DNA methylation at *AHRR* as a master predictor of smoke exposure and a biomarker for sleep and exercise

Ewelina Pośpiech[1*], Joanna Rudnicka[2], Rezvan Noroozi[2,4], Aleksandra Pisarek-Pacek[3,5], Bożena Wysocka[6], Aleksander Masny[6], Michał Boroń[6], Kamila Migacz-Gruszka[7], Paulina Pruszkowska-Przybylska[8], Magdalena Kobus[9], Dagmara Lisman[1], Grażyna Zielińska[1], Sandra Cytacka[1], Aleksandra Iljin[10], Joanna A. Wiktorska[11], Małgorzata Michalczyk[12], Piotr Kaczka[12], Michał Krzysztofik[13], Aneta Sitek[8], Magdalena Spólnicka[14], Andrzej Ossowski[1] and Wojciech Branicki[5,15]

Abstract

Background DNA methylation profiling may provide a more accurate measure of the smoking status than self-report and may be useful in guiding clinical interventions and forensic investigations. In the current study, blood DNA methylation profiles of nearly 800 Polish individuals were assayed using Illuminia EPIC and the inference of smoking from epigenetic data was explored. In addition, we focused on the role of the *AHRR* gene as a top marker for smoking and investigated its responsiveness to other lifestyle behaviors.

Results We found > 450 significant CpGs associated with cigarette consumption, and overrepresented in various biological functions including cell communication, response to stress, blood vessel development, cell death, and atherosclerosis. The model consisting of cg05575921 in *AHRR* ($p = 4.5 \times 10^{-32}$) and three additional CpGs (cg09594361, cg21322436 in *CNTNAP2* and cg09842685) was able to predict smoking status with a high accuracy of AUC = 0.8 in the test set. Importantly, a gradual increase in the probability of smoking was observed, starting from occasional smokers to regular heavy smokers. Furthermore, former smokers displayed the intermediate DNA methylation profiles compared to current and never smokers, and thus our results indicate the potential reversibility of DNA methylation after smoking cessation. The *AHRR* played a key role in a predictive analysis, explaining 21.5% of the variation in smoking. In addition, the *AHRR* methylation was analyzed for association with other modifiable lifestyle factors, and showed significance for sleep and physical activity. We also showed that the epigenetic score for smoking was significantly correlated with most of the epigenetic clocks tested, except for two first-generation clocks.

Conclusions Our study suggests that a more rapid return to never-smoker methylation levels after smoking cessation may be achievable in people who change their lifestyle in terms of physical activity and sleep duration. As cigarette smoking has been implicated in the literature as a leading cause of epigenetic aging and *AHRR* appears to be modifiable by multiple exogenous factors, it emerges as a promising target for intervention and investment.

Keywords Smoke exposure, Sleep, Exercise, DNA methylation, Epigenetic aging, Lifestyle biomarkers, AHRR

*Correspondence:
Ewelina Pośpiech
ewelina.pospiech@pum.edu.pl
Full list of author information is available at the end of the article

Pośpiech *et al. Clinical Epigenetics*   (2024) 16:147

Page 2 of 12

Background

Cigarette consumption is associated with numerous adverse health effects and accounts for more than 8 million deaths worldwide each year, according to the WHO. All forms of tobacco use are known to be harmful, including exposure to second-hand smoke and exposure in fetal life [1]. A growing body of research agrees that habitual smoking leaves a significant signature on DNA methylation [2, 3], and there is increasing evidence that some of the diseases associated with smoking may be mediated by smoke-induced changes in DNA methylation [4–6]. Smoke is also known to be a major cause of increased epigenetic aging [7–9]. In our recent study, we showed that those who smoke are, on average, four years older epigenetically than those who have never smoked [10]. The mechanism by which smoking affects DNA methylation is still under investigation, but current evidence suggests that it involves smoke-induced DNA damage and recruitment of DNA methyltransferases that methylate adjacent cytosines in CpG dinucleotides, or nicotine-induced downregulation of DNA methyltransferases [3, 11].

Monitoring the dynamics of changes in the DNA methylation profile with smoking exposure or after smoking cessation may be of great interest in diagnostics or clinical interventions to describe a patient's smoking status better than self-report and to estimate the risk of smoking-related diseases such as myocardial infarction, lung cancer, diabetes, or chronic obstructive pulmonary disease (COPD) [12, 13]. DNA methylation is a promising biomarker of health and biological aging due to the potential to revert. However, cigarette-sensitive markers may differ in their ability to return to the pre-exposure methylation state [14–17]. An association between methylation profile and time since smoking cessation has also been reported [16–18]. Moreover, inferring an individual's smoking status can also be a useful piece of information in a criminal investigation to complement an offender's genetic profile [19–21].

Numerous CpG sites (CpGs) have been associated with tobacco smoking in blood, but the strongest smoking-induced epigenetic response has been reported for the *AHRR* gene [14, 16, 22–24], with cg05575921 included in most of predictive models available in the literature [21, 25–29]. A protein encoded by the *AHRR* gene participates in the aryl hydrocarbon receptor (AhR) signaling cascade, which mediates the degradation of environmental toxins, and regulates cell growth and differentiation. *AHRR* was found to be associated with smoking in Europeans and Asians, although population differences in methylation levels for *AHRR* and other markers were observed and an interaction between smoking status and ethnic group was identified at the *AHRR* locus [25].

Importantly, *AHRR* methylation was found to be informative for smoking inference in both blood and saliva tissues [27]. *AHRR* is also a predictor of epigenetic aging, included in the pace of aging model [30] and mortality risk score models [8], but it is not clear whether its role in epigenetic aging is limited to smoking exposure only. In the literature, *AHRR* has been associated with smoking-related diseases such as COPD incidence, lung function, lung cancer [31, 32] and atherosclerosis [33].

In the current study, blood DNA methylation profiles of nearly 800 individuals were assayed using Illuminia EPIC arrays. By conducting a novel epigenome-wide association study, we described epigenetic markers of smoking in the Polish population and focused on the role of the *AHRR* gene as a top marker of smoking and other lifestyle habits. We compared the methylation profiles of smokers, never smokers, and former smokers, which allowed us to infer the temporality of DNA methylation at smoke-sensitive markers. Furthermore, we developed a new compact 4-CpG prediction model for smoking, which may be of practical importance for forensic and diagnostic purposes.

Methods

Study population and smoking information

The study comprised of a total of 772 participants. First, whole blood samples were collected from 737 Polish unrelated individuals aged ≥ 20 years old (age mean $\pm$ SD $= 46.4 \pm 14.8$) along with information on cigarette smoking history (Table S1). Participants were asked whether they were current or former cigarette smokers. Current smokers were asked to define their smoking frequency and were categorized as occasional smokers ($\leq 1–2$ times per week), regular light smokers (< 5 cigarettes per day), regular medium smokers (5–20 cigarettes per day) or regular heavy smokers (> 20 cigarettes per day). Former smokers were asked to define the number of years since quitting. For the purpose of statistical tests, current smokers were all included in one group of smokers or in two groups of light smokers or heavy smokers, with light smokers including individuals who smoke occasionally or regularly in small amounts, while the second group included individuals who smoke regularly in medium amounts or regularly in large amounts. Blood samples were also collected from a group of $N = 35$ children under the age of 13 who were treated as an additional validation group, although no smoking questionnaire was available at the time of sample collection.

Written informed consent was obtained from all participants, and the study was approved by the Bioethics Committee of the Jagiellonian University in Krakow (decision no. 1072.6120.132.2018).

Pośpiech *et al. Clinical Epigenetics*   (2024) 16:147

Page 3 of 12

Description of other lifestyle habits

Information on other lifestyle and socio-demographic characteristics was also available in this study ($N=737$) and included education (university degree, primary school, high school, or vocational school); socioeconomic status (SES) (high vs. above average vs. average vs. low vs. very low); job type (physical/manual, mental work partially sedentary (up to 4 h per day; sedentary mental; SM), mental work only sedentary (more than 4 h per day; long sedentary mental; LSM), or retired/unemployed); workplace exposures (low/high workplace temperature, exposure to pesticides/chemicals, toxins/heavy metals/ air pollution, ionizing radiation, sun, and stress); physical activity (e.g., exercise, jogging, cycling, yoga, etc.); frequency of alcohol consumption (non-drinkers, occasional drinkers (drinking once a week), and frequent drinkers (drinking at least three times a week)); number of meals per day; servings of fruits and vegetables; frequency of fish and meat consumption; cups of coffee per day; and hours of sleep. Frequencies of lifestyle-related characteristics in the study population are shown in Table S2.

DNA methylation data analysis

DNA was extracted from blood samples using an automated method and the Maxwell RSC Blood DNA Kit (Promega Corporation), and its quantity and quality were evaluated using NanoDrop (Thermo Scientific, MA, USA) and the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). The amount of 500 ng DNA per sample was subjected to bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research Corp., CA, USA) according to the manufacturer's instructions for Infinium assays. DNA methylation profiles were then generated using the Illumina Infinium Methylation EPIC microarray (Illumina, San Diego, CA, USA) [34]. To minimize batch effects, samples were analyzed at similar times and randomized on microarray plates prior to analysis using the web-based RANDOMIZE application [35]. In addition, at the data analysis stage, principal component analysis was used to detect potential batch effects, and the associated scatter plot of the top PCs did not indicate a high source of variation. Primary quality control of raw array data, filtering of low-quality samples and probes was performed as described in our previous study [10]. Functional normalization (FunNorm) on filtered data (750 821 probes) was performed using the minfi package in R version 4.3.1 [36].

Statistical data analysis strategy

Using specific sample sets and methylation data, a series of statistical analyses were performed, including an epigenome-wide association study (EWAS) for categorical outcomes and associated functional annotation analysis, followed by model building using logistic regression, again for categorical outcomes and model validation, and finally an association analysis of lifestyle factors with DNA methylation at the AHRR cg05575921 and since DNA methylation was considered the dependent variable in this type of analysis, linear regression was used. In the EWAS analysis, all current smokers ($N=171$) were compared to all never smokers ($N=404$). In turn, a subset of the EWAS sample was used for prediction model training (70%, $N=238$) and testing (30%, $N=100$), and samples within each group were matched for age and sex. Former smokers ($N=162$) and children ($N=35$) were used as external validation groups. The association of sociodemographic and lifestyle characteristics with DNA methylation at the *AHRR* was assessed in the dataset of blood samples from $N=737$ individuals, including the EWAS sample set and former smokers. Summary statistics of the study samples are shown in Table S1.

Epigenome-wide association study

Differentially methylated positions (DMPs) for smoking, defined in a binary manner as smoking ($N=171$) vs. nonsmoking ($N=404$), were identified by a EWAS study using the limma package [37]. The results of the EWAS analysis were adjusted for age, sex, and blood cell counts (T-cells: CD8+, CD4+, CD8 naive, CD4 naïve; NK, B-cells; monocytes; granulocytes) predicted with a Houseman method [38]. Associations between differentially methylated CpGs and Gene Ontology (GO) and KEGG terms were investigated using the missMethyl R package [39]. The Benjamini–Hochberg false discovery rate (FDR) method was used to correct for multiple comparisons, and results with FDR $p < 0.05$ were considered significant.

Development and validation of predictive model

Predictive model training was performed on 238 samples, including 119 non-smokers and 119 current smokers, matched for age and sex. No significant differences in sex (Pearson's chi-square $p=1.0$) and age (Mann–Whitney U $p=0.999$) between the groups were confirmed with appropriate statistical tests (Table S1). Variable selection was performed on the DMPs identified in the EWAS study by stepwise logistic regression, using the significance of the score statistic as the criterion for predictor entry and the probability of a likelihood ratio statistic based on maximum partial likelihood estimates as the test of exclusion. Nagelkerke's R^2 statistic was calculated, which is an approximation of the R^2 statistic for linear regression. The R^2 statistic measures the goodness of fit of a model and summarizes the proportion of variance explained by each predictor. Bootstrap analysis

Pośpiech *et al. Clinical Epigenetics*　　(2024) 16:147

Page 4 of 12

with $n = 10,000$ permutation tests was applied to evaluate the robustness of the model. Two types of models were developed, the binomial model for comparing two categories (current smokers vs. never smokers) and the multinomial model, where three categories were considered (never smokers vs. light smokers vs. heavy smokers). The performance of the developed models was tested on an independent set of 100 blood samples from 50 smokers and 50 never smokers and the accuracy was described by AUC (the area under the ROC curve), sensitivity (true positive rate, for the binomial model it is the percentage of smokers correctly identified as smokers), specificity (true negative rate, for the binomial model it is the percentage of never smokers correctly identified as never smokers), and total number of correct predictions (the percentage of correct scores calculated for the combined group of smokers and never smokers). The newly developed biomarker for smoking was further validated using two additional sample sets, former smokers ($N = 162$) and children ($N = 35$). The epigenetic score for smoking, i.e. the probability of smoking and its distribution in different groups, was analyzed. Prediction modelling was done using IBM SPSS Statistics 29.

Analysis of the correlation between epigenetic score for smoking and epigenetic age

In our previous study, we showed that in our database of Polish samples, epigenetic clocks correlate strongly with lifestyle habits, including self-reported smoking [10]. Here, we tested whether the epigenetic score for smoking correlates with different measures of epigenetic age acceleration (EAA) as determined by different clocks, including Horvath 2013 [40], Hannum 2013 [41], Horvath Skin&Blood [42], PhenoAge [43], GrimAge [7], FitAge [44], Mortality Risk Score (MRS) [8] and DunedinPACE [45]. For all clocks except the latter two, epigenetic age acceleration (EAA) was calculated as described in our previous study [10]. The pairwise correlation between the epigenetic score for smoking measured by our new binomial model and different EAAs, MRS or PACE was analyzed by Pearson's correlation test.

Association tests for AHRR and lifestyle

We tested for association between DNA methylation at cg05575921 in *AHRR* and lifestyle data. DNA methylation at cg05575921 was treated as the outcome, while each socio-demographic and lifestyle characteristic was independently entered into the linear regression model as an independent variable. The strength of association was interpreted using standardized beta coefficients, and the R^2 statistic was used to calculate the proportion of variation in DNA methylation at *AHRR* explained by lifestyle factors. Results were always adjusted for age and sex,

smoking status, blood cell components, and DNA methylation pack-years calculated with the Horvath online tool (https://dnamage.clockfoundation.org/); and results $p < 0.05$ were considered statistically significant.

Results

Epigenome-wide association study for smokers vs. never smokers

The EWAS analysis for smoking yielded 459 age- and sex-adjusted CpG associations in a group of 575 individuals (Fig. 1). The top 12 CpGs (FDR $p < 1 \times 10^{-7}$) are shown in Table 1, while the full list of FDR-significant CpGs (FDR $p < 0.05$; raw $p < 5 \times 10^{-5}$) is shown in Table S3. When the results were additionally corrected for several lifestyle factors, including socioeconomic status, education, coffee consumption, and sleep duration, 99 CpGs remained significant (Table S4). The highest ranking hit was **cg05575921** *AHRR* ($p = 4.5 \times 10^{-32}$ adjusted for sex and age; $p = 8.7 \times 10^{-27}$ adjusted for age, sex and lifestyle factors), and 6 other CpGs in *AHRR* also reached EWAS significance (Table S3). Downstream pathway analysis revealed enrichment in multiple biological functions (FDR $p < 0.05$), including among others organism development, cell communication, regulation of signaling, response to stress, blood vessel development, and cell death (Table S5). When KEGG pathways were tested in the overrepresentation analysis, although no significant groups of genes were revealed after FDR correction, the top significant (raw $p < 0.05$) terms included pathways in cancer, lipid and atherosclerosis, purine metabolism, and axon guidance (Table S6). When all 459 EWAS-significant CpG markers were analyzed in an independent group of former smokers, 117 of them were found to be statistically significantly (nominal $p < 0.05$) associated with time since quitting, with 21 reaching $p < 0.0001$ (Table S7).

Predictive analysis of smoking status

Prediction models were developed using logistic regression and stepwise marker selection. Two types of data categorization were used, including a simple binary definition of smoking and including 3 smoking categories defined as non-smokers, occasional smokers, and regular smokers. Four CpGs (cg05575921 in *AHRR*, cg09594361 at chr1: 54905423, cg21322436 in *CNTNAP2* and cg09842685 at chr12: 4492769) were selected to explain a total of Nagelkerke $R^2 = 38.9\%$ of the observed variation in smoking. The analysis of multicollinearity showed low values of variance inflation factor (VIF) for the predictors in the model (between 1.1 and 1.4). The characteristics of a binomial regression model are shown in Table 2.

The binomial model predicted smoking with high accuracy described by the AUC parameter at the level

Pośpiech *et al. Clinical Epigenetics*   (2024) 16:147

Page 5 of 12
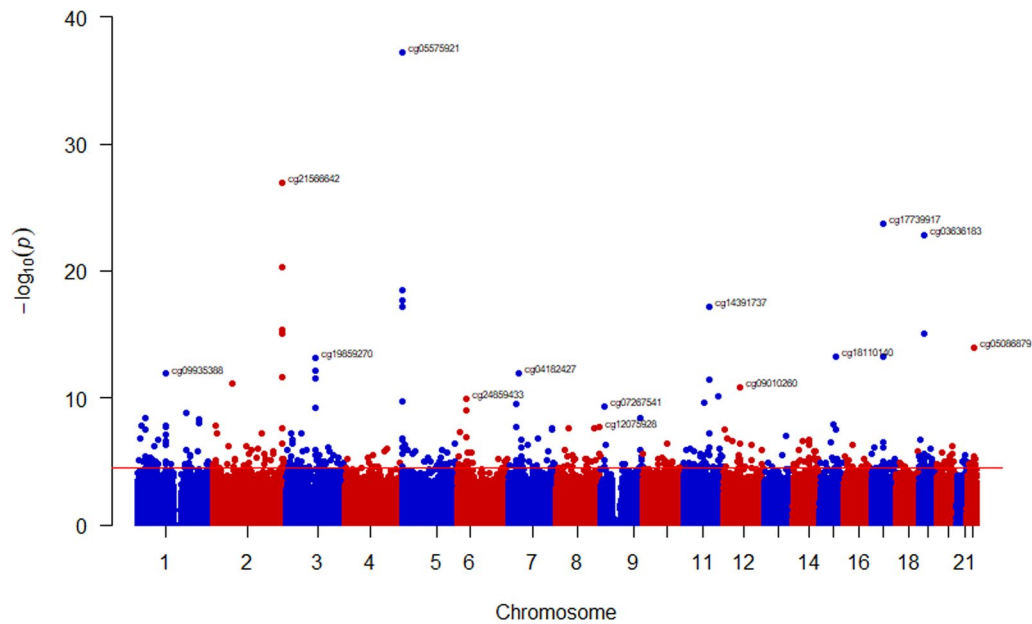
Fig. 1 Manhattan plot of EWAS for smoking adjusted for age, sex and blood cell count. Red horizontal line corresponds to FDR $p < 0.05$, with the top annotated markers meeting this threshold

Table 1 Top 12 (FDR $p < 10^{-7}$) independent EWAS-based CpGs associated with smoking

CpG ID	Chr	hg19	Gene name	Gene group	p	FDR p
cg05575921	chr5	373378	*AHRR*	Body	5.95E−38	4.47E−32
cg21566642	chr2	233284661	NA	NA	1.08E−27	4.05E−22
cg17739917	chr17	38477572	*RARA*	5'UTR	1.77E−24	4.44E−19
cg03636183	chr19	17000585	*F2RL3*	Body	1.53E−23	2.86E−18
cg14391737	chr11	86513429	*PRSS23*	5'UTR;Body	7.14E−18	5.96E−13
cg05086879	chr22	39861490	*MGAT3*	5'UTR	1.04E−14	6.02E−10
cg18110140	chr15	75350380	NA	NA	6.18E−14	3.09E−09
cg19859270	chr3	98251294	*GPR15*	1stExon	7.33E−14	3.44E−09
cg02657160	chr3	98311063	*CPOX*	Body	7.32E−13	3.23E−08
cg04182427	chr7	32315577	*PDE1C*	Body	1.14E−12	4.75E−08
cg09935388	chr1	92947588	*GFI1*	Body	1.28E−12	5.08E−08
cg02978227	chr3	98292027	NA	NA	2.67E−12	9.56E−08

Table 2 Variables in the binomial logistic regression model and bootstrap analysis (n permutation tests = 10,000)

CpG in a model	B	Bias	SE	p	Bootstrap 95% CI	
					Lower	Upper
cg05575921	− 12.947	− 0.482	2.540	< 0.001	− 18.918	− 8.891
cg09594361	− 19.051	− 0.473	7.173	0.004	− 34.293	− 6.033
cg21322436	− 8.577	− 0.465	5.250	0.086[*]	− 19.659	0.768
cg09842685	− 14.905	− 0.632	6.869	0.022	− 29.346	− 2.082
Constant	35.634	1.228	7.302	< 0.001	24.072	52.685

SE Standard error, *CI* Confidence interval

[*] cg21322436 was retained in a model because it improved prediction accuracy in both, training and test sets, as shown in Table S8

Pośpiech *et al. Clinical Epigenetics* (2024) 16:147

Page 6 of 12

of 0.80, in both training ($N=238$) and test ($N=100$) sets (Table 3). The sensitivity and specificity were 66% and 82%, respectively. The final number of correct classifications was 74% in the test set. The epigenetic score for smoking in the smoker and never smoker categories was compared and shown in Fig. 2. When the distribution of probabilities was analyzed in a larger number of categories, taking into account the frequency of smoking, a gradual increase in the probability of smoking was observed, starting from occasional smokers to regular smokers who smoke in large quantities. Importantly, the mean score (probability) of smoking was higher for former smokers than for never smokers, but lower than for current smokers (Fig. 3). In addition, the specificity of the model was tested in children aged < 13 years and, as expected, very low probabilities of smoking were achieved in this group. Notably, the mean score for smoking was higher in the never smokers' group (0.37 ± 0.19) than in the children (0.20 ± 0.15).
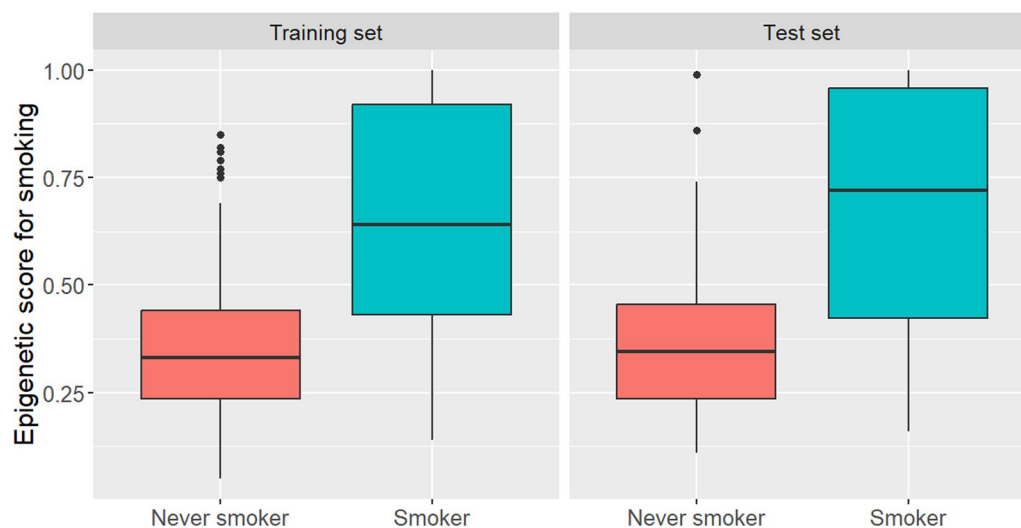
In the next step, the categories of occasional smokers and regular light smokers were combined into one category of light smokers, while regular medium smokers and regular heavy smokers were combined into another category of heavy smokers. In the multinomial model, the AUC values were very high for never smokers and heavy smokers (0.8 and 0.88, respectively), and lower, but close to 0.7 level of prediction accuracy was achieved for the light smoker category, intermediate between never smokers and heavy smokers. High sensitivity was obtained for the never smoker category (84%), which was significantly lower for light smokers (44%), while high specificity of prediction was achieved for the light smoker and heavy smoker categories (80% and 97%, respectively).

All four CpGs selected in the models were significantly correlated with smoking status in both types of EWAS analyses (Table S3 and S4), and positively influenced the AUC value in both the training and test sets (Table S8). The single *AHRR* cg05575921 was found to explain 21.5% of the variation in smoking and, when used alone, predicted smoking at the AUC level of 0.76 in the test set (Table 3). The *AHRR* gene was also found to have the highest observed change in DNA methylation

Table 3 Predictive performance for binomial and multinomial models for smoking status

Model type	Smoking category	Training set	Test set				
		AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
AHRR cg05575921	Smoker vs. never smoker	0.75	0.56	0.82	0.76	0.60	0.82
Binomial 4-CpG	Smoker vs. never smoker	0.80	0.64	0.82	0.80	0.66	0.82
Multinomial 4-CpG	Cat. 1 Never smoker	0.80	0.84	0.57	0.80	0.84	0.66
	Cat. 2 Light smoker	0.68	0.39	0.82	0.69	0.44	0.80
	Cat. 3 Heavy smoker	0.89	0.56	0.97	0.88	0.52	0.97

Fig. 2 Epigenetic score (probability) for smoking for never smokers and current smokers within training and test sets, as determined by a binomial regression model

Pośpiech *et al. Clinical Epigenetics*   (2024) 16:147

Page 7 of 12
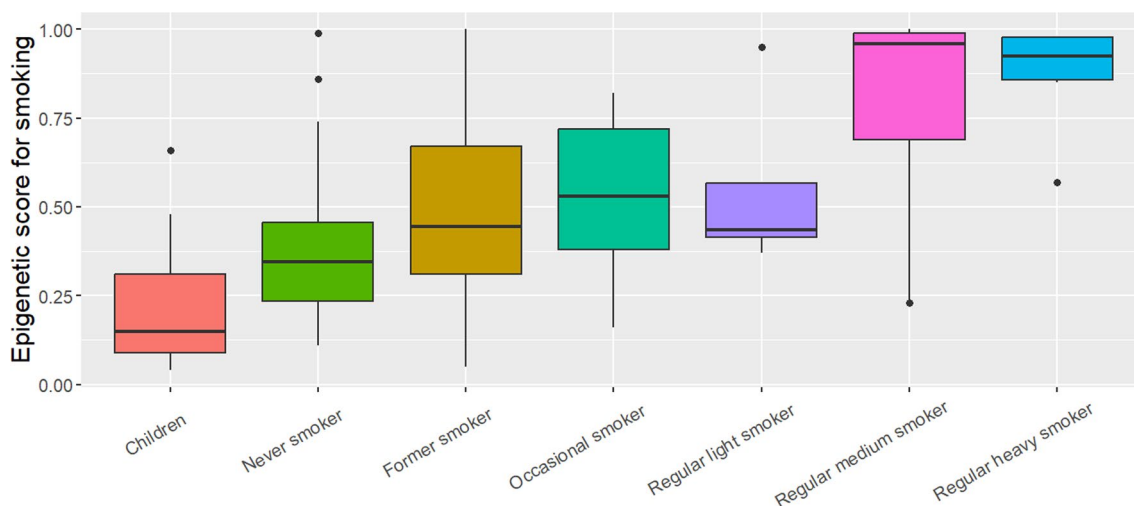
Fig. 3 Epigenetic score (probability) for smoking for different sample groups as determined by a binomial regression model. Smokers were divided into four groups based on frequency of smoking: occasional smokers, regular light smokers, regular medium smokers and regular heavy smokers

when comparing current smokers to never smokers with a mean difference in DNA methylation beta of 0.13 ± 0.01 (Fig. 4). Importantly, the *AHRR* cg05575921 and the other three DMPs included in the prediction models were all hypomethylated in smokers.

To analyze the dynamics of DNA methylation in smoke associated CpGs, we also compared non-smokers and

current smokers with former smokers and found that there were no significant differences in DNA methylation beta levels between never smokers and former smokers for all CpGs in the model except cg05575921 in *AHRR* (Fig. 4). However, the methylation profile of former smokers at cg05575921 was more different from current smokers ($p < 0.001$) than from never smokers ($p = 0.006$).
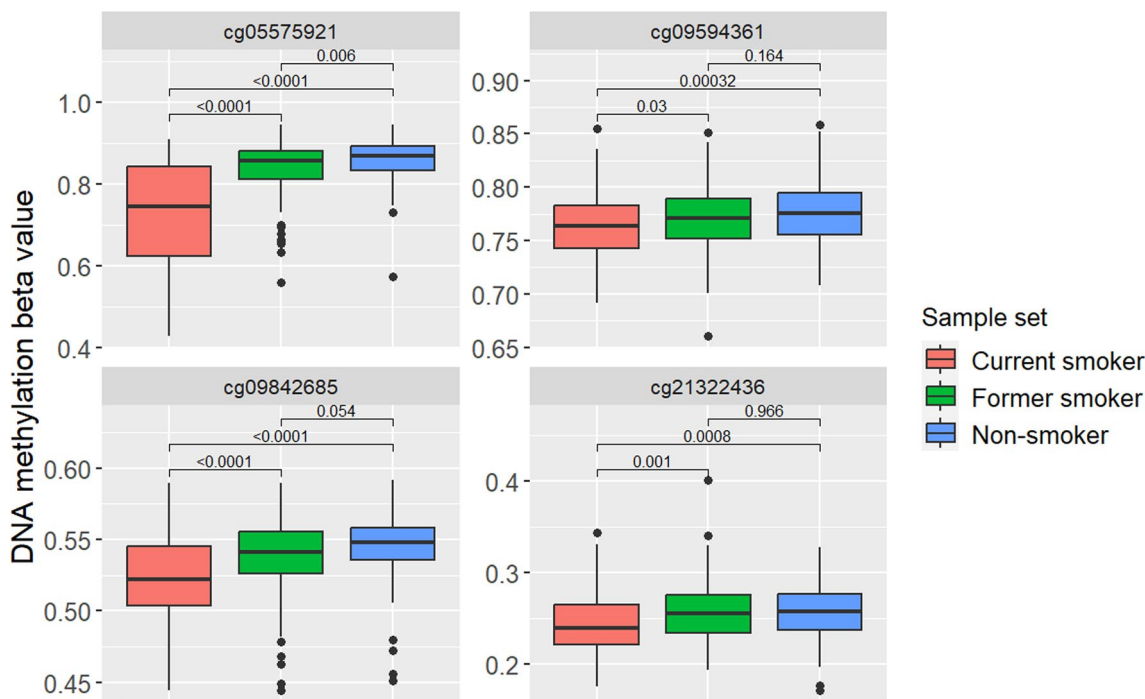
Fig. 4 Boxplots of methylation levels for 4 CpGs included in the predictive model for smoking and comparison between current smokers, former smokers and never smokers. Groups were matched for age and mean age of each group was 46.3

Pośpiech *et al. Clinical Epigenetics*   (2024) 16:147

Page 8 of 12

This finding may indicate the potential recovery of DNA methylation after smoking cessation and suggests that the rate of these changes may differ for different loci in the genome.

Epigenetic score for smoking correlates with epigenetic age acceleration

Analysis of the correlation between the epigenetic score for smoking (the probability of smoking) and various measures of epigenetic age acceleration (EAA) showed a significant effect for most clocks, except Horvath 2013 and Horvath Skin&Blood, which belong to the first generation of epigenetic clocks trained solely on chronological age. The strongest correlation and highest significance was observed for the GrimAge clock (Pearson's $R = 0.665$, $p = 1.25E-63$) and the Mortality Risk Score (Pearson's $R = 0.525$, $p = 6.30E-36$). Results are shown in Table S9.

Associations between *AHHR* and various lifestyle factors

Methylation of *AHRR* cg05575921 decreases slightly with age ($p = 1.89 \times 10^{-5}$) and is lower in males ($p = 1.67 \times 10^{-6}$). Of all habits examined, cigarette smoking was the most strongly correlated with methylation at cg05575921 ($p = 4.71 \times 10^{-41}$). Higher methylation was observed in physically active individuals who exercised daily, but not less than once a week, and the results were significant after adjustment for age, sex, smoking status, blood cell count and DNA methylation pack-years ($p = 0.028$). People who slept at least 8 h per night also showed a more favorable methylation profile in cg05575921 compared to people with a sleep deficit ($p = 0.027$, Table 4). The

full association tests for all lifestyle factors are shown in Table S10.

Discussion

To address the problem of the aging population and as a prevention strategy for age-related diseases, epigenetic reprogramming is a promising solution [46, 47]. As cigarette smoking has been implicated as a major cause of epigenetic aging [9, 10, 48], and literature data show that smoke-induced changes in DNA methylation patterns may mediate disease development [4–6], smoking-responsive markers appear to be promising targets for intervention and investment. DNA methylation-based smoking inference is also of interest in the field of forensic DNA phenotyping, which aims to describe the physical appearance, biogeographic ancestry, age, and lifestyle information of a perpetrator or human remains subject to identification [49, 50].

DNA methylation at a number of genomic loci has been associated with smoking exposure, and a clear trend towards hypomethylation of the genome has been observed in smokers [22]. This phenomenon was also observed for four markers included in our new smoking prediction model. The largest magnitude of effect was observed for the aryl hydrocarbon receptor repressor gene (*AHRR*), which has an established role in the response to smoking [23–25, 33, 51]. Reduced methylation at *AHRR* cg05575921 (chr5: 373378) results in increased expression of an AHRR-encoded protein [52] that represses the AhR receptor, thereby negatively impacting toxic clearance processes [53]. AhRR

Table 4 List of lifestyle behaviors significantly associated with DNA methylation at *AHRR* cg05575921 in a group of *N* = 737 individuals

Lifestyle/demographic factor	*AHRR* cg05575921 association test				
	B	Std. error	Std. B	R²	p
Age	−0.001	0.0002	−0.155	0.020	1.89E−05
Sex male (ref. female)	−0.031	0.006	−0.174	0.026	1.67E−06
Smoking					
Current smoker (ref. never smoker)	−0.104	0.007	−0.516	0.252	4.71E−41
Current smoker (ref. former smoker)	−0.092	0.011	−0.424	0.100	6.24E−15
Former smoker (ref. never smoker)	−0.028	0.005	−0.221	0.065	6.65E−08
Lifestyle risk factors					
Sleep hours ≥ 8 h (ref. < 8 h.)	0.021	0.010	0.045	0.012	0.027
Physical activity (PA) (ref. no PA)	0.009	0.004	0.040	0.027	0.028
PA freq. (ref. none)					
Once a month	−0.001	0.008	−0.003	0.007	0.930
Once a week	0.011	0.005	0.053	0.038	0.037
Every day	0.012	0.005	0.058	0.034	0.028

Table shows the results of univariate linear regression analysis with DNA methylation at AHRR as dependent variable and each lifestyle factor introduced in separate models as independent variable, adjusted for age, sex, smoking status and blood cell count

Reference categories are provided for each lifestyle/demographic factors tested

protein can also affect other signaling pathways and regulates inflammatory responses [54]. We noted the potential reversibility of DNA methylation following smoking cessation, although the DNA methylation pattern in *AHRR* in former smokers was still significantly different from that in never smokers, unlike the other three markers in the model for which we observed that methylation profiles did not differ significantly between never-smokers and ex-smokers. Different markers may require different times to return to levels seen in never smokers, and the literature suggests that methylation at *AHRR* returns to normal after 5 years of abstinence [22]. Our study suggests that this time may be longer or population dependent, as the mean time since quitting smoking in our group of former smokers was 14.6 ± 10.5 years. The obtained data also suggest that a faster return to methylation levels characteristic of never-smokers after quitting could potentially be achieved in people who change their lifestyle in terms of physical activity, sleep hours and diet. This shows the potential of interactions in shaping the final profile of DNA methylation.

