Author/s:
Penha, RCC;Onwuka, JU;Langdon, R;Sandanger, TM;Nøst, TH;Vineis, P;Johansson, M;Milne, RL;Dugué, P-A;Relton, C;Suderman, M;McKay, J;Johansson, M

Title:
DNA methylation-based clocks, tobacco smoking, and lung cancer risk.

Date:
2025-12-17

Citation:
Penha, R. C. C., Onwuka, J. U., Langdon, R., Sandanger, T. M., Nøst, T. H., Vineis, P., Johansson, M., Milne, R. L., Dugué, P. -A., Relton, C., Suderman, M., McKay, J. & Johansson, M. (2025). DNA methylation-based clocks, tobacco smoking, and lung cancer risk.. BMC Med, 24 (1), pp.40-. https://doi.org/10.1186/s12916-025-04542-9.

Persistent Link:
https://hdl.handle.net/11343/368201

BMC Medicine

RESEARCH

Open Access

DNA methylation-based clocks, tobacco smoking, and lung cancer risk

Ricardo Cortez Cardoso Penha[1†], Justina Ucheojor Onwuka[1†], Ryan Langdon[1], Torkjel M. Sandanger[2], Therese Haugdahl Nøst[2], Paolo Vineis[3], Mikael Johansson[4], Roger L. Milne[5,6,7], Pierre-Antoine Dugué[5,6,7], Caroline Relton[8], Matthew Suderman[9], James McKay[1] and Mattias Johansson[1*]

Abstract

Background Biological age, estimated by DNA methylation-based (DNAm) clocks, has been reported to be associated with lung cancer risk. However, the extent to which tobacco smoking behaviours can explain this association and the extent to which DNAm clocks and their components can inform risk assessment for lung cancer remains to be elucidated. This study aimed to evaluate the relationship between DNAm clocks, smoking, and lung cancer risk.

Methods We analyzed four prospective cohorts (MCCS, Australia, 324 cases/324 controls; NSHDS, Sweden, 190 cases/190 controls; EPIC, Italy, 160 cases/107 controls; and NOWAC, Norway, 115 case/70 controls) with blood samples collected before lung cancer diagnosis. Study participants were restricted to those with a history of smoking. Incidence sampling was used to match one control to each of the lung cancer cases by cohort, sex, date of blood collection, age, and smoking status in MCCS and NSHDS. The risk discriminative performance of age-adjusted DNAm clocks and their components was compared with that of the Prostate, Lung, Colorectal, and Ovarian model 2012 ($PLCO_{m2012}$) lung cancer risk model.

Results We found several DNAm clocks positively associated with lung cancer risk (Hannum: OR = 1.13, 95% CI = 1.02–1.26; PhenoAge: OR = 1.25, 95% CI = 1.12–1.40; DunedinPACE: OR = 1.44, 95% CI = 1.29–1.62; PCGrimAge (a principal component-denoised GrimAge): OR = 1.79, 95% CI = 1.56–2.06), after adjustment for age and tobacco smoking. Tobacco smoking explained a modest proportion of variance in most age-adjusted DNAm clocks ($R^2 < 11\%$), except for PCGrimAge, where it accounted for ~ 30% of variance in both lung cancer cases and controls. Detailed smoking adjustments attenuated the PCGrimAge association with lung cancer risk by 13%. In a secondary analysis adjusting for PCGrimAge components and the $PLCO_{m2012}$ score, DNA methylation-predicted packyears emerged as an independent predictor of lung cancer risk (OR = 2.23, 95% CI = 1.58–3.14). The area under the receiver operating characteristic curve (AUC) for the $PLCO_{m2012}$ model was 0.66 (95% CI = 0.61–0.71) compared with 0.72 (95% CI = 0.67–0.77) for the PCGrimAge model ($P_{difference} = 0.03$). Combining PCGrimAge with $PLCO_{m2012}$ provided similar risk discrimination as PCGrimAge alone (AUC = 0.72, 95% CI = 0.67–0.77).

Conclusions Methylation-based biological clocks capture epigenetic marks left by exposure to tobacco smoke, and some clocks may inform lung cancer risk assessment by complementing or replacing traditional prediction models.

†Ricardo Cortez Cardoso Penha and Justina Ucheojor Onwuka contributed equally.

*Correspondence:
Mattias Johansson
johanssonm@iarc.who.int
Full list of author information is available at the end of the article

Keywords Biomarkers, Lung cancer, Epidemiology, DNA methylation clocks, Tobacco smoking

Background

Screening high-risk individuals with low-dose computed tomography (LDCT) reduces mortality from lung cancer [1]. The US Preventive Services Task Force (USPSTF) eligibility criteria for LDCT are based on chronological age (number of years an individual has been alive since birth; 50–80 years) and smoking history (≥20 packyears for both current smokers and former smokers who quit less than 15 years prior) [2]. However, many incident lung cancer cases occur in individuals with smoking history who are ineligible for screening based on these criteria [3]. In addition to defining target groups based on the USPSTF eligibility criteria, absolute risk thresholds from risk-prediction models such as the $PLCO_{m2012}$ model can identify individuals at high risk of developing lung cancer [4, 5].

The $PLCO_{m2012}$ lung cancer risk model, developed from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, represents the current standard for risk-based screening eligibility [6]. This model integrates demographic factors (age, sex, race); smoking history (pack-years, quit duration); medical history (family history, COPD, prior cancer); and socioeconomic variables (education, BMI) to predict 6-year lung cancer risk [2]. Despite good performance [6], $PLCO_{m2012}$ relies on self-reported variables that may not accurately capture individual's historical smoking exposure, leaving opportunities to improve risk assessment. To improve the identification of high-risk individuals who will benefit from screening, the USPSTF recommended the development of biomarkers for early lung cancer diagnosis [2]. The $PLCO_{m2012}$ model is already used in some screening settings, but many lung cancer cases remain ineligible, and incorporating biomarkers offer an opportunity to improve risk assessment and screening efficiency. DNA methylation aging clocks offer objective, blood-based biomarkers that integrate cumulative biological aging effects from genetic, environmental, and lifestyle factors [7, 8]. These epigenetic markers may complement traditional risk factors by providing objective and more precise measures of smoking-related biological damage to identify high-risk individuals missed by conventional models [9, 10].

Biological age, which reflects a physiological state of individuals over time, can be estimated by DNA methylation-based (DNAm) clocks. These clocks are built on sparse linear regression models trained on DNA methylation levels of CpG sites that vary with chronological age and/or are also associated with clinical biomarkers [11]. Multiple DNAm clocks have been developed to capture different facets of aging, and they can be divided into two general categories [11]. The first-generation clocks (e.g. Horvath's [12]) were trained on CpG sites that vary along with chronological age. The second-generation clocks were built to identify functional features linked to biological age [13] (e.g. mitotic age as for epiTOC2 [14]) or to predict all-cause mortality risk (e.g. GrimAge [15]), while DunedinPACE uniquely measures the pace of aging—how fast an individual is aging rather than their current biological age [11, 15]. The difference between the biological and chronological age of individuals (age acceleration) has been used as a biomarker to predict risk of aging-related diseases such as cancer, as well as all-cause and cancer-specific mortality [16, 17]. Age acceleration, estimated by DNAm or clinical blood-based biomarkers, has been reported to be associated with increased lung cancer risk in prospective studies (Melbourne Collaborative Cohort Study/MCCS study [16]; UK Biobank/UKBB cohort [18, 19]).

However, the proportion of variance in DNAm clocks explained by smoking remains to be elucidated [20, 21], as well as the extent to which DNAm clocks are associated with lung cancer risk. Identifying which components of these clocks are linked to tobacco smoking is crucial for understanding the relationship between aging processes and lung cancer risk. $PLCO_{m2012}$ primarily incorporates smoking-related metrics and other factors indirectly capturing smoking exposure [4]. It is not known whether DNAm clocks provide additional information on lung cancer risk. If they do, a key question is whether this improvement stems from smoking-related signals embedded in the clocks or from other biological features they capture, independently of smoking. Addressing this will help clarify both the added value of DNAm clocks in lung cancer risk prediction and the broader interplay between smoking, epigenetic aging, and lung cancer development.

In this study, we aimed to evaluate the relationship between biological age, as estimated by DNAm clocks, tobacco smoking, and lung cancer risk using pre-diagnostic methylation data from four case–control studies nested in prospective cohorts.

Methods

Data source

To assess whether DNAm clocks are associated with lung cancer risk, we used four pre-diagnostic datasets with DNA methylation data from previous studies, including

the Melbourne Collaborative Cohort Study (MCCS), the Northern Sweden Health and Disease Study (NSHDS), the European Investigation into Cancer and Nutrition (EPIC-Italy), and the Norwegian Women and Cancer (NOWAC) [22]. DNA methylation profiling and data preprocessing protocols for the MCCS, NSHDS, EPIC-Italy, and NOWAC cohorts have been previously described [23]. Briefly, genome-wide methylation profiling was conducted using pre-diagnostic blood specimens analyzed with the Illumina Infinium Human Methylation 450 K BeadChip array. We grouped cohorts into tobacco smoking matched (MCCS-Australia and NSHDS-Sweden) and not matched on tobacco smoking (NOWAC-Norway and EPIC-Italy) sets for the analyses, as described previously [24]. Briefly, in the MCCS and NSHDS cohorts, each case was matched to one control based on sex, age, blood collection date, and smoking variables. In EPIC-Italy, matching was done by sex, age, study centre, and enrolment date. For the NOWAC cohort, controls were matched to cases based on year of birth and time since blood collection. The smoking matching variables assessed at blood draw included whether individuals quit smoking less than 10 years prior and whether they smoked ≥ 15 or < 15 cigarettes per day. For all cohorts included, participants were cancer-free at enrolment and lung cancer cases were defined as all invasive cancers coded C34.0 to C34.9 in the International Classification of Diseases for Oncology, Third Edition. [25].

DNAm clocks and DNAm-based smoking scores

We analyzed several classes of DNAm clocks that differ in their training targets and interpretation. Chronological age clocks (Horvath [12], Hannum [26], and Zhang [27]) output predicted age in years. Mitotic clocks (DNAmTL for telomere length [28] and epiTOC2 for cell division rate estimators [14]) capture cell replication dynamics. All-cause mortality risk clocks includes PhenoAge [29] (predicted age in years), DunedinPACE [30] (a measure of the pace of aging, scaled so that 1.0 reflects 1 year of biological aging per chronological year), and principal component-denoised version of GrimAge/PCGrimAge [31] (originates as a log hazard ratio of mortality risk, which is then linearly rescaled to resemble chronological age in years for intuitive interpretation). More recently, causality-enriched clocks [32] based on epigenome-wide Mendelian randomization analyses have been proposed, including DamAge (detrimental effects of aging in years) and AdaptAge (beneficial effects against aging or healthy aging in years). We used the PCGrimAge clock instead of GrimAge2 because calculating GrimAge2 requires uploading DNA methylation data to an external web tool (https://dnamage.clockfoundation.org), and we did not have authorization from the individual studies to share these data externally. The DNAm clocks were obtained using the R package methylCIPHER (v0.2.0; https://github.com/MorganLevineLab/methylCIPHER), except for PCGrimAge, which was calculated using the code provided by the original study (https://github.com/MorganLevineLab/PC-Clocks) [31]. This code additionally provides the values of subcomponents of PCGrimAge, also estimated using DNA methylation data, including packyears and blood circulating protein levels (TIMP1: TIMP Metallopeptidase Inhibitor 1; GDF15: Growth Differentiation Factor 15; ADM: Adrenomedullin; B2M: Beta-2-Microglobulin; PAI1/ SERPINE1: Serpin Family E Member 1; Cystatin C; Leptin). A deconvolution method (Houseman algorithm) applied to DNAm data was used to infer the proportion of immune blood cells (monocytes, natural killer/NK, B, CD4 and CD8 T cells) in all studies [22, 33]. Residuals from the regression of DNAm clocks on chronological age were used in the subsequent analyses (age-adjusted DNAm clocks).

Three DNAm-based smoking scores were calculated to quantify smoking exposure. The McCartney score [34] was computed as a weighted sum of methylation β-values across 233 CpG sites previously associated with current smoking status. The Elliott score [35], representing cumulative smoking exposure (packyears), was derived using the Smoking Score (SSc) method implemented in the EpiSmokEr R [36] (v0.1.0) package, based on 187 CpG sites. The Joehanes score [10] was based on 11,267 CpG sites associated with current smoking status and demonstrated a dose–response relationship with packyears.

Statistical analysis

Pearson correlation and variance explained in DNAm clocks

Pearson correlations between age-adjusted DNAm clocks were calculated for participants with smoking history. Adjusted R^2, derived from linear regression models, was used to estimate the percentage of cumulative variance in the age-adjusted DNAm clocks explained by the available set of traits in sequential order: (1) sex (male vs. female); (2) smoking status (current vs. former smokers); (3) detailed smoking information (smoking duration, time of quitting smoking, and cigarettes per day) by observational studies. The analyses were conducted for controls and lung cancer cases separately.

Relative risk of lung cancer and DNAm clocks

To estimate odds ratios (OR) for lung cancer risks per 1 standard deviation (s.d) increased in age-adjusted DNAm clocks, we fitted a series of unconditional logistic regression models across the four combined cohorts (overall), the tobacco smoking matched (MCCS-Australia and NSHDS-Sweden) and not matched on smoking (NOWAC-Norway and EPIC-Italy) sets. First, we fitted

a model minimally adjusted for age, sex, and cohort. Second, we fitted a tobacco smoking model, additionally adjusted for detailed smoking information (duration of smoking, cigarettes per day, and years since quit smoking). Finally, we additionally adjusted for estimated proportion of immune cells to account for potential confounders (e.g. heterogeneity of blood cell types), and body mass index as covariates, as previously reported [16]. In the stratified analyses, cut-offs for tobacco smoking intensity and duration, and packyears were calculated based on the mean values of these variables among the cohorts not matched on tobacco smoking.

Lung cancer discrimination analysis

We used receiver operating characteristic (ROC) curves to evaluate whether age-adjusted DNAm clocks and $PLCO_{m2012}$ alone or combined improved lung cancer discrimination between individuals with smoking history. We fitted unconditional logistic regression models on the smoking matched datasets (MCCS-Australia and NSHDS-Sweden) and the ones not matched on smoking (NOWAC-Norway and EPIC-Italy) separately. We calculated the area under the curve (AUC) using three models, adjusting for age and sex, and cohort: (a) age-adjusted DNAm clock alone (AUC_{Clock}); (b) $PLCO_{m2012}$ model alone (AUC_{PLCO}); and (c) a combined age-adjusted DNAm clock and $PLCO_{m2012}$ model ($AUC_{Combined}$).

All statistical analyses were conducted using R (v4.2.1). The ROC curves were plotted using the R package pROC (v1.18.5) and their statistical differences ($P_{difference}$) were estimated by DeLong's Z-test using the roc.test function of the same package.

Results

Baseline characteristics of studies

The characteristics of the participants included in this study are shown in Table 1. We included 1480 participants with a history of smoking in the tobacco smoking matched (514 case–control pairs from MCCS-Australia and NSHDS-Sweden) and not matched on smoking (275 cases and 177 controls from EPIC-Italy and NOWAC-Norway) sets. The cases and controls in the set not matched on smoking were predominantly female (64% and 61%, respectively, vs. 39% in both groups of the

Table 1 Characteristics of the participants included in the studies at baseline

Characteristics	Smoking-matched set		Smoking-unmatched set	
	Lung Cancer (N = 514)	Control (N = 514)	Lung Cancer (N = 275)	Control (N = 177)
Sex. N (%)				
Female	202 (39%)	202 (39%)	176 (64%)	108 (61%)
Male	312 (61%)	312 (61%)	99 (36%)	69 (39%)
Age. Mean (SD)	57.7 (7.1)	57.6 (7.1)	55.4 (5.6)	55.0 (5.4)
Cohort. N (%)				
MCCS	324 (63%)	324 (63%)		
NSHDS	190 (37%)	190 (37%)		
EPIC-Italy			160 (58%)	107 (60%)
NOWAC			115 (42%)	70 (40%)
Smoking status. N (%)				
Current	287 (56%)	287 (56%)	182 (66%)	78 (44%)
Former	227 (44%)	227 (44%)	93 (34%)	99 (56%)
Years smoked. Mean (SD)	35.9 (10.3)	33.6 (12.0)	33.4 (10.2)	27.6 (11.2)
Average cigarettes smoked per day. Mean (SD)	22.0 (13.6)	19.4 (12.8)	14.6 (7.2)	11.6 (7.4)
Quit years. Mean (SD)	4.3 (7.4)	5.7 (9.7)	3.6 (7.1)	8.1 (10.2)
Mean Pre-diagnosis lead time (SD)	9.3 (4.9)		5.7 (3.5)	
3-year	71 (14%)		67 (24%)	
5-year	50 (9.7%)		63 (23%)	
≥ 5-years	393 (76%)		145 (53%)	

smoking matched set) and had a lower mean age at baseline (mean (SD) = 55.4 years [5.6] and 55.0 years [5.4] for cases and controls, respectively, compared to 57.7 years [7.1] and 57.6 years [7.1] in the set not matched on smoking). Regarding smoking status, the set not matched on smoking had a higher proportion of current smokers among cases at baseline (66% vs. 56% in the smoking matched set) but a lower proportion among controls (44% vs. 56%). Additionally, participants in the set not matched on smoking reported fewer years of smoking (mean (SD) = 33.4 [10.2] years for cases vs. 27.6 [11.2] years for controls) compared to those in the smoking matched set (mean (SD) = 35.9 [10.3] vs. 33.6 [12.0] years, respectively). Daily cigarette consumption was also lower in the set not matched on smoking (mean (SD) = 14.6 [7.2] vs. 11.6 [7.4] cigarettes/day for cases and controls) than in the smoking matched set (mean (SD) = 22.0 [13.6] vs. 19.4 [12.8] cigarettes/day). Among lung cancer cases, the smoking matched set had a longer mean pre-diagnosis lead time—the interval between blood draw and cancer diagnosis—compared to the set not matched on smoking (mean (SD) = 9.3 [4.9] vs. 5.7 [3.5] years). In both sets, most lung cancer cases were diagnosed more than 5 years after blood collection (76% and 53%, respectively).

Overview of the DNAm clocks

Chronological age was consistently associated with unadjusted DNAm clocks in both controls and lung cancer cases (Additional file 1: Table S1). Pairwise correlations between age-adjusted DNAm clocks were highly consistent across four cohorts, whether matched or unmatched for smoking. For instance, we observed a strong positive correlation between PCGrimAge vs. DunedinPACE ($r = 0.67$, SD = 0.06) and an inverse correlation between AdaptAge vs. DamAge ($r = -0.50$, SD = 0.05) (Additional file 1: Fig. S1). We evaluated the extent to which sex and smoking behaviours explained variability in age-adjusted DNAm clocks in blood samples of individuals with smoking history and lung cancer (Fig. 1A, Additional file 1: Material S1). Sex alone accounted for < 1% of variance in Zhang (2019), PhenoAge, epiTOC2, DunedinPACE, DamAge, and AdaptAge, and 5–6% in PCGrimAge and DNAmTL. Adding smoking status (current vs. former) increased explained

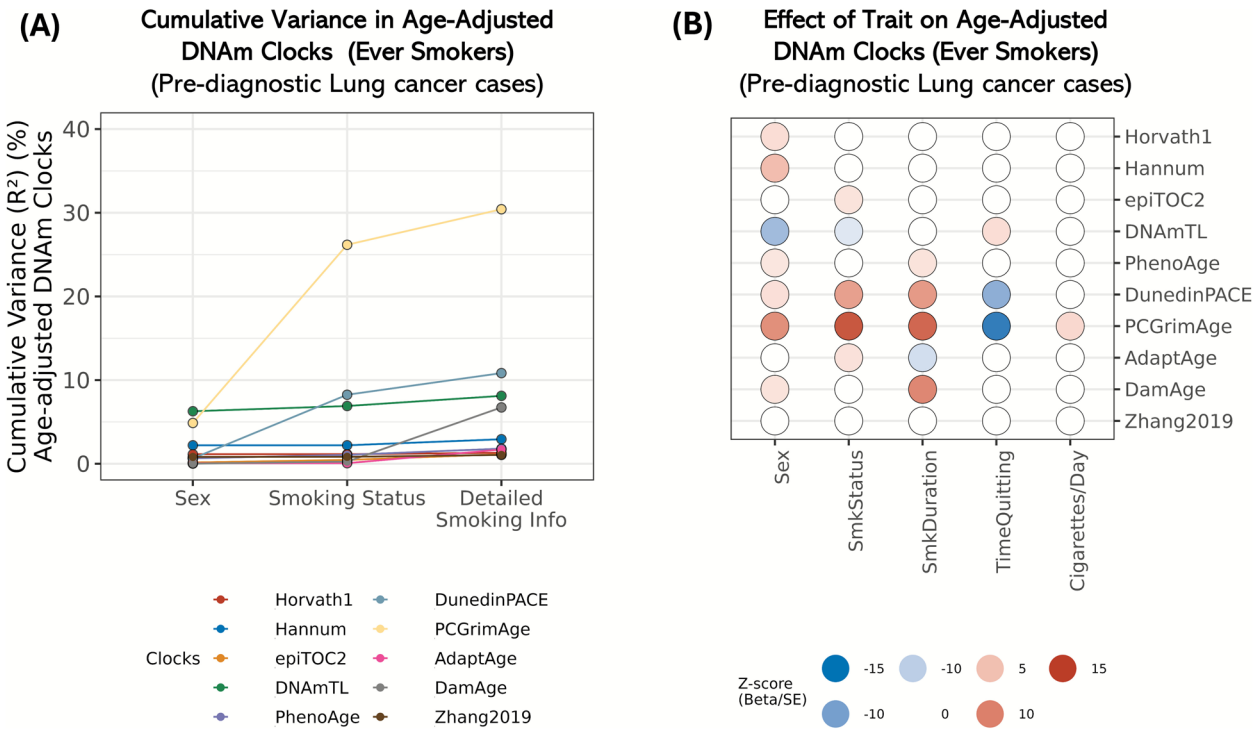

Fig. 1 Relationship between DNAm clocks, epidemiological and molecular traits among individuals with tobacco smoking history and lung cancer. **A** Multivariate logistic regression models to estimate the proportion of cumulative variance (R^2) in age-adjusted DNAm clocks explained by a sequential ordered trait: (1) sex (male vs. female/REF); (2) smoking status (current vs. former smokers/REF); (3) detailed smoking info (cigarettes/day, smoke duration, and time quitting smoking). **B** Heatmaps representing Z-scores (beta divided by standard errors) of the associations between age-adjusted DNAm clocks (outcome) and above variables (predictors). Linear regression models included age, sex (whenever applicable), and study name as covariates. Z-score > 0 (shades of red) and Z-score < 0 (shades of blue). Associations with $p < 0.05$ were considered statistically significant and coloured while the ones with $p > 0.05$ were omitted (empty cells)

variance in DunedinPACE to ~ 8% and PCGrimAge to 26%. Incorporating duration of smoking, cigarettes per day, and years since quitting smoking further explained variance, with the largest contribution for PCGrimAge clock ($R^2 = 30\%$) and the lowest for Zhang (2019), Horvath, and epiTOC2 ($R^2 = \sim 1\%$).

Among individuals with incident lung cancer and smoking history, males tended to have higher values for most age-adjusted DNAm clocks, particularly PCGrimAge ($\beta_{males} = 0.63 \pm 0.07$), whereas females exhibited higher values of DNAmTL ($\beta_{males} = -0.52 \pm 0.07$) (Fig. 1B, Additional file 1: Material S2). DunedinPACE and PCGrimAge increased with current smoking (DunedinPACE: $\beta_{current} = 0.52 \pm 0.07$; PCGrimAge: $\beta_{current} = 0.82 \pm 0.06$); smoking duration (DunedinPACE: $\beta = 0.27 \pm 0.03$; PCGrimAge: $\beta = 0.37 \pm 0.03$), and packyears (DunedinPACE: $\beta = 0.13 \pm 0.04$; PCGrimAge: $\beta = 0.22 \pm 0.04$) (Fig. 1B, Additional file 1: Material S2). Additional associations with smoking duration were observed for DamAge ($\beta = 0.10 \pm 0.01$) and AdaptAge ($\beta = -0.03 \pm 0.01$). Similar results were observed for blood samples of individuals with smoking history and without lung cancer (Additional file 1: Fig. S2, Material S2).

DNAm clocks and lung cancer risk

In the minimally-adjusted model (adjusted for age, sex, and cohort), higher age-adjusted DNAm clocks were associated with increased lung cancer risk in the combined four studies (Fig. S3, Table S2). The strongest associations were observed for DunedinPACE (OR = 1.44, 95% CI = 1.29–1.62) and PCGrimAge (OR = 1.79, 95% CI = 1.56–2.06), with more modest but significant associations for PhenoAge (OR = 1.25, 95% CI = 1.12–1.40) and Hannum (OR = 1.13, 95% CI = 1.02–1.26) (Fig. S3, Table S2). After additionally accounting for detailed smoking metrics (smoking adjusted model), the effect of DunedinPACE (OR = 1.38, 95% CI = 1.23–1.55) on lung cancer risk was attenuated by 12% and by 13% for PCGrimAge (OR = 1.66, 95% CI = 1.44–1.91) (Additional file 1: Fig. S3, Table S2). Further adjustment for DNAm-inferred immune cell proportions ("fully adjusted" model) strengthened the PCGrimAge association (OR = 1.89, 95% CI = 1.59–2.27) (Fig. 2, Additional file 1: Table S2).

Comparing smoking matched (MCCS + NSHDS) vs. unmatched (EPIC-Italy + NOWAC) sets revealed modest heterogeneity in clock-risk associations (Additional file 1: Table S2, Table S3), likely reflecting differences in the design of these cohorts by matching on smoking. In

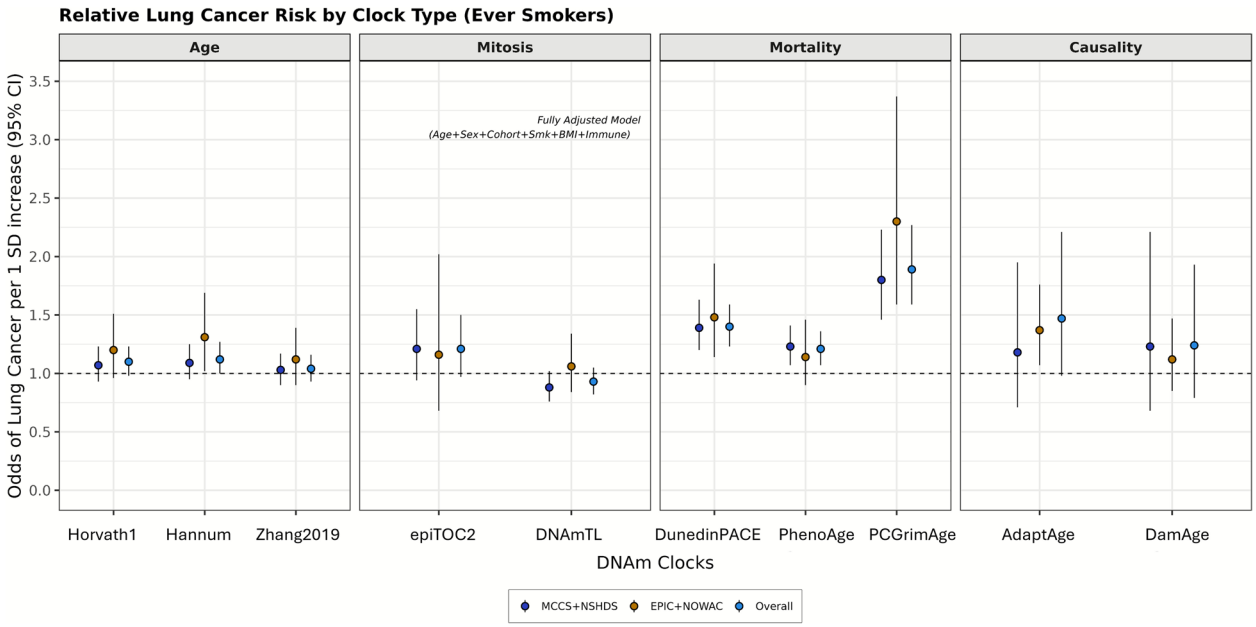

Fig. 2 Association between DNAm clocks and lung cancer risk. Odds ratios of lung cancer risk per 1 unit s.d increase in age-adjusted DNA clocks in MCCS + NSHDS (smoking-matched; dark blue dots), EPIC-Italy + NOWAC (smoking unmatched; orange dots), and overall studies (light blue dots) among individuals with smoking history. Clocks were originally trained on chronological age (Horvath, Hannum, and Zhang2019); cellular mitosis (epiTOC2 and DNAmTL); and mortality risk (PhenoAge, DunedinPACE, and PCGrimAge). Additional causality-enriched clocks were included (DamAge and AdaptAge). Logistic regression models included age, sex, cohort/study, smoking status, cigarettes per day, smoke duration, time of quitting smoking, body mass index (BMI), and proportions of immune cells inferred by DNA methylation data (monocytes, natural killers, CD4T, CD8T, and B cells) as covariates. Models include matched variables: age, sex, matching variables. Detailed information about estimates, including OR [95%CI], *p*-values, and AUC by cohorts and overall are in the Additional file 1: Table S2

the minimally-adjusted models, association effects of clocks on lung cancer risk in the unmatched set were 20% higher for DunedinPACE, 28% for Horvath, 31% for Hannum, and 40% for PCGrimAge ($OR_{matched}$ = 1.65, 95% CI = 1.41–1.93 vs. $OR_{unmatched}$ = 2.31, 95% CI = 1.71–3.16; $P_{Heterogeneity}$ = 0.055) than the matched set. After accounting for detailed smoking metrics and DNAm-inferred immune cell proportions, this between set heterogeneity was substantially attenuated (Additional file 1: Table S3). Overall, the OR and AUC estimates for DNAm clocks were largely consistent between the minimally and fully adjusted models, suggesting that our results were less likely to be influenced by cell type heterogeneity and potential batch effects across plates and studies. For this reason, most subsequent analyses were performed using the minimally adjusted model.

The association effects of some DNAm clocks on lung cancer risk differed by age, sex, and smoking status strata (Additional file 1: Table S4). Relatively higher lung cancer risk was observed for older individuals with DunedinPACE (≥ 60 years: OR = 3.65, 95% CI = 2.07–7.12 vs. < 60 years: OR = 1.66, 95% CI = 1.32–2.13; $P_{Heterogeneity}$ = 0.02) and DamAge (≥ 60 years: OR = 2.09, 95% CI = 1.26–3.73 vs. < 60 years: OR = 1.01, 95% CI = 0.81–1.26; $P_{Heterogeneity}$ = 0.01), being male with PhenoAge (OR = 1.69, 95% CI = 1.21–2.44 vs. Female: OR = 1.10, 95% CI = 0.86–1.41; $P_{Heterogeneity}$ = 0.050), and current smoker with epiTOC2 (OR = 1.87, 95% CI = 1.28–2.83 vs. former smoker: OR = 0.83, 95% CI = 0.59–1.16; $P_{Heterogeneity}$ = 0.002).

Discriminatory performance of lung cancer risk models

We used AUC and ROC curves to evaluate whether the DNAm clocks captured additional information about lung cancer risk beyond the $PLCO_{m2012}$ model in discriminating individuals with lung cancer by fitting a model combining DNAm clocks and $PLCO_{m2012}$.The AUC for $PLCO_{m2012}$ model was 0.66 (95% CI = 0.61–0.71) in the set not matched on smoking (EPIC-Italy + NOWAC) (Fig. 3). Among the four DNAm clocks associated with lung cancer risk (Hannum, PhenoAge, DunedinPACE, and PCGrimAge), PCGrimAge clock alone (AUC = 0.72, 95% CI = 0.67–0.77) slightly outperformed $PLCO_{m2012}$ (AUC_{Clock} vs. AUC_{PLCO}, $P_{difference}$ = 0.03), while the other DNAm clocks (Hannum, PhenoAge, and DunedinPACE) showed similar or lower AUC values when compared to $PLCO_{m2012}$ model ($P_{difference}$ > 0.05). Combining the PCGrimAge and $PLCO_{m2012}$ yielded similar risk discriminatory performance as the PCGrimAge alone (AUC = 0.72, 95% CI = 0.67–0.77) (Fig. 3). Similar AUC estimates were noted for the PCGrimAge model alone or combined with $PLCO_{m2012}$ models between age, sex, smoking status, years of blood sample collection, eligibility criteria for LDCT by USPSTF 2021 and $PLCO_{m2012}$, and cohort strata (Additional file 1: Table S5). Additionally, no association between pre-diagnosis lead time and PCGrimAge values were observed among lung cancer cases (Additional file 1: Fig. S4). Comparable results were observed in the smoking matched set (MCCS + NSHDS), with the PCGrimAge providing higher discrimination than the $PLCO_{m2012}$ model ($P_{difference}$ = 0.03; Additional file 1: Fig. S5). Sensitivity analyses using the values of PCGrimAge unadjusted by chronological age showed identical AUCs for lung cancer discrimination in both smoking matched and unmatched sets (Additional file 1: Table S6).

Secondary analysis of PCGrimAge components

To explore which PCGrimAge components provided information on lung cancer risk beyond the $PLCO_{m2012}$ model to discriminate lung cancer cases, we conducted multivariate regression models for each DNAm-inferred component using the set not matched on smoking (Table 2). After accounting for $PLCO_{m2012}$ score and covariates (age, sex, and cohort), several PCGrimAge components remained significantly associated with lung cancer risk, including packyears (OR = 2.53, 95% CI = 1.93–3.30); cystatin C (OR = 2.45, 95% CI = 1.53–3.95); GDF15 (OR = 2.53, 95% CI = 1.59–4.03); PAI1 (OR = 1.48, 95% CI = 1.13–1.94); and TIMP1 (OR = 2.04, 95% CI = 1.09–3.81). In a combined model including all PCGrimAge components, DNAm-based packyears remained an independent predictor of lung cancer risk ($OR_{DNAm\text{-}packyears}$: 2.23, 95% CI = 1.58–3.14), after adjustments for self-reported packyears (Table 2). Other DNAm-based smoking scores—trained to predict either packyears or smoking status—showed comparable associations and discriminatory performance to DNAm-based packyears included in the PCGrimAge (Additional file 1: Table S7).

To evaluate the characteristics of research participants deemed at high risk based on the PCGrimAge and $PLCO_{m2012}$ model, we identified the PCGrimAge cut-off (72.6 years) that selected the same number of controls (n = 148) as a 1% cut-off based on the $PLCO_{m2012}$ model (Table 3). Based on these cut-offs, the $PLCO_{m2012}$ model identified 210 cases in our total study sample and PCGrimAge identified 248 cases. We found that these groups of study participants were largely similar, but cases and controls selected based on PCGrimAge reported less smoking exposure than those selected based on the $PLCO_{m2012}$ model (p-value < 0.001) (Table 3).

Discussion

In this study, we investigated the relationship between DNAm clocks, their components, and lung cancer risk among individuals with a history of tobacco smoking, using pre-diagnostic blood samples from four prospective

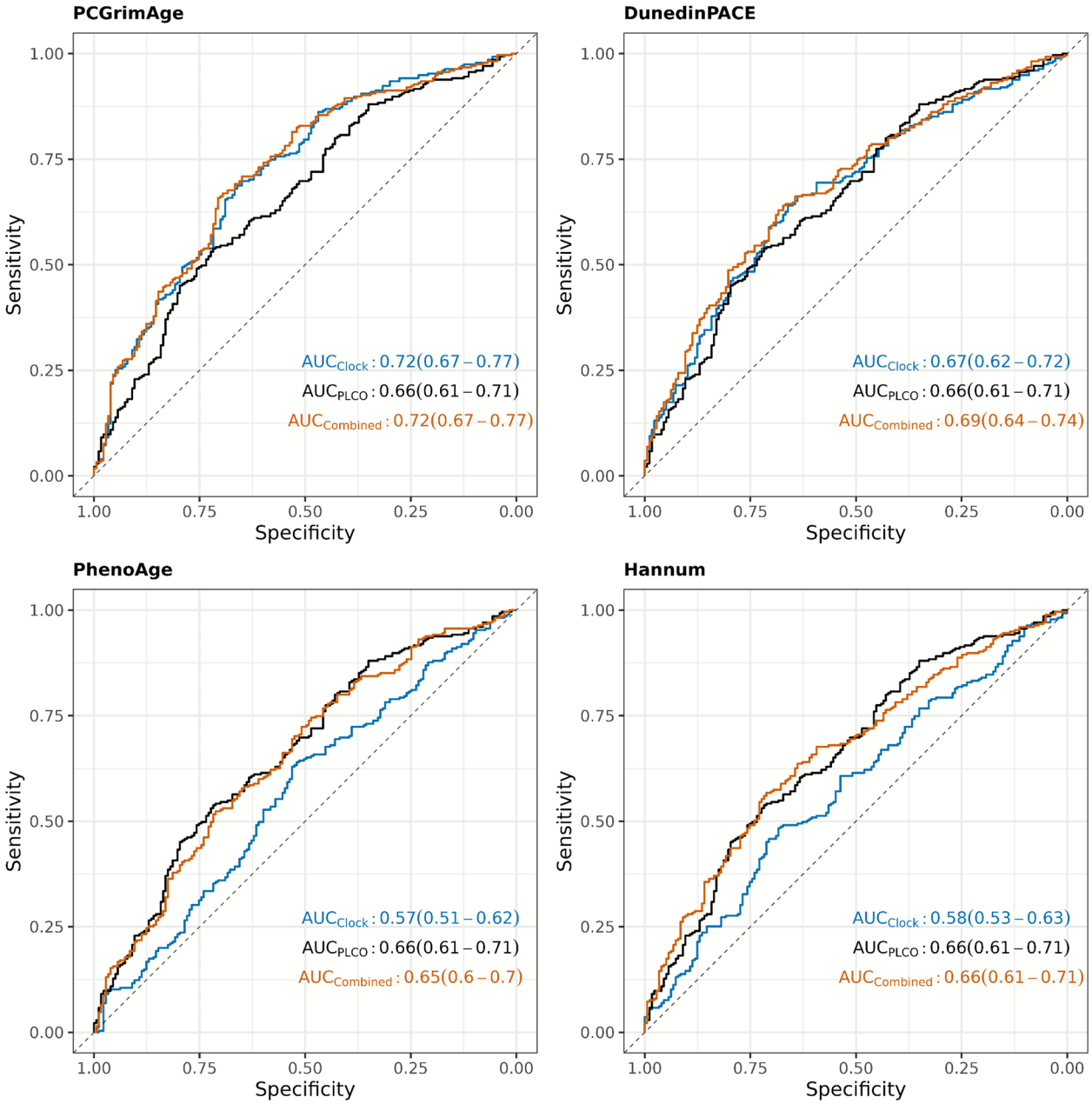

Fig. 3 Risk-discriminatory performance of lung cancer models using ROC curves in the smoking unmatched set. The three models developed on the smoking unmatched set (NOWAC and EPIC-Italy studies) were (1) age-adjusted DNAm clock (AUC$_{Clock}$; coloured in blue); (2) PLCO$_{m2012}$ (AUC$_{PLCO}$; coloured in black), adjusting by the matching variables (age and sex) and study name; (3) a combination of age-adjusted DNAm clock and PLCO$_{m2012}$ (AUC$_{Combined}$; orange), adjusting by the matching variables (age and sex) and study name. Hannum, PhenoAge, DunedinPACE, and PCGrimAge clocks were analyzed. ROC, receiver operating characteristic. *X*-axis represents specificity and *Y*-axis sensitivity

cohorts. We assessed whether smoking behaviour influenced these associations and evaluated the ability of DNAm clocks to improve lung cancer risk prediction beyond the established PLCO$_{m2012}$ model. Across both smoking matched and unmatched sets, we found that DNAm clocks trained on mortality risk—particularly PCGrimAge—offered additional predictive value for lung cancer risk to that from the well-established PLCO$_{m2012}$ model.

