



OPEN

## Pre-diagnostic DNA methylation in blood leucocytes in cutaneous melanoma; a nested case–control study within the Norwegian Women and Cancer cohort

Christian M. Page<sup>1,2,3✉</sup>, Therese H. Nøst<sup>4,5</sup>, Vera Djordjilović<sup>6</sup>, Magne Thoresen<sup>7</sup>, Arnaldo Frigessi<sup>1,7</sup>, Torkjel M. Sandanger<sup>4</sup> & Marit B. Veierød<sup>7</sup>

The prognosis of cutaneous melanoma depends on early detection, and good biomarkers for melanoma risk may provide a valuable tool to detect melanoma development at a pre-clinical stage. By studying the epigenetic profile in pre-diagnostic blood samples of melanoma cases and cancer free controls, we aimed to identify DNA methylation sites conferring melanoma risk. DNA methylation was measured at 775,528 CpG sites using the Illumina EPIC array in whole blood in incident melanoma cases (n = 183) and matched cancer-free controls (n = 183) in the Norwegian Women and Cancer cohort. Phenotypic information and ultraviolet radiation exposure were obtained from questionnaires. Epigenome wide association (EWAS) was analyzed in future melanoma cases and controls with conditional logistic regression, with correction for multiple testing using the false discovery rate (FDR). We extended the analysis by including a public data set on melanoma (GSE120878), and combining these different data sets using a version of covariate modulated FDR (AdaPT). The analysis on future melanoma cases and controls did not identify any genome wide significant CpG sites ( $0.85 \leq p_{\text{adj}} \leq 0.99$ ). In the restricted AdaPT analysis, 7 CpG sites were suggestive at the FDR level of 0.15. These CpG sites may potentially be used as pre-diagnostic biomarkers of melanoma risk.

Incidence of cutaneous melanoma, the most aggressive and lethal form of skin cancer, continues to increase in fair-skinned populations<sup>1</sup>. Ultraviolet radiation (UVR) exposure is the main risk factor of melanoma<sup>2</sup>. By sequencing lesions in different evolutionary stages, Shain et al.<sup>3</sup> found strong signature UVR mutations detectable at all stages, from benign lesions to invasive melanoma<sup>3</sup>. In some melanoma patients, the tumour suppressor gene *TP53* has occasionally shown mutations with a UVR-signature<sup>4,5</sup>. Still, the most common mutations found in melanomas are the *BRAF* V600E and V600K mutations<sup>6</sup> although these mutations do not carry the typical UVR signature<sup>7</sup>. The biological pathway from UVR exposure to melanoma onset is not yet fully explained, and could be due to epigenetic changes after UVR exposure.

Epigenetic modifications, such as DNA methylation (DNAm), have been associated with all cancers, including melanoma<sup>8</sup>. However, to our knowledge, previous studies on the epigenetics of melanoma have mostly been based on post-diagnostic sampling<sup>9</sup>. Epigenetic biomarkers have been proposed for both survival and response to treatment in melanoma patients<sup>10,11</sup>. However, the predictive value of these markers for melanoma risk itself has not been evaluated. Since most melanomas appear without association with a precursor lesion<sup>12</sup>, evaluation of pre-diagnostic markers should be based on samples of skin, blood, or saliva stored in biobanks. The DNA from these mixed tissues such as blood or saliva could be confounded by the composition of the different tissues.

<sup>1</sup>Oslo Centre for Biostatistics and Epidemiology, Division for Research Support, Oslo University Hospital, Oslo, Norway. <sup>2</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway. <sup>3</sup>Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway. <sup>4</sup>Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway. <sup>5</sup>K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. <sup>6</sup>Department of Economics, Ca' Foscari University of Venice, Venice, Italy. <sup>7</sup>Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway. ✉email: christmp@math.uio.no

Based on epigenetic data, it is possible to estimate the relative contribution from each cell type observed in the tissue mix<sup>13</sup>.

The population-based Norwegian Women and Cancer (NOWAC) cohort study<sup>14</sup> has been used to study the importance of host factors and UVR exposure in melanoma risk<sup>15–17</sup>, and to identify pre-diagnostic epigenetic markers for lung and breast cancer<sup>18–20</sup>. In a nested case–control study within the NOWAC cohort, we aimed to identify biomarkers of melanoma risk in pre-diagnostic blood samples of melanoma cases and cancer free controls, in an epigenome wide association study (EWAS), as well as a subset EWAS on case specific characteristics. To complement our analyses, EWAS was also performed in an open source data set from an independent study on melanoma<sup>21</sup> and combined with our results.