AHRR is recognized as a strong marker of smoking in Europeans and Asians [55], but also as a biomarker of epigenetic aging [30], mortality [8], smoking-associated diseases and lung cancer [31, 32, 56]. In our study, we demonstrated that *AHRR* is also sensitive to lifestyle factors other than cigarette smoking. Favorable effect and increased DNA methylation was observed for enough sleep hours and physical activity (at least once a week). However, the significance of the association and the magnitude of the effect were smaller than for smoking. These results suggest the role of *AHRR* in epigenetic aging via mechanisms independent of smoking consumption and its power as a general biomarker of health and fitness. Importantly, members of the AhR-signaling pathway have been linked in the literature to the circadian rhythms [57, 58], and SNP polymorphisms in the *AHRR* gene associated with insomnia and early awakening [59]. These findings fit in well with our results showing that DNA methylation at *AHRR* is sensitive to sleep duration. The beneficial effects of physical activity on epigenetic aging have been widely described [60, 61] and confirmed in our recent study [10]. Daily exercise was associated with reduced epigenetic aging as measured by most of the known epigenetic clocks. Studies showed that exercises can lead to mobilization of natural killer cells involved in cell cytotoxicity via AhR/IDO pathway [62]. A recent study showed that acute exercise can affect AhR signaling, which in turn can influence the expression of the programmed cell death protein (PD-1), a promising target in cancer therapy [63]. Physical activity is therefore emerging as an important factor in immune regulation.

We showed that the *AHRR* cg05575921 was the main predictor of smoking in our study population, allowing high accuracy of smoking prediction described by the AUC parameter at the level of 0.76 when used alone. The value of AUC considered acceptable for implementation is 0.7 [64, 65]. To augment the performance and accuracy of the model, larger number of predictors is needed. In our study, three additional methylation markers were selected that independently contributed to the accuracy of the smoking inference. Two of them have a well-established role in smoking, *CNTNAP2* cg21322436 (chr7:145812842) [16, 66–68] and cg09842685 (chr12: 4492769) [69], and their association with the risk of COPD and lung cancer has also been described [70]. *CNTNAP2* is one of the largest genes in the human genome and encodes a contactin-associated protein-like 2a, which is a type of neurexin protein involved in cell adhesion and nervous system development. *CNTNAP2* has been implicated in many neurodevelopmental diseases. Maternal smoking during pregnancy was found out to impact DNA methylation at *CNTNAP2* and decreased level of methylation was suggested as a protective mechanism against adverse effects of smoking [71]. CpG cg09594361 (chr1:54905423) is a novel marker whose role in smoking has not been described so far. Importantly, this CpG was not covered by an older Infinium 450K array.

The developed 4-CpG model is characterized by high prediction accuracy with AUC = 0.8 and number of correct classifications at 74% in the test set. This result is more or less comparable to other published prediction models, using from 1 to 13 CpG markers for model training [21, 26, 29], although it should be noted that AUCs reaching 0.9 which is higher than in our study were also reported [21]. Our model allows to balance predictive accuracy and compactness of the marker panel, avoiding the overfitting often discussed in the literature [21, 29] and fulfilling the criteria often set in forensic genetics when dealing with difficult and degraded DNA material. The small number of markers also allows the use of alternative, less expensive and more available methods of DNA methylation analysis, such as pyrosequencing or high-throughput targeted DNA sequencing [72]. Furthermore, in our study we provided sufficiently accurate smoking inference at a higher prediction resolution by considering three possible categories of smoking status (never smoker vs. light smoker vs. heavy smoker), with former smokers showing similar methylation profiles to light smokers. Lower predictive accuracy for former smokers category and misclassification of former smokers as never smokers have been observed in previously published prediction models [21, 29, 74]. Furthermore, as we have shown in our study that *AHRR* is modifiable

by other lifestyle factors, this may also have an impact on the accuracy of smoking prediction. Importantly, high predictive accuracy was observed in our study for a group of children used as an additional set for model validation. The probabilities of smoking were lower for children compared to never-smokers, suggesting that passive smoking and the cumulative effects of other exogenous agents may affect DNA methylation at smoke-associated CpGs over time. However, although second-hand smoke exposure is a known risk factor for diseases, passive smoking was shown to be much less pronounced in the DNA methylation pattern than active smoking [74].

Our study does not come without its limitations. The sample groups used to train and test the predictive models were part of the sample set used to perform the EWAS analysis. Therefore, the results should be validated in the future. On the other hand, the use of two additional independent groups, i.e. former smokers and children, strengthens the conclusions of the study. Another weakness of the study is the lack of questionnaire data on smoking among those under 13 years of age. Therefore, the assumption of non-smoking in this group is biased. In addition, given the large number of smoking-associated markers known in the literature and discovered in our project, alternative methods of variable selection could be applied in the future [75, 76] to improve the selection of the most relevant variants, facilitate the detection of epistatic effects [77], and ultimately further improve the accuracy of prediction.

Conclusions

We confirmed that DNA methylation is highly predictive of smoking status and revealed the role of > 400 CpGs, including novel loci involved in smoking. We confirmed *AHRR* as the top locus associated with smoking and provided novel data showing that changes in DNA methylation at the *AHRR* can be achieved by multiple lifestyle factors. Although there are reports in the literature suggesting the role of *AHRR* polymorphism in insomnia, this is the first study to link DNA methylation at *AHRR* with sleep duration. We developed a competitive model for smoking inference from blood consisting of only 4 CpGs. High predictive accuracy was obtained for the binomial model, but importantly, the multinomial model was able to predict three categories of smoking with reasonable accuracy. We also demonstrated a high correlation between the epigenetic score for smoking and epigenetic age acceleration. We provided novel evidence for DNA methylation reversion after smoking cessation, which is of utmost importance given that DNA methylation at the *AHRR* and other loci has been associated with the risk of developing lung cancer and other diseases in the literature.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13148-024-01757-0.

Additional file 1

Additional file 2

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study was approved by the Bioethics Committee of the Jagiellonian University in Krakow (decision no. 1072.6120.132.2018).

Consent for publication
Written informed consent forms were obtained from all participants.

Competing interests
The authors declare that they have no conflicts of interest.

Author details
[1]Department of Forensic Genetics, Pomeranian Medical University in Szczecin, Powstańców Wlkp. 72, 70-111 Szczecin, Poland. [2]Doctoral School of Exact and Natural Sciences, Jagiellonian University, Krakow, Poland. [3]Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland. [4]Johns Hopkins University School of Medicine, Baltimore, USA. [5]Institute of Zoology and Biomedical Research of the Jagiellonian University, Krakow, Poland. [6]Central Forensic Laboratory of the Police, Warsaw, Poland. [7]Department of Dermatology, Jagiellonian University Medical College, Krakow, Poland. [8]Department of Anthropology, Faculty of Biology and Environmental Protection, University of Lodz, Lodz, Poland. [9]Institute of Biological Sciences, Faculty of Biology and Environmental Sciences, Cardinal Stefan Wyszynski University in Warsaw, Warsaw, Poland. [10]Department of Plastic, Reconstructive and Aesthetic Surgery, Medical University of Lodz, Lodz, Poland. [11]MelissaMed M. Gauer, General Partnership, Lodz, Poland. [12]Department of Sport Nutrition, The Jerzy Kukuczka Academy of Physical Education in Katowice, Katowice, Poland. [13]Institute of Sports Sciences, The Jerzy Kukuczka Academy of Physical Education in Katowice, Katowice, Poland. [14]Center for Forensic Science, University of Warsaw, Warsaw, Poland. [15]Institute of Forensic Research, Krakow, Poland.

Pośpiech *et al. Clinical Epigenetics*   (2024) 16:147

Page 11 of 12

References

1. Kobus M, Sitek A, Antoszewski B, Rożniecki JJ, Pełka J, Żądzińska E. The impact of exposure to tobacco smoking and maternal trauma in fetal life on risk of migraine. Front Neurosci. 2023. https://doi.org/10.3389/fnins.2023.1191091.
2. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet. 2011;88:450–7.
3. Satta R, Maloku E, Zhubi A, Pibiri F, Hajos M, Costa E, et al. Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. Proc Natl Acad Sci U S A. 2008;105:16356–61. https://doi.org/10.1073/pnas.0808699105.
4. Maas SCE, Mens MMJ, Kühnel B, van Meurs JBJ, Uitterlinden AG, Peters A, et al. Smoking-related changes in DNA methylation and gene expression are associated with cardio-metabolic traits. Clin Epigenet. 2020;12:1–16. https://doi.org/10.1186/s13148-020-00951-0.
5. Zhang H, Kalla R, Chen J, Zhao J, Zhou X, Adams A, et al. Altered DNA methylation within DNMT3A, AHRR, LTA/TNF loci mediates the effect of smoking on inflammatory bowel disease. Nature Commun. 2024;15:1–14.
6. Fragou D, Pakkidi E, Aschner M, Samanidou V, Kovatsi L. Smoking and DNA methylation: correlation of methylation with smoking behavior and association with diseases and fetus development following prenatal exposure. Food Chem Toxicol. 2019;129:312–27.
7. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging. 2019;11(2):303.
8. Zhang Y, Wilson R, Heiss J, Breitling LP, Saum KU, Schöttker B, et al. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. Nat Commun. 2017. https://doi.org/10.1038/ncomms14617.
9. Klopack ET, Carroll JE, Cole SW, Seeman TE, Crimmins EM. Lifetime exposure to smoking, epigenetic aging, and morbidity and mortality in older adults. Clin Epigenet. 2022. https://doi.org/10.1186/s13148-022-01286-8.
10. Noroozi R, Rudnicka J, Pisarek A, Wysocka B, Masny A, Boroń M, et al. Analysis of epigenetic clocks links yoga, sleep, education, reduced meat intake, coffee, and a SOCS2 gene variant to slower epigenetic aging. GeroScience. 2023. https://doi.org/10.1007/s11357-023-01029-4.
11. Lee KWK, Pausova Z. Cigarette smoking and DNA methylation. Front Genet. 2013. https://doi.org/10.3389/fgene.2013.00132.
12. 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.
13. Zhang Y, Elgizouli M, Schöttker B, Holleczek B, Nieters A, Brenner H. Smoking-associated DNA methylation markers predict lung cancer incidence. Clin Epigenet. 2016. https://doi.org/10.1186/s13148-016-0292-4.
14. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet. 2016;9:436–47.
15. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenet. 2015. https://doi.org/10.1186/s13148-015-0148-3.
16. Dugué P-A, Jung C-H, Joo JE, Wang X, Ming Wong E, Makalic E, et al. Smoking and blood DNA methylation: an epigenome-wide association study and assessment of reversibility. Epigenetics. 2019. https://doi.org/10.1080/15592294.2019.1668739.
17. Guida F, Sandanger TM, Castagné R, Campanella G, Polidoro S, Palli D, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. Hum Mol Genet. 2015;24:2349–59.
18. Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, et al. Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. PLoS One. 2016;11(12):0166486.
19. Vidaki A, Planterose Jiménez B, Poggiali B, Kalamara V, van der Gaag KJ, Maas SCE, et al. Targeted DNA methylation analysis and prediction of smoking habits in blood based on massively parallel sequencing. Forensic Sci Int Genet. 2023;65:102878.
20. Bollepalli S, Korhonen T, Kaprio J, Anders S, Ollikainen M. EpiSmokEr: a robust classifier to determine smoking status from DNA methylation data. Epigenomics. 2019;11:1469–86.
21. Maas SCE, Vidaki A, Wilson R, Teumer A, Liu F, van Meurs JBJ, et al. Validated inference of smoking habits from blood with a finite DNA methylation marker set. Eur J Epidemiol. 2019;34(11):1055.
22. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One. 2013;8(5):63812.
23. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics. 2016;8:599–618.
24. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European prospective investigation into cancer and nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22:843–51.
25. 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 Epigenet. 2014. https://doi.org/10.1186/1868-7083-6-4.
26. Alghanim H, Wu W, McCord B. DNA methylation assay based on pyrosequencing for determination of smoking status. Electrophoresis. 2018;39:2806–14.
27. Philibert R, Dogan M, Beach SRH, Mills JA, Long JD. AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. Am J Med Genet B Neuropsychiatr Genet. 2020;183:51–60.
28. Chamberlain JD, Nusslé S, Chapatte L, Kinnaer C, Petrovic D, Pradervand S, et al. Blood DNA methylation signatures of lifestyle exposures: tobacco and alcohol consumption. Clin Epigenet. 2022;14(1):155.
29. Ambroa-Conde A, de Cal MC, Gómez-Tato A, Robinson O, Mosquera-Miguel A, de la Puente M, et al. Inference of tobacco and alcohol consumption habits from DNA methylation analysis of blood. Forensic Sci Int Genet. 2024;70:103022.
30. Belsky DW, Caspi A, Arseneault L, Baccarelli A, Corcoran D, Gao X, et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. Elife. 2020;9:1–56.
31. Hillary RF, McCartney DL, Bernabeu E, Gadd DA, Cheng Y, Chybowska AD, et al. Blood-based epigenome-wide analyses on the prevalence and incidence of nineteen common disease states. medRxiv. 2023. https://doi.org/10.1101/2023.01.10.23284387v1.
32. Imboden M, Wielscher M, Rezwan FI, Amaral AFS, Schaffner E, Jeong A, et al. Epigenome-wide association study of lung function level and its change. Eur Respir J. 2019;54(1):1900457.
33. Reynolds LM, Wan M, Ding J, Taylor JR, Lohman K, Su D, et al. DNA methylation of the aryl hydrocarbon receptor repressor associations with cigarette smoking and subclinical atherosclerosis. Circ Cardiovasc Genet. 2015;8(5):707.
34. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8:389–99.
35. Wani AH, Dahrendorff J, Uddin M. RANDOMIZE: a web server for data randomization. Archiv Proteomics Bioinfo. 2020. https://doi.org/10.1101/2020.04.02.013656.
36. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.
37. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
38. 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 Bioinfo. 2012;13:1–16.
39. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. Genome Biol. 2021;22(1):173.
40. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.
41. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda SV, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49:359–67.

Pośpiech *et al. Clinical Epigenetics*   (2024) 16:147

Page 12 of 12

42. Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, et al. Epigenetic clock for skin and blood cells applied to hutchinson gilford progeria syndrome and ex vivo studies. Aging. 2018;10:1758–75.

43. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging. 2018;10:573–91.

44. McGreevy KM, Radak Z, Torma F, Jokai M, Lu AT, Belsky DW, et al. DNAm-FitAge: biological age indicator incorporating physical fitness. Aging. 2023;15:3904–38.

45. 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.

46. Noroozi R, Ghafouri-Fard S, Pisarek A, Rudnicka J, Spólnicka M, Branicki W, et al. DNA methylation-based age clocks: From age prediction to age reversion. Ageing Res Rev. 2021;68:101314. https://doi.org/10.1016/j.arr.2021.101314.

47. Pereira B, Correia FP, Alves IA, Costa M, Gameiro M, Martins AP, et al. Epigenetic reprogramming as a key to reverse ageing and increase longevity. Ageing Res Rev. 2024;95:102204.

48. Gao X, Zhang Y, Breitling LP, Brenner H. Relationship of tobacco smoking and smoking-related DNA methylation with epigenetic age acceleration. Oncotarget. 2016;7:46878–89.

49. Kayser M, Branicki W, Parson W, Phillips C. Recent advances in Forensic DNA Phenotyping of appearance, ancestry and age. Forensic Sci Int Genet. 2023;65:102870. https://doi.org/10.1016/j.fsigen.2023.102870.

50. Vidaki A, Kayser M. From forensic epigenetics to forensic epigenomics: broadening DNA investigative intelligence. Genome Biol. 2017. https://doi.org/10.1186/s13059-017-1373-1.

51. Grieshober L, Graw S, Barnett MJ, Thornquist MD, Goodman GE, Chen C, et al. AHRR methylation in heavy smokers: Associations with smoking, lung cancer risk, and lung cancer mortality. BMC Cancer. 2020. https://doi.org/10.1186/s12885-020-07407-x.

52. Grieshober L, Graw S, Barnett MJ, Thornquist MD, Goodman GE, Chen C, et al. AHRR methylation in heavy smokers: associations with smoking, lung cancer risk, and lung cancer mortality. BMC Cancer. 2020. https://doi.org/10.1186/s12885-020-07407-x.

53. De Vries M, Van Der Plaat DA, Nedeljkovic I, Nynke Verkaik-Schakel R, Kooistra W, Amin N, et al. From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function. Respir Res. 2018. https://doi.org/10.1186/s12931-018-0904-y.

54. Vogel CFA, Haarmann-Stemmann T. The aryl hydrocarbon receptor repressor – More than a simple feedback inhibitor of AhR signaling: clues for its role in inflammation and cancer. Curr Opin Toxicol. 2017;2:109–19.

55. 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 Epigenet. 2014. https://doi.org/10.1186/1868-7083-6-4.

56. Fragou D, Pakkidi E, Aschner M, Samanidou V, Kovatsi L. Smoking and DNA methylation: correlation of methylation with smoking behavior and association with diseases and fetus development following prenatal exposure. Food Chem Toxicol. 2019;129:312–27.

57. Rannug A, Fritsche E. The aryl hydrocarbon receptor and light. Biol Chem. 2006;387:1149–57.

58. Pendergast JS, Yamazaki S. The mammalian circadian system is resistant to dioxin. J Biol Rhythms. 2012;27:156–63.

59. Ziv-Gal A, Flaws JA, Mahoney MM, Miller SR, Zacur HA, Gallicchio L. Genetic polymorphisms in the aryl hydrocarbon receptor-signaling pathway and sleep disturbances in middle-aged women. Sleep Med. 2013;14:883–7.

60. Kresovich JK, Garval EL, Martinez Lopez AM, Xu Z, Niehoff NM, White AJ, et al. Associations of body composition and physical activity level with multiple measures of epigenetic age acceleration. Am J Epidemiol. 2021;190:984–93. https://doi.org/10.1093/aje/kwaa251.

61. Fox FAU, Liu D, Breteler MMB, Aziz NA. Physical activity is associated with slower epigenetic ageing-findings from the Rhineland study. Aging Cel. 2023. https://doi.org/10.1111/acel.13828.

62. Pal A, Schneider J, Schlüter K, Steindorf K, Wiskemann J, Rosenberger F, et al. Different endurance exercises modulate NK cell cytotoxic and inhibiting receptors. Eur J Appl Physiol. 2021;121(3379):3387.

63. Schenk A, Joisten N, Walzik D, Koliamitra C, Schoser D, Bloch W, et al. Acute exercise impacts AhR and PD-1 levels of CD8+ T-cells—exploratory results from a randomized cross-over trial comparing endurance versus resistance exercise. Eur J Appl Physiol. 2021;121(637):644.

64. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010;5:1315–6.

65. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med. 2000;45:23–41.

66. Sikdar S, Joehanes R, Joubert BR, Xu CJ, Vives-Usano M, Rezwan FI, et al. Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. Epigenomics. 2019;11:1487–500.

67. Christiansen C, Castillo-Fernandez JE, Domingo-Relloso A, Zhao W, El-Sayed Moustafa JS, Tsai PC, et al. Novel DNA methylation signatures of tobacco smoking with trans-ethnic effects. Clin Epigenet. 2021. https://doi.org/10.1186/s13148-021-01018-4.

68. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics. 2016;8:599–618.

69. Domingo-Relloso A, Riffo-Campos AL, Haack K, Rentero-Garrido P, Ladd-Acosta C, Fallin DM, et al. Cadmium, smoking, and human blood DNA methylation profiles in adults from the strong heart study. Environ Health Perspect. 2020. https://doi.org/10.1289/EHP6345.

70. Hillary RF, McCartney DL, Bernabeu E, Gadd DA, Cheng Y, Chybowska AD, et al. Blood-based epigenome-wide analyses on the prevalence and incidence of nineteen common disease states. medrxiv. 2023. https://doi.org/10.1101/2023.01.10.23284387v1.

71. Witt SH, Frank J, Gilles M, Lang M, Treutlein J, Streit F, et al. Impact on birth weight of maternal smoking throughout pregnancy mediated by DNA methylation. BMC Genomics. 2018;19(1):290.

72. Pośpiech E, Pisarek A, Rudnicka J, Noroozi R, Boroń M, Masny A, et al. Introduction of a multiplex amplicon sequencing assay to quantify DNA methylation in target cytosine markers underlying four selected epigenetic clocks. Clin Epigenet. 2023. https://doi.org/10.1186/s13148-023-01545-2.

73. Shenker NS, Ueland PM, Polidoro S, Van Veldhoven K, Ricceri F, Brown R, et al. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology. 2013;24:712–6.

74. Hulls PM, de Vocht F, Bao Y, Relton CL, Martin RM, Richmond RC. DNA methylation signature of passive smoke exposure is less pronounced than active smoking: the Understanding Society study. Environ Res. 2020;190:109971. https://doi.org/10.1016/j.envres.2020.109971.

75. Frommlet F, Bogdan M, Ramsey D. Phenotypes and Genotypes. London: Springer, 2016, https://doi.org/10.1007/978-1-4471-5310-8

76. Li C, Li H. Network-constrained regularization and variable selection for analysis of genomic data. Bioinformatics. 2008;24:1175–82.

77. Zhou F, Ren J, Lu X, Ma S, Wu C. Gene–Environment Interaction: A Variable Selection Perspective. Methods in Molecular Biology. Humana Press Inc.; 2021 191–223.

Publisher's Note