The identification of biomarkers for early detection of lung cancer could assist in selecting high-risk individuals for lung cancer screening with LDCT [3]. Our study reinforces the link between lung cancer risk and age-adjusted DNAm clocks, with PCGrimAge emerging

Table 2 Association between PCGrimAge components and $PLCO_{m2012}$ score on lung cancer risk

Predictor	Minimally-adjusted model					$PLCO_{m2012}$-adjusted model					Mutually-adjusted model				
	OR	L95	U95	P	AUC [95%CI]	OR	L95	U95	P	AUC [95%CI]	OR	L95	U95	P	AUC [95%CI]
$PLCO_{m2012}$	1.58	1.26	2.00	1×10^{-04}	0.66 [0.61-0.71]						0.95	0.89	1.03	0.213	
PCGrimAge	2.53	1.97	3.25	4×10^{-13}	0.72 [0.67-0.77]	2.37	1.82	3.10	2×10^{-10}	0.72 [0.67-0.77]					
Age	1.08	0.89	1.31	0.460	0.52 [0.47-0.58]	0.99	0.80	1.22	0.915	0.66 [0.61-0.71]	1.22	0.48	3.12	0.682	
Sex	0.86	0.51	1.45	0.575	0.52 [0.47-0.58]	0.65	0.37	1.13	0.129	0.66 [0.61-0.71]	0.06	0.00	1.99	0.117	
Packyears (Epi)	2.26	1.74	2.92	7×10^{-10}	0.69 [0.64-0.74]	2.23	1.61	3.07	1×10^{-06}	0.69 [0.64-0.74]	1.67	1.21	2.31	0.002	
Packyears (DNAm)	2.61	2.04	3.34	2×10^{-14}	0.74 [0.69-0.78]	2.53	1.93	3.30	1×10^{-11}	0.74 [0.69-0.78]	2.23	1.58	3.14	5×10^{-06}	
ADM (DNAm)	1.47	1.02	2.12	0.040	0.56 [0.51-0.62]	1.39	0.95	2.01	0.087	0.65 [0.60-0.70]	1.10	0.53	2.26	0.802	0.75 [0.71-0.80]
B2M (DNAm)	1.58	1.00	2.49	0.048	0.56 [0.50-0.61]	1.55	0.98	2.46	0.062	0.66 [0.61-0.71]	1.14	0.59	2.22	0.701	
Cystatin C (DNAm)	2.77	1.73	4.43	2×10^{-05}	0.62 [0.56-0.67]	2.45	1.53	3.95	2×10^{-04}	0.67 [0.62-0.72]	2.06	0.83	5.15	0.121	
Leptin (DNAm)	0.64	0.19	2.19	0.478	0.53 [0.48-0.59]	0.65	0.19	2.27	0.502	0.66 [0.61-0.71]	0.42	0.07	2.63	0.353	
GDF15 (DNAm)	2.97	1.88	4.68	3×10^{-06}	0.63 [0.58-0.69]	2.53	1.59	4.03	9×10^{-05}	0.67 [0.62-0.72]	0.59	0.24	1.40	0.229	
PAI1 (DNAm)	1.57	1.20	2.06	0.001	0.59 [0.54-0.65]	1.48	1.13	1.94	0.005	0.65 [0.60-0.70]	1.32	0.86	2.04	0.201	
TIMP1 (DNAm)	2.26	1.23	4.17	0.009	0.57 [0.51-0.62]	2.04	1.09	3.81	0.026	0.65 [0.60-0.70]	0.52	0.13	2.12	0.362	

Unconditional logistic regression models using tobacco smoking unmatched studies (NOWAC and EPIC) (275 lung cancer cases, 177 controls). Outcome: Lung Cancer. Minimally-adjusted model adjusted by age, sex, and cohort. $PLCO_{m2012}$-adjusted model additionally adjusted by $PLCO_{m2012}$ logit score. Multivariate Mutually-adjusted model included adjustment between variables, excluding PCGrimAge clock due to collinearity. DNAm (inferred by DNA methylation data). Epi (epidemiological measured data). Area under the curve derived from ROC curves for unconditional logistic regression models. Odds ratio (OR) scales as 1 unit s.d increase PCGrimAge: Principal component-denoised version of GrimAge by Higgins-Chen, et al. [31]

Table 3 Characteristics of lung cancer cases and controls by $PLCO_{m2012}$ and PCGrimAge eligibility screening thresholds

Characteristics	Lung Cancer			Control		
	PCGrimAge (≥72.6y) (N = 248)*	$PLCO_{m2012}$ (≥1%) (N = 210)	P-value**	PCGrimAge (≥72.6y) (N = 148)*	$PLCO_{m2012}$ (≥1%) (N = 148)*	P-value**
Age. Mean (SD)	60.3 ± 3.8	60.6 ± 3.8	0.658	60.9 ± 3.8	61.0 ± 3.7	0.790
Sex. N (%)						
Female	108 (43.5%)	82 (39.0%)	0.343	52 (35.1%)	54 (36.5%)	0.904
Male	140 (56.5%)	128 (61.0%)		96 (64.9%)	94 (63.5%)	
Smoking status. N (%)						
Current	226 (91.1%)	193 (91.9%)	0.867	137 (92.6%)	136 (91.9%)	1.000
Former	22 (8.9%)	17 (8.1%)		11 (7.4%)	12 (8.1%)	
Years smoked. Mean (SD)	41.6 ± 6.8	43.0 ± 5.4	0.091	42.4 ± 7.6	43.5 ± 5.9	0.350
Average cigarettes smoked per day. Mean (SD)	17.8 ± 9.4	23.6 ± 9.7	8×10^{-13}	19.5 ± 11.3	23.7 ± 10.1	4×10^{-06}
Packyears. Mean (SD)	36.2 ± 23.0	49.3 ± 22.7	2×10^{-12}	42.1 ± 29.7	51.5 ± 26.8	5×10^{-05}
Smoking quit years. Mean (SD)	0.7 ± 2.7	0.4 ± 1.6	0.701	0.8 ± 3.6	0.6 ± 2.5	0.856
Pre-diagnosis lead time. N (%)						
< 5 years	156 (62.9%)	140 (66.7%)	0.433			
≥ 5 years	92 (37.1%)	70 (33.3%)				

*Mean ± SD; n (%); **Kruskal-Wallis rank sum test; Fisher's exact test

as the most informative predictor among the analyzed DNAm clocks. These findings are consistent with previous studies, such as those using the MCCS cohort, which also reported associations between GrimAge and lung cancer risk [16]. Importantly, we observed no relationship between PCGrimAge and pre-diagnosis lead time, and the strength of its association with lung cancer was similar among cases diagnosed within 3 years of blood draw and those diagnosed later. These findings suggest that PCGrimAge does not accelerate as diagnosis approaches, reinforcing its potential as a biomarker for early detection of lung cancer.

The added value of PCGrimAge to the $PLCO_{m2012}$ risk model appears to be largely driven by its smoking-related

components, especially DNAm-inferred packyears. This component remained an independent predictor of lung cancer after adjustment for both $PLCO_{m2012}$ scores and self-reported packyears. We further observed similar associations using other independent DNAm-based smoking scores [10, 34, 35], reinforcing the relevance of DNAm marks of smoking in lung cancer risk stratification.

Whereas we found that the association between some of the DNAm clocks and lung cancer risk was only partly attenuated after accounting for detailed smoking information, from an etiological perspective, we believe that it is unlikely that the DNAm clocks exert a causal influence on lung cancer development. Rather, these observed associations can likely be explained by residual confounding by smoking, as well as by capturing facets of tobacco exposure not easily assessed through questionnaires. This interpretation is in line with prior Mendelian randomization analyses that have not found evidence in support a causal role of DNAm clocks in lung cancer aetiology [37].

Self-reported smoking data, which is integral to $PLCO_{m2012}$ risk model, is often subject to recall bias and misreporting [38]. In contrast, PCGrimAge incorporates molecular markers that reflect cumulative and biologically relevant tobacco smoke exposure, and thus PCGrimAge may serve as a valuable complementary tool for assessing lung cancer risk. Indeed, smoking explained 30% of the variance in PCGrimAge in our study, a finding consistent with a previous study among Costa Rican adults [20]. Interestingly, while high-risk individuals identified by either PCGrimAge or $PLCO_{m2012}$ scores had similar characteristics, those classified by PCGrimAge included relatively fewer heavy smokers (≥ 20 packyears). Additionally, daily cigarette consumption was similar across PCGrimAge-defined groups. These findings suggest that PCGrimAge may help identify lung cancer cases among light-to-moderate smokers (<20 packyears) who might be missed by $PLCO_{m2012}$ risk prediction model.