## Materials and methods

**Material (NOWAC).** The NOWAC cohort includes over 172,000 women aged 30–70 years at recruitment in 1991–2006 (response 54%)<sup>14</sup>. Information on host characteristics and lifestyle factors was collected through baseline questionnaires and up to two follow-up questionnaires. The NOWAC study has high external validity, with no major selection bias<sup>14,22</sup>. Approximately 50,000 women (46–63 years) constitute the post genome cohort within NOWAC and donated a blood sample at inclusion or at the second follow-up (2003–2006)<sup>23</sup>. By using the unique identity number of Norwegian citizens, NOWAC is linked to the Cancer Registry of Norway (CRN) for follow-up of cancer incidence and vital status. We included all incident melanoma cases ( $n = 183$ ) with an isolated DNA and RNA sample in the biobank per December 31, 2013, and matched each case with one cancer free control, based on time since blood sampling and year of birth (1943–1947, 1948–1952, 1953–1957). The Norwegian Malignant Melanoma Registry (NMMR) was established under the CRN in 2008, and information on tumour thickness for incident cases since 2008 was obtained from the NMMR. For melanoma cases diagnosed before 2008, information on tumour thickness was extracted manually by the CRN's experienced melanoma registrars from histopathological reports in the CRN archive<sup>24</sup>.

All participants gave written informed consent and the Medical Ethical Committees of North Norway has approved the NOWAC study, the storage of human biological material, as well as this study (2016/976/REK Nord). All methods in this study were performed in accordance with the relevant ethical guidelines and regulations.

**DNA methylation.** Details of the DNAm quality control has been described elsewhere<sup>25</sup>. Briefly, DNA were treated with bi-sulfite and hybridised to the Illumina Infinium MethylationEPIC array according to the manufacturer's protocol. Background subtraction and control normalization were performed with *minfi* to reduce background noise and dye bias<sup>26</sup>. Type I and Type II probes were normalized using the Beta mixture quantile normalization method from the *wateRmelon* R-package<sup>27</sup>. After quality control, 775 528 CpG probes remained in the data set. White blood cell composition was estimated using the Houseman algorithm<sup>13,28</sup>.

To complement our analysis, we included an open source data set from an independent study by Conway et al.<sup>21</sup> (hereafter referred to as the GSE120878 study) which compared the epigenetic profiles of melanoma biopsies ( $n = 89$ ) and nevi biopsies ( $n = 73$ ), all from suspected melanoma biopsies from different patients using logistic regression. Their DNA methylation data were deposited at the GEO database in April 2019 (accession number GSE120878)<sup>21</sup>. In this study, DNA methylation was measured on the Illumina Infinium HumanMethylation450 BeadChip array and processed with the *minfi* R package.

**Covariates.** Baseline variables include age at recruitment, birth cohort and recruitment year. Age at blood sample and time in freezer were recorded. Participants reported, by questionnaire, education ( $\leq 10$ , 11–13,  $\geq 14$  years), smoking (never, former, current smoker), hair color (dark brown/black, brown, blond/yellow, red), freckling after sunbathing (yes, no), and the number of asymmetric nevi  $> 5$  mm on the legs (0, 1, 2–3, 4–6, 7–12, 13–24,  $\geq 25$ ; categorized as 0, 1,  $\geq 2$ ). On the basis of average ambient ultraviolet radiation<sup>29</sup>, region of residence (latitudes  $70^\circ$ – $58^\circ$ ) was categorized as low UVR exposure (north Norway), medium–low (central Norway), medium (southwest Norway), and highest UVR exposure (southeast Norway)<sup>30</sup>. In the baseline and follow-up questionnaires, participants reported history of severe sunburns (never, 1, 2–3, 4–5,  $\geq 6$  times/year), average number of weeks per year spent on sunbathing vacations (never, 1, 2–3, 4–6,  $\geq 7$  weeks/year), and average use of an indoor tanning device (never; rarely; 1, 2, or 3–4 times/month;  $> 1$  time/week) in childhood ( $\leq 9$  years), adolescence (10–19 years), and adulthood ( $> 19$  years)<sup>30</sup>. The reported frequencies of sunburns, sunbathing vacations and indoor tanning were transformed and multiplied by the length of each interval for each questionnaire<sup>15</sup>. The participants were then classified into five categories; non-exposed and quartiles. To capture the tail of the distribution, the upper quartile was further divided into two equally sized groups (i.e. six categories in total) as described in Page et al.<sup>25</sup> Cumulative UVR exposure was constructed by summarizing the categories (i.e. scores 0–5) for indoor tanning and sunbathing vacations<sