large association for Alc-459 with risk of B-cell lymphoma RR=1.37, 95%CI: 1.11-1.69. For cancer overall, the risk was increased by 13% (4% to 18%) and 11% (6% to 21%) per SD increase in Alc-450 and Alc-459, respectively, after adjustment for cohort-collected sociodemographic, anthropometric and lifestyle variables; these associations were virtually the same as in unadjusted models, without adjustment for cell proportions, or using the scores of Liu et al. (**Supplementary Tables 6 and 7**). For the scores of Liu et al., the associations did not appear larger for the scores including more CpGs,

for example for cancer overall: Alc-144: RR=1.08, 95%CI: 1.02-1.15; Alc-78: RR=1.12, 95%CI: 1.06-1.19; Alc-23: RR=1.15, 95%CI: 1.08-1.22 and Alc-5: RR=1.14, 95%CI: 1.07-1.21 (**Supplementary Table 7**).

Results for the contribution of methylation scores to the prediction of cancer risk are shown in **Table 5**. The AUCs of models with methylation scores instead of traditionally assessed smoking, BMI, and alcohol consumption appeared similar for all cancer types and were somewhat higher for lung cancer (0.60 vs 0.59) and mature B-cell neoplasms (0.59 vs 0.57). The addition of methylation scores to variables that were collected via questionnaires or measured in the MCCS nevertheless resulted in relatively small AUC gains: models including two methylation scores for each of smoking, BMI, and alcohol consumption, *GrimAge* and *PhenoAge* compared with models without: colorectal cancer: AUC=0.59 vs 0.56, gastric: 0.64 vs 0.61, kidney: 0.70 vs 0.67, lung: 0.66 vs 0.59, mature B-cell neoplasms: 0.66 vs 0.57, prostate: 0.56 vs 0.55, urothelial: 0.63 vs 0.60.

DISCUSSION

We calculated various methylation scores for smoking, alcohol consumption and BMI, the main lifestyle factors that cause changes to DNA methylation and increase the risk of cancer.[12-14, 39-41] Several of these were associated with certain types of cancer and associations were only moderately attenuated after adjustment for sociodemographic, anthropometric and lifestyle factors, including several questionnaire-collected variables relating to smoking history (pack-years, age at starting, time since quitting). These associations were nevertheless generally relatively small, except for smoking scores and lung and urothelial cancer risk, for which we reported some of the findings previously.[42] Reasonably large associations were also

10

observed for methylation scores for BMI and alcohol consumption with risk of mature B-cell lymphoma (RR per SD~1.3-1.4), but these should be interpreted with some caution given the strongly modified blood methylation profiles found in B-cell neoplasms, [26] and the fact that lifestyle is thought to only moderately increase risk for these types of cancer.[43] For risk of cancer overall, the methylation scores based on weights derived from EWAS conducted in the MCCS [12-14] produced somewhat stronger associations with cancer risk than those obtained via regularised regression,[19] being estimated in comprehensively adjusted models at 9%, 10%, and 13% increase, compared with 6%, 2% and 11%, per SD of smoking, BMI, and alcohol methylation scores, respectively. That associations appeared stronger with these scores than with those of McCartney et al. [19] may be explained in part by the way the scores were built (i.e. 'cumulative damage' provided by the weighted average for all CpGs identified as top-ranked associations between health risk factors and DNA methylation, instead of 'best predictor', obtained via regularised regression, which aims to maximise prediction), and perhaps more importantly because methylation at all CpGs they included was found to be strongly associated with these lifestyle factors in the samples used in this study.[12-14]. Consistent with this, the predictor found by Liu et al. [11] to be the best predictor of alcohol consumption in external data (based on 144 CpGs) performed less well than the reduced versions of it in terms of strength of association with cancer risk. In line with the relatively small adjusted associations we observed, these scores, in addition to *PhenoAge* [30, 32] and *GrimAge,*[31, 32] only provided moderate added discrimination between cases and controls (~3% for colorectal, gastric, kidney and urothelial cancers, ~7% for lung cancer and ~8% for mature B-cell lymphomas).

The two main strengths of this study are i) its prospective design using incidence density sampling with improved control for confounding and efficiency by carefully matching cases and controls on several cancer-associated factors, including smoking history

11

for the lung cancer study, as well as very good control for batch effects provided by placing matched pairs on the same slide of the array; and ii) its large sample size (3,123 cancer cases) and the inclusion of several cancer types. The main limitation of our study is that it does not provide a 'single best' or integrated / synthetic methylation score representative of each lifestyle factor. It could also be that while methylation scores provided a good summary of the effect of lifestyle on the methylome, they were not able to capture the methylation patterns specific to risk for each cancer, as it is likely that methylation changes in certain genomic regions may have different implications regarding susceptibility to different cancers.

To our knowledge, our study is novel in the field of cancer, and reports of similar attempts to assess other health outcomes are limited. In the study by McCartney et al.,[19] the authors examined all-cause mortality and showed it was associated with their BMI methylation score (HR=1.23, 95%CI: 1.06-1.42, after adjustment for age, sex, white blood cell proportions, smoking methylation score, BMI polygenic score and measured BMI). Somewhat weaker associations were observed for waist-to-hip ratio and body fat percentage, which we did not include in our study. A negative association was observed for methylation-predicted total cholesterol. The association for their alcohol consumption methylation score was similar to those observed in our study (HR=1.11, 95%CI: 0.97-1.28). A strong association was observed for the smoking methylation score, in particular after adjustment for BMI methylation score (HR=1.57, 95%CI: 1.39-1.78). Other studies published by the same group showed increased prediction of several health-related outcomes for their BMI methylation score [44] and their smoking methylation score [45]. In the study by Langdon et al., [46] methylation scores were evaluated in participants with oropharyngeal cancer and moderately large associations (HR~1.2-1.3) with survival were observed for some scores in adjusted models, with modest improvements in terms of prediction. Three other scores were evaluated in that study [10, 47, 48] as well as just methylation at cg05575921 (*AHRR*, smoking), [12] for which

the results were similar to those we considered. Other attempts were also made to use individual, or a restricted set of, smoking-associated methylation marks to predict lung cancer or other smoking-related outcomes, focusing on the strongest observed associations (e.g. *AHRR*), for example in the studies by Baglietto et al. including MCCS data, [23] Zhang et al., [49, 50] and others. [51-54]

We conclude that methylation scores for smoking, BMI, and alcohol consumption show some associations with risk of cancer, independently of many health-related variables, but these were generally relatively small. While these scores could potentially replace the variables routinely collected in cohorts, their added contribution to the prediction of cancer is modest.

Data Availability Statement: The methylation array data that support the findings of this study will be deposited to an appropriate (restricted-access) public repository for public release upon publication.

Ethics Statement: Participants provided informed consent in accordance with the Declaration of Helsinki and the study was approved by Cancer Council Victoria's Human Research Ethics Committee.

Conflicts of interest: None declared.

REFERENCES

1. Wilson LF, Antonsson A, Green AC, Jordan SJ, Kendall BJ, Nagle CM, Neale RE, Olsen CM, Webb PM, Whiteman DC. How many cancer cases and deaths are potentially preventable? Estimates for Australia in 2013. *Int J Cancer* 2018;**142**: 691-701.

2. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013;**14**: R115.

3. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* 2015;**16**: 25.

4. Dugué PA, Bassett JK, Joo JE, Jung CH, Ming Wong E, Moreno-Betancur M, Schmidt D, Makalic E, Li S, Severi G, Hodge AM, Buchanan DD, et al. DNA methylation-based biological aging and cancer risk and survival: Pooled analysis of seven prospective studies. *Int J Cancer* 2018;**142**: 1611-9.

5. Dugué PA, Bassett JK, Joo JE, Baglietto L, Jung CH, Wong EM, Fiorito G, Schmidt D, Makalic E, Li S, Moreno-Betancur M, Buchanan DD, et al. Association of DNA Methylation-Based Biological Age With Health Risk Factors and Overall and Cause-Specific Mortality. *Am J Epidemiol* 2018;**187**: 529-38.

6. Dugué PA, Bassett JK, Wong EM, Joo JE, Li S, Yu C, Schmidt DF, Makalic E, Doo NW, Buchanan DD, Hodge AM, English DR, et al. Biological Aging Measures Based on Blood DNA Methylation and Risk of Cancer: A Prospective Study. *JNCI Cancer Spectr* 2021;**5**: pkaa109.

7. Dugué PA, Li S, Hopper JL, Milne RL. DNA Methylation-Based Measures of Biological Aging. In: Tollefsbol TO. *Epigenetics in Human Disease (Second Edition)* ed., vol. 6. London: Academic Press, 2018: 39-64.

8. Ambatipudi S, Horvath S, Perrier F, Cuenin C, Hernandez-Vargas H, Le Calvez-Kelm F, Durand G, Byrnes G, Ferrari P, Bouaoun L, Sklias A, Chajes V, et al. DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. *Eur J Cancer* 2017;**75**: 299-307.

9. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai PC, Ried JS, Zhang W, Yang Y, Tan S, Fiorito G, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* 2017;**541**: 81-6.

10. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S, Moreno-Macias H, Smith JA, et al. Epigenetic Signatures of Cigarette Smoking. *Circulation: Cardiovascular Genetics* 2016;**9**: 436-47.

11. Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai PC, Reynolds LM, Just AC, Duan Q, Boer CG, Tanaka T, Elks CE, Aslibekyan S, et al. A DNA methylation biomarker of alcohol consumption. *Molecular psychiatry* 2016;**23**: 422-33.

12. Dugué PA, Jung CH, Joo JE, Wang X, Wong EM, Makalic E, Schmidt DF, Baglietto L, Severi G, Southey MC, English DR, Giles GG, et al. Smoking and blood DNA methylation: an epigenome-wide association study and assessment of reversibility. *Epigenetics* 2020;**15**: 358-68.

13. Dugué PA, Wilson R, Lehne B, Jayasekara H, Wang X, Jung CH, Joo JE, Makalic E, Schmidt DF, Baglietto L, Severi G, Gieger C, et al. Alcohol consumption is associated with widespread changes in blood DNA methylation: Analysis of cross-sectional and longitudinal data. *Addiction Biol* 2019: e12855.

14. Geurts YM, Dugué PA, Joo JE, Makalic E, Jung CH, Guan W, Nguyen S, Grove ML, Wong EM, Hodge AM, Bassett JK, FitzGerald LM, et al. Novel associations between blood DNA methylation and body mass index in middle-aged and older adults. *Int J Obes (Lond)* 2018;**42**: 887-96.

15. Barrow TM. Epigenetic Epidemiology for Cancer Risk *Epigenetic Mechanisms in Cancer*ed.: Elsevier, 2018: 195-222.

16. Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. *Genome Med* 2017;**9**: 96.

17. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013;**9**: e1003348.

18. Widschwendter M, Jones A, Evans I, Reisel D, Dillner J, Sundstrom K, Steyerberg EW, Vergouwe Y, Wegwarth O, Rebitschek FG, Siebert U, Sroczynski G, et al. Epigenome-based cancer risk prediction: rationale, opportunities and challenges. *Nat Rev Clin Oncol* 2018;**15**: 292-309.

19. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, Morris SW, Bermingham ML, Campbell A, Murray AD, Whalley HC, Gale CR, et al. Epigenetic prediction of complex traits and death. *Genome Biol* 2018;**19**: 136.

20. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC scientific publications* 2002;**156**: 69-70.

21. Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK, Bruinsma FJ, Lynch BM, Dugué PA, Jayasekara H, Brinkman MT, Popowski LV, et al. Cohort Profile: The Melbourne Collaborative Cohort Study (Health 2020). *Int J Epidemiol* 2017;**46**: 1757-i.

22. Dugué PA, Brinkman MT, Milne RL, Wong EM, FitzGerald LM, Bassett JK, Joo JE, Jung C-H, Makalic E, Schmidt DF, Park DJ, Chung J, 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. Baglietto L, Ponzi E, Haycock P, Hodge A, Bianca Assumma M, Jung CH, Chung J, Fasanelli F, Guida F, Campanella G, Chadeau-Hyam M, Grankvist K, 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.

24. FitzGerald LM, Naeem H, Makalic E, Schmidt DF, Dowty JG, Joo JE, Jung CH, Bassett JK, Dugué PA, Chung J, Lonie A, Milne RL, et al. Genome-Wide Measures of Peripheral Blood Dna Methylation and Prostate Cancer Risk in a Prospective Nested Case-Control Study. *Prostate* 2017;**77**: 471-8.

25. Bodelon C, Ambatipudi S, Dugué PA, Johansson A, Sampson JN, Hicks B, Karlins E, Hutchinson A, Cuenin C, Chajès V, Southey MC, Romieu I, et al. Blood DNA methylation and breast cancer risk: a meta-analysis of four prospective cohort studies. *Breast Cancer Res* 2019;**21**: 62.

26. Wong Doo N, Makalic E, Joo JE, Vajdic CM, Schmidt DF, Wong EM, Jung C-H, Severi G, Park DJ, Chung J, Baglietto L, Prince HM, et al. Global measures of peripheral blood-derived DNA methylation as a risk factor in the development of mature B-cell neoplasms. *Epigenomics* 2016;**8**: 55-66.

27. Chamberlain JA, Dugué PA, Bassett JK, Milne RL, Joo JE, Wong EM, Brinkman MT, Stuart GW, Boussioutas A, Southey MC. DNA methylation in peripheral blood and risk of gastric cancer: a prospective nested case-control study. *Cancer Prev Res* 2020. doi: 10.1158/1940-6207.CAPR-20-0003

28. Victorian Cancer Registry, Cancer in Victoria: Statistics & Trends 2018. Cancer Council Victoria, Melbourne, Victoria, 2019.

29. Dugué PA, English DR, MacInnis RJ, Jung CH, Bassett JK, FitzGerald LM, Wong EM, Joo JE, Hopper JL, Southey MC, Giles GG, Milne RL. Reliability of DNA methylation measures from dried blood spots and mononuclear cells using the HumanMethylation450k BeadArray. *Sci Rep* 2016;**6**: 30317.

30. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* 2018;**10**: 573-91.

31. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, Whitsel EA, Assimes TL, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* 2019;**11**: 303-27.

32. Dugué PA, Bassett JK, Wong EM, Joo JE, Li S, Yu C, Schmidt DF, Makalic E, Doo NW, Buchanan DD. Biological aging measures based on blood DNA methylation and risk of cancer: a prospective study. *medRxiv* 2020.

33. Pearce N. What does the odds ratio estimate in a case-control study? *Int J Epidemiol* 1993;**22**: 1189-92.

34. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;**13**: 86.

35. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome biology* 2014;**15**: R31.

36. MacInnis RJ, English DR, Hopper JL, Haydon AM, Gertig DM, Giles GG. Body size and composition and colon cancer risk in men. *Cancer Epidemiol Biomarkers Prev* 2004;**13**: 553-9.

37. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, Stampfer MJ, Willett WC. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr* 2012;**142**: 1009-18.

38. Pink B, Socio-economic indexes for areas (SEIFA). Australian Bureau of Statistics, 2013.

39. Chamberlain JA, Dugué PA, Bassett JK, Hodge AM, Brinkman MT, Joo JE, Jung CH, Makalic E, Schmidt DF, Hopper JL, Buchanan DD, English DR, et al. Dietary intake of one-carbon metabolism nutrients and DNA methylation in peripheral blood. *Am J Clin Nutr* 2018;**108**: 611-21.

40. van Roekel EH, Dugué PA, Jung CH, Joo JE, Makalic E, Wong EEM, English DR, Southey MC, Giles GG, Lynch BM, Milne RL. Physical Activity, Television Viewing Time, and DNA Methylation in Peripheral Blood. *Med Sci Sports Exerc* 2019;**51**: 490-8.

41. Li S, Dugué PA, Baglietto L, Severi G, Wong EM, Nguyen TL, Stone J, English DR, Southey MC, Giles GG, Hopper JL, Milne RL. Genome-wide association study of

peripheral blood DNA methylation and conventional mammographic density measures. *Int J Cancer* 2019;**145**: 1768-73.

42. Dugué PA, Hodge AM, Wong EM, Joo JE, Jung C-H, Hopper JL, English DR, Giles GG, Milne RL, Southey MC. Methylation marks of prenatal exposure to maternal smoking and risk of cancer in adulthood. *Int J Epidemiol* 2020.

43. Flowers CR, Skibola CF. Identifying risk factors for B-cell lymphoma. *Blood* 2016;**127**: 10-1.

44. Hamilton OKL, Zhang Q, McRae AF, Walker RM, Morris SW, Redmond P, Campbell A, Murray AD, Porteous DJ, Evans KL, McIntosh AM, Deary IJ, et al. An epigenetic score for BMI based on DNA methylation correlates with poor physical health and major disease in the Lothian Birth Cohort. *Int J Obes (Lond)* 2019;**43**: 1795-802.

45. Corley J, Cox SR, Harris SE, Hernandez MV, Maniega SM, Bastin ME, Wardlaw JM, Starr JM, Marioni RE, Deary IJ. Epigenetic signatures of smoking associate with cognitive function, brain structure, and mental and physical health outcomes in the Lothian Birth Cohort 1936. *Transl Psychiatry* 2019;**9**: 248.

46. Langdon RJ, Beynon RA, Ingarfield K, Marioni RE, McCartney DL, Martin RM, Ness AR, Pawlita M, Waterboer T, Relton C, Thomas SJ, Richmond RC. Epigenetic prediction of complex traits and mortality in a cohort of individuals with oropharyngeal cancer. *Clin Epigenetics* 2020;**12**: 58.

47. Banos DT, McCartney DL, Patxot M, Anchieri L, Battram T, Christiansen C, Costeira R, Walker RM, Morris SW, Campbell A. Bayesian reassessment of the epigenetic architecture of complex traits. *Nat Commun* 2020;**11**: 1-14.

48. Zhang Y, Florath I, Saum K-U, Brenner H. Self-reported smoking, serum cotinine, and blood DNA methylation. *Environ Res* 2016;**146**: 395-403.

49. Zhang Y, Elgizouli M, Schottker B, Holleczek B, Nieters A, Brenner H. Smoking-associated DNA methylation markers predict lung cancer incidence. *Clin Epigenetics* 2016;**8**: 127.

50. Zhang Y, Schöttker B, Florath I, Stock C, Butterbach K, Holleczek B, Mons U, Brenner H. Smoking-Associated DNA Methylation Biomarkers and Their Predictive Value for All-Cause and Cardiovascular Mortality. *Environ Health Perspectives* 2016;**124**: 67-74.

51. Kodal JB, Kobylecki CJ, Vedel-Krogh S, Nordestgaard BG, Bojesen SE. AHRR hypomethylation, lung function, lung function decline and respiratory symptoms. *Eur Resp J* 2018;**51**.

52. Morrow JD, Make B, Regan E, Han M, Hersh CP, Tal-Singer R, Quackenbush J, Choi AM, Silverman EK, DeMeo DL. DNA methylation is predictive of mortality in current and former smokers. *American Journal of Respiratory and Critical Care Medicine* 2020;**201**: 1099-109.

53. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax* 2017;**72**: 646-53.

54. Kemp Jacobsen K, Johansen JS, Mellemgaard A, Bojesen SE. AHRR (cg05575921) methylation extent of leukocyte DNA and lung cancer survival. *PloS one* 2019;**14**: e0211745.

Table 1. Correlations between methylation scores

	R²,a	Smk-233	Smk-1061	Bmi-1109	Bmi-85	Alc-450	Alc-459	Alc-144	Alc-78	Alc-23	Alc-5	*PhenoAge*	*GrimAge*
Age		-0.01	0.17	0.02	0.15	0.00	0.17	-0.02	0.03	0.14	0.17	0.00	0.00
Smk-233	56%	1	0.62	0.05	0.13	0.16	0.13	0.07	0.10	0.11	0.08	0.08	0.58
Smk-1061	28%		1	0.05	0.34	0.23	0.45	0.11	0.20	0.34	0.31	0.13	0.42
Bmi-1109	15%			1	0.46	0.03	-0.03	-0.02	0.01	0.04	0.07	0.06	0.12
Bmi-85	11%				1	0.01	-0.10	-0.13	-0.06	0.14	0.15	0.06	0.05
Alc-450	11%					1	0.27	0.26	0.33	0.42	0.45	0.06	0.19
Alc-459	5%						1	0.47	0.55	0.53	0.50	0.25	0.33
Alc-144	11%							1	0.92	0.59	0.43	0.07	0.08
Alc-78	12%								1	0.75	0.59	0.09	0.12
Alc-23	10%									1	0.94	0.09	0.12
Alc-5	9%										1	0.12	0.16
PhenoAge	0%											1	0.35
GrimAge	0%												1

[a] The variance explained R^2 is given for each lifestyle factor; Smk-233 and Smk-1061: smoking status (current / former / never); Bmi-1109 and Bmi-85: BMI (continuous); Alc-450, Alc-459, Alc-144, Alc-78, Alc-23 and Alc-5: alcohol consumption (continuous). *PhenoAge* and *GrimAge* are independent of age because the measures were regressed on age. The variance explained before adjustment for age was 48% and 63%, respectively.

Table 2. Association (rate ratios, 95% confidence intervals) between methylation scores for smoking and cancer risk.

Cancer type (N cases)	Model 1[a]			Model 2[a]		
Colorectal (N=814)	**RR**	**CI95**	**P**	**RR**	**CI95**	**P**
Smk-233	1.06	0.94-1.19	0.32	0.91	0.75-1.1	0.34
Smk-1061	1.06	0.94-1.20	0.31	0.98	0.84-1.15	0.83
Gastric (N=166)						
Smk-233	1.01	0.79-1.29	0.95	1.10	0.73-1.67	0.65
Smk-1061	0.95	0.72-1.27	0.74	0.90	0.59-1.38	0.63
Kidney (N=139)						
Smk-233	1.09	0.84-1.42	0.50	0.82	0.46-1.47	0.50
Smk-1061	0.97	0.73-1.30	0.85	0.78	0.47-1.31	0.35
Lung (N=327)						
Smk-233	1.84	1.44-2.36	2×10^{-6}	1.68	1.29-2.19	0.0001
Smk-1061	1.59	1.26-2.01	0.0001	1.49	1.17-1.90	0.001
MBCN (N=426)						
Smk-233	0.97	0.80-1.17	0.74	0.93	0.69-1.24	0.61
Smk-1061	1.12	0.93-1.35	0.22	1.21	0.96-1.52	0.11
Prostate (N=847)						
Smk-233	0.89	0.79-1.00	0.04	0.90	0.75-1.08	0.25
Smk-1061	0.88	0.78-1.00	0.05	0.89	0.76-1.04	0.15
Urothelial (N=404)						
Smk-233	1.35	1.15-1.58	0.0002	1.16	0.9-1.48	0.25
Smk-1061	1.47	1.24-1.75	8×10^{-6}	1.32	1.06-1.63	0.01
All types (N=3,000)						
Smk-233	1.09	1.02-1.15	0.006	1.06	0.97-1.16	0.20
Smk-1061	1.11	1.04-1.18	0.0009	1.09	1.01-1.18	0.02

[a] RRs were calculated using conditional logistic regression models. Cases and controls were matched on age, sex, country of birth, sample type and smoking status (for the lung cancer study), Matched pairs were placed consecutively on a same chip of the assay, at random positions.

Model 1: adjusted for white blood cell composition

Model 2: additionally adjusted for smoking status (current/former/never), alcohol consumption (in g/day) and BMI (in kg/m2), smoking pack-years, age at starting (never smoked; age 16 or less; between age 17 and 21; after age 21 years); years since quitting (never smoked; more than 10 years without smoking, between 5 and 10 years without smoking; less than 5 years without smoking), a score for physical activity, the alternate healthy eating index to reflect overall diet quality, education (score ranging from 1 to 8), socioeconomic status (score ranging from 1 to 10) and height (continuous, in meters)

Table 3. Association (rate ratios, 95% confidence intervals) between methylation scores for BMI and cancer risk.

Cancer type (N cases)	Model 1[a]			Model 2[a]		
Colorectal (N=814)	RR	95% CI	P	RR	95% CI	P
Bmi-1109	1.14	1.03-1.27	0.01	1.11	0.98-1.26	0.10
Bmi-85	1.18	1.03-1.35	0.01	1.11	0.96-1.29	0.18
Gastric (N=166)						
Bmi-1109	1.27	0.98-1.64	0.07	1.29	0.97-1.72	0.08
Bmi-85	1.15	0.83-1.59	0.40	1.14	0.79-1.63	0.49
Kidney (N=139)						
Bmi-1109	1.30	1.03-1.63	0.03	1.22	0.92-1.62	0.17
Bmi-85	1.16	0.82-1.66	0.39	0.97	0.62-1.54	0.91
Lung (N=327)						
Bmi-1109	0.90	0.76-1.06	0.22	0.99	0.83-1.19	0.94
Bmi-85	0.97	0.78-1.21	0.78	1.08	0.84-1.38	0.57
MBCN (N=426)						
Bmi-1109	0.87	0.74-1.01	0.07	0.81	0.68-0.97	0.02
Bmi-85	1.25	1.02-1.53	0.03	1.32	1.04-1.66	0.02
Prostate (N=847)						
Bmi-1109	0.99	0.89-1.11	0.92	0.97	0.86-1.10	0.61
Bmi-85	0.90	0.78-1.05	0.18	0.90	0.77-1.05	0.19
Urothelial (N=404)						
Bmi-1109	1.05	0.89-1.23	0.58	1.03	0.87-1.23	0.72
Bmi-85	1.35	1.10-1.66	0.004	1.23	0.98-1.55	0.07
All types (N=3,000)						
Bmi-1109	1.03	0.98-1.09	0.25	1.02	0.96-1.08	0.57
Bmi-85	1.11	1.04-1.20	0.003	1.10	1.02-1.19	0.01

[1] RRs were calculated using conditional logistic regression models. Cases and controls were matched on age, sex, country of birth, sample type and smoking status (for the lung cancer study), Matched pairs were placed consecutively on a same chip of the assay, at random positions.

Model 1: adjusted for white blood cell composition

Model 2: additionally adjusted for smoking status (current/former/never), alcohol consumption (in g/day) and BMI (in kg/m2), smoking pack-years, age at starting (never smoked; age 16 or less; between age 17 and 21; after age 21 years); years since quitting (never smoked; more than 10 years without smoking, between 5 and 10 years without smoking; less than 5 years without smoking), a score for physical activity, the alternate healthy eating index to reflect overall diet quality, education (score ranging from 1 to 8), socioeconomic status (score ranging from 1 to 10) and height (continuous, in meters)

Table 4. Association (rate ratios, 95% confidence intervals) between methylation scores for alcohol consumption and cancer risk

Cancer type (N cases)	Model 1[a]			Model 2[a]		
Colorectal (N=814)	**RR**	**95% CI**	**P**	**RR**	**95% CI**	**P**
Alc-450	1.12	1.01-1.25	0.04	1.11	0.99-1.25	0.07
Alc-459	1.13	0.99-1.29	0.08	1.09	0.95-1.26	0.23
Gastric (N=166)						
Alc-450	1.04	0.80-1.34	0.79	0.98	0.73-1.32	0.90
Alc-459	0.88	0.63-1.23	0.45	0.8	0.53-1.21	0.28
Kidney (N=139)						
Alc-450	1.02	0.78-1.32	0.91	1.14	0.83-1.57	0.43
Alc-459	0.80	0.56-1.14	0.21	0.86	0.55-1.34	0.51
Lung (N=327)						
Alc-450	1.39	1.17-1.66	0.0002	1.37	1.12-1.67	0.002
Alc-459	1.32	1.07-1.63	0.008	1.36	1.08-1.70	0.008
MBCN (N=426)						
Alc-450	1.01	0.86-1.18	0.93	1.03	0.87-1.23	0.72
Alc-459	1.29	1.06-1.57	0.01	1.37	1.11-1.69	0.003
Prostate (N=847)						
Alc-450	1.04	0.93-1.16	0.53	1.04	0.92-1.18	0.55
Alc-459	1.02	0.89-1.17	0.80	1.01	0.87-1.17	0.92
Urothelial (N=404)						
Alc-450	1.12	0.96-1.31	0.15	1.06	0.89-1.26	0.54
Alc-459	1.14	0.97-1.33	0.11	1.05	0.89-1.25	0.55
All types (N=3,000)						
Alc-450	1.11	1.05-1.17	0.0002	1.11	1.04-1.18	0.0007
Alc-459	1.14	1.07-1.21	8×10^{-5}	1.13	1.06-1.21	0.0004

[1] RRs were calculated using conditional logistic regression models. Cases and controls were matched on age, sex, country of birth, sample type and smoking status (for the lung cancer study), Matched pairs were placed consecutively on a same chip of the assay, at random positions.

Model 1: adjusted for white blood cell composition

Model 2: additionally adjusted for smoking status (current/former/never), alcohol consumption (in g/day) and BMI (in kg/m2), smoking pack-years, age at starting (never smoked; age 16 or less; between age 17 and 21; after age 21 years); years since quitting (never smoked; more than 10 years without smoking, between 5 and 10 years without smoking; less than 5 years without smoking), a score for physical activity, the alternate healthy eating index to reflect overall diet quality, education (score ranging from 1 to 8), socioeconomic status (score ranging from 1 to 10) and height (continuous, in meters)

Table 5. AUCs for various combinations of methylation scores and traditionally-assessed lifestyle variables.

Cancer type	Colorectal	Gastric	Kidney	Lung	MBCN	Prostate	Urothelial
Model used	**AUC**	**AUC**	**AUC**	**AUC**	**AUC**	**AUC**	**AUC**
Model 0: Age, sex, country of birth, physical activity, diet quality, socioeconomic status, education and height	0.53	0.60	0.64	0.53	0.56	0.54	0.55
Model 1: Model 0 + smoking, alcohol, BMI	0.56	0.61	0.67	0.59	0.57	0.55	0.60
Model 0 + methylation scores smoking, alcohol, and BMI	0.55	0.61	0.66	0.60	0.59	0.55	0.60
Model 2: Model 1 + methylation scores smoking, alcohol, BMI	0.57	0.61	0.69	0.63	0.59	0.55	0.61
Model 3: Model 2 + two methylation scores for smoking, alcohol, BMI	0.58	0.63	0.69	0.64	0.65	0.55	0.62
Model 4: Model 3 + *PhenoAge*, and *GrimAge*	0.59	0.64	0.70	0.66	0.66	0.56	0.63