Furthermore, we also found an association between the mitotic clock epiTOC2 and lung cancer risk among current smokers, consistent with prior studies reporting increased tick rates of this mitotic clock in normal buccal tissues of smokers [39]. This observation suggests that mitotic-driven epigenetic alterations may contribute to lung carcinogenesis in active smokers, warranting further investigation.

In contrast to our previous work, where we evaluated a five-CpG DNA methylation-based risk score trained on lung cancer risk using the same dataset with similar endpoints [24], this current study investigated established DNAm clocks in the context of lung cancer prediction. Our previous findings showed that the five-CpG DNA methylation-based risk score provided similar discriminatory performance to $PLCO_{m2012}$, with no improvement when combined [24]. In the current study, we found that PCGrimAge provided additional risk discrimination compared to the $PLCO_{m2012}$ model. This suggests that comprehensive epigenetic aging markers trained on tobacco smoking intensity captures information relevant to lung cancer risk assessment. Importantly, DNAm clocks capture more than just smoking exposure. Prior studies have shown that GrimAge acceleration was associated with lung cancer risk among individuals without smoking history [40], underscoring the broader relevance of this clock as biomarker for early detection of lung cancer, independent of smoking.

Despite these promising findings, our study has some limitations. The lack of racial and ethnic diversity in our cohorts, with most participants being of European ancestry, limits the generalizability of our results. Previous studies have shown that biological aging markers, including GrimAge clock, may vary based on socioeconomic status and ancestry [41, 42]. Future studies incorporating diverse populations are needed to explore whether the relationship between DNAm clocks and lung cancer risk differs across demographic groups. Additionally, the matching variables in our case–control study design (age and sex, as well as smoking history in two cohorts) influenced the discrimination metrics which are not comparable to previous cohort studies of lung cancer risk models. Future studies evaluating the importance of DNA methylation markers and lung cancer risk should be conducted using a study design that is representative of the background study population.

Conclusions

In conclusion, our findings demonstrate that methylation-based biological clocks capture epigenetic marks left by exposure to tobacco smoke, and some may inform lung cancer risk assessment by complementing or replacing traditional prediction models. These findings highlight opportunities for the use of methylation-based biomarkers to identify high-risk individuals who may benefit from lung cancer screening with low-dose CT.

Abbreviations

AUC	Area under the curve
DNAm	DNA methylation
EPIC	European Investigation into Cancer and Nutrition
LDCT	Low-dose computed tomography
$PLCO_{m2012}$	Prostate, Lung, Colorectal, and Ovarian model 2012
MCCS	Melbourne Collaborative Cohort Study
NSHDS	Northern Sweden Health and Disease Study
NOWAC	Norwegian Women and Cancer
OR	Odds ratios
ROC	Receiver operating characteristic
s.d	Standard deviation
UKBB	UK Biobank
USPSTF	US Preventive Services Task Force

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-04542-9.

Additional file 1: Figures S1–S5, Materials S1–S2, and Tables S1–S7. Fig. S1. Pairwise correlations between DNAm clocks. Pearson correlation coefficients (R) between residuals of DNAm clocks adjusted by chronological age (adj) were represented by study (the Melbourne Collaborative Cohort Study/MCCS, the Northern Sweden Health and Disease Study/NSHDS, the European Investigation into Cancer and Nutrition/EPIC-Italy, and the Norwegian Women and Cancer/NOWAC). R > 0 (shades of red) and R < 0 (shades of blue). Correlations with $p < 0.05$ were considered statistically significant and coloured while the ones with $p > 0.05$ were omitted (empty cells). Fig. S2. Relationship between DNAm clocks, epidemiological and molecular traits among individuals with tobacco smoking history and without lung cancer. (A) Multivariate logistic regression models to estimate the proportion of cumulative variance (R^2) in age-adjusted DNAm clocks explained by a sequential ordered trait: 1) Sex (male vs. female/REF); 2) Smoking status (current vs. former smokers/REF); 3) Detailed smoking info (cigarettes/day, smoke duration, and time quitting smoking). (B) Heatmaps representing Z-scores (beta divided by standard errors) of the associations between age-adjusted DNAm clocks (outcome) and above variables (predictors). Linear regression models included age, sex (whenever applicable), and study name as covariates. Z-score > 0 (shades of red) and Z-score < 0 (shades of blue). Associations with $p < 0.05$ were considered statistically significant and coloured while the ones with $p > 0.05$ were omitted (empty cells). Fig. S3. Association between DNAm clocks and lung cancer risk. Odds ratios of lung cancer risk per 1 unit s.d increase in age-adjusted DNA clocks in MCCS + NSHDS (smoking-matched; red dots), EPIC-Italy + NOWAC (smoking unmatched; blue dots), and Overall studies (orange dots) among individuals with smoking history. Clocks were originally trained on chronological age (Horvath, Hannum, and Zhang2019), cellular mitosis (epiTOC2 and DNAmTL), and mortality risk (PhenoAge, DunedinPACE, and PCGrimAge). Additional causality-enriched clocks were included (DamAge and AdaptAge). Two logistic regression models were used: (A) Minimally adjusted (age, sex, and cohort/study); (B) Smoking adjusted that additionally included (smoking status, cigarettes per day, smoke duration, and time of quitting smoking) as covariates. The two models included matched variables: age, sex, matching variables. Detailed information about estimates, including OR [95%CI], *P*-values, and AUC by cohorts and overall are in the Table S1. Fig. S4. Relationship between the PCGrimAge clock and time to lung cancer diagnosis among individuals with smoking history. Panel A shows data from the set not matched on smoking (EPIC and NOWAC; $n = 275$) and B (MCCS and NSHDS; $n = 514$) for the smoking-matched set. The y-axis represents age-adjusted PCGrimAge values scaled per one standard deviation increase, while the x-axis denotes time to lung cancer diagnosis in years (calculated from blood draw to diagnosis). The regression line was derived using the geom_smooth function in R with the loess method, modeling time to diagnosis as the outcome and PCGrimAge as the predictor. Annotations in the upper-right corner display the regression coefficients (β), standard errors (SE), and *p*-values from linear models adjusted for sex and cohort. Panel B presents the same analysis in smoking-matched cohorts (MCCS and NSHDS; $n = 514$). Fig. S5. Risk-discriminatory performance of lung cancer models using ROC curves in the smoking matched set. The three models developed on the smoking unmatched set (MCCS and NSHDS studies) were: 1) DNAm clock (AUCClock; coloured in blue); 2) $PLCO_{m2012}$ (AUCPLCO; coloured in black); 3) a combination of DNAm clock and $PLCO_{m2012}$ (AUCCombined; orange), adjusting by the matching variables (age, and sex) and study name. Hannum, PhenoAge, DunedinPACE, and PCGrimAge clocks were analyzed. ROC, receiver operating characteristic. X-axis represents specificity and Y-axis sensitivity. Material S1. Source data for Fig. 1A and figure S2A. Material S2. Source data for Fig. 1B and figure S2B. Table S1. Correlation between DNAm clocks and chronological age. Table S2. Source data of Fig. 2 and Figure S4. Table S3. Heterogeneity test by cohorts. Table S4. Stratified analyses of lung cancer risk and DNAm clocks. Table S5. Risk-discriminatory performance estimated using area under the curve (AUC) in the smoking unmatched set. Table S6. Lung cancer discrimination of unadjusted PCGrimAge and PLCOm2012 risk prediction models. Table S7. DNA methylation smoking scores and $PLCO_{m2012}$ risk prediction models.

Acknowledgements
Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council (NHMRC) grants 209057, 396414, and 1074383 and by infrastructure provided by Cancer Council Victoria. The DNA methylation and lung cancer nested case control study was funded by NHMRC grant 1050198. Cases and their vital status were ascertained through the Victorian Cancer Registry.

WHO disclaimer
Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

Authors' contributions
Conception and design: RCCP and MJ Conceptualized and designed the study. Investigation Methodology: RCCP, JUO, and MJ developed the methodology. Data curation: RCCP, JUO, and MJ curated the data. Statistical analysis: RCCP and JUO conducted the statistical analysis. Provision of study resources: TMS, THN, PV, MJ, RLM, PAD, CR, MS, JM, MJ. Acquisition of funding: MJ acquired funding. Project management and administration: RCCP and JUO did project management and administration. Supervision: MJ provided supervision. Writing–original draft: RCCP, JUO, and MJ drafted the manuscript. Writing – review and editing: RCCP, JUO, RL, JM, and MJ, reviewed and edited the manuscript. Final approval of manuscript: All authors read and approved the final manuscript.

Authors' Twitter handles'
X: @cortezr1987; @tinauchojo; @Mattias31.

Funding
This work was supported by a Cancer Research UK programme grant (the Integrative Cancer Epidemiology Programme, C18281/A29019).

Data availability
Data are available upon reasonable request. The four case-control studies analysed are not publicly available for confidentiality reasons. However, the corresponding author may make the anonymized version available upon request.

Declarations

Ethics approval and consent to participate
The matched case-controls studies involved human participants, and they were approved by the International Agency for Research on Cancer (IARC) Ethics Committee (IEC) (No. 11–13). All research participants provided written, informed consent. Participants gave informed consent to participate in the study before taking part.

Consent for publication
All authors have agreed to the submission and publication of this research.

Competing interests
The authors declare no competing interests.

Author details
[1]Genomic Epidemiology Branch, International Agency for Research On Cancer, Lyon, France. [2]Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromso, Troms, Norway. [3]MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [4]Department of Radiation Sciences Oncology, Umeå University, Umea, Sweden. [5]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia. [6]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia. [7]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia. [8]Research and Academic Development, London School of Hygiene and Tropical Medicine, London, UK. [9]MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School , University of Bristol, Bristol, UK.

Received: 4 July 2025 Accepted: 20 November 2025
Published online: 17 December 2025

References

1. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409.
2. Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, Caughey AB, et al. Screening for lung cancer: US preventive services task force recommendation statement. JAMA. 2021;325(10):962–70.
3. Wood DE. POINT: should lung cancer screening be expanded to persons who don't currently meet accepted criteria set forth by the CHEST guidelines on lung cancer screening? Yes. Chest. 2018;153(6):1299–302.
4. Tammemagi MC, Ruparel M, Tremblay A, Myers R, Mayo J, Yee J, et al. USPSTF2013 versus PLCOm2012 lung cancer screening eligibility criteria (International Lung Screening Trial): interim analysis of a prospective cohort study. Lancet Oncol. 2022;23(1):138–48.
5. Robbins HA, Alcala K, Swerdlow AJ, Schoemaker MJ, Wareham N, Travis RC, et al. Comparative performance of lung cancer risk models to define lung screening eligibility in the United Kingdom. Br J Cancer. 2021;124(12):2026–34.
6. Tammemagi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med. 2013;368(8):728–36.
7. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, et al. DNA methylation aging clocks: challenges and recommendations. Genome Biol. 2019;20(1):249.
8. Liang R, Tang Q, Chen J, Zhu L. Epigenetic clocks: beyond biological age, using the past to predict the present and future. Aging Dis. 2024;16(6):3520–45.
9. Andersen AM, Philibert RA, Gibbons FX, Simons RL, Long J. Accuracy and utility of an epigenetic biomarker for smoking in populations with varying rates of false self-report. Am J Med Genet B Neuropsychiatr Genet. 2017;174(6):641–50.
10. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet. 2016;9(5):436–47.
11. Rutledge J, Oh H, Wyss-Coray T. Measuring biological age using omics data. Nat Rev Genet. 2022;23(12):715–27.
12. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.
13. Teschendorff AE, Horvath S. Epigenetic ageing clocks: statistical methods and emerging computational challenges. Nat Rev Genet. 2025;26(5):350–68.
14. Teschendorff AE. A comparison of epigenetic mitotic-like clocks for cancer risk prediction. Genome Med. 2020;12(1):56.
15. Lu AT, Binder AM, Zhang J, Yan Q, Reiner AP, Cox SR, et al. DNA methylation GrimAge version 2. Aging (Albany NY). 2022;14(23):9484–549.
16. Dugue 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 Spectr. 2020;5(1):pkaa109.
17. Dugue PA, Bassett JK, Joo JE, Jung CH, Ming Wong E, Moreno-Betancur M, et al. DNA methylation-based biological aging and cancer risk and survival: pooled analysis of seven prospective studies. Int J Cancer. 2018;142(8):1611–9.
18. Ma Z, Zhu C, Wang H, Ji M, Huang Y, Wei X, et al. Association between biological aging and lung cancer risk: cohort study and Mendelian randomization analysis. iScience. 2023;26(3):106018.
19. Mak JKL, McMurran CE, Kuja-Halkola R, Hall P, Czene K, Jylhava J, et al. Clinical biomarker-based biological aging and risk of cancer in the UK Biobank. Br J Cancer. 2023;129(1):94–103.
20. Cardenas A, Ecker S, Fadadu RP, Huen K, Orozco A, McEwen LM, et al. Epigenome-wide association study and epigenetic age acceleration associated with cigarette smoking among Costa Rican adults. Sci Rep. 2022;12(1):4277.
21. Dugué PA, Yu C, Hodge AM, Wong EM, Joo JE, Jung CH, et al. Methylation scores for smoking, alcohol consumption and body mass index and risk of seven types of cancer. Int J Cancer. 2023;153(3):489–98.
22. Battram T, Richmond RC, Baglietto L, Haycock PC, Perduca V, Bojesen SE, et al. Appraising the causal relevance of DNA methylation for risk of lung cancer. Int J Epidemiol. 2019;48(5):1493–504.
23. 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;6:10192.
24. Onwuka JU, Guida F, Langdon R, Johansson M, Severi G, Milne RL, et al. Blood-based DNA methylation markers for lung cancer prediction. BMJ Oncol. 2024;3(1):e000334.
25. World Health O. International classification of diseases for oncology (ICD-O). 3rd ed., 1st revision ed. Geneva: World Health Organization. 2013:2013.
26. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49(2):359–67.
27. Zhang Q, Vallerga CL, Walker RM, Lin T, Henders AK, Montgomery GW, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med. 2019;11(1):54.
28. Lu AT, Seeboth A, Tsai PC, Sun D, Quach A, Reiner AP, et al. DNA methylation-based estimator of telomere length. Aging (Albany NY). 2019;11(16):5895–923.
29. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018;10(4):573–91.
30. Belsky DW, Caspi A, Corcoran DL, Sugden K, Poulton R, Arseneault L, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. Elife. 2022. https://doi.org/10.7554/eLife.73420.
31. Higgins-Chen AT, Thrush KL, Wang Y, Minteer CJ, Kuo PL, Wang M, et al. A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. Nat Aging. 2022;2(7):644–61.
32. Ying K, Liu H, Tarkhov AE, Sadler MC, Lu AT, Moqri M, et al. Causality-enriched epigenetic age uncouples damage and adaptation. Nat Aging. 2024;4(2):231–46.
33. 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.
34. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, et al. Epigenetic prediction of complex traits and death. Genome Biol. 2018;19(1):136.
35. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clin Epigenetics. 2014;6(1):4.
36. Bollepalli S, Korhonen T, Kaprio J, Anders S, Ollikainen M. EpiSmokEr: a robust classifier to determine smoking status from DNA methylation data. Epigenomics. 2019;11(13):1469–86.
37. Morales Berstein F, McCartney DL, Lu AT, Tsilidis KK, Bouras E, Haycock P, et al. Assessing the causal role of epigenetic clocks in the development of multiple cancers: a Mendelian randomization study. Elife. 2022. https://doi.org/10.7554/eLife.75374.
38. Wan ES, DeMeo DL. Translating genomics into risk prediction. Thorax. 2017;72(7):598–9.
39. Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD, et al. Correlation of an epigenetic mitotic clock with cancer risk. Genome Biol. 2016;17(1):205.
40. Rahman ML, Breeze CE, Shu XO, Wong JYY, Blechter B, Cardenas A, et al. Epigenome-wide association study of lung cancer among never smokers in two prospective cohorts in Shanghai, China. Thorax. 2024;79(8):735–44.
41. Hicken MT, Dou J, Kershaw KN, Liu Y, Hajat A, Bakulski KM. Racial and ethnic residential segregation and monocyte DNA methylation age acceleration. JAMA Netw Open. 2023;6(11):e2344722.
42. Crimmins EM, Thyagarajan B, Levine ME, Weir DR, Faul J. Associations of age, sex, race/ethnicity, and education with 13 epigenetic clocks in a nationally representative U.S. sample: the health and retirement study. J Gerontol A Biol Sci Med Sci. 2021;76(6):1117–23.

Publisher's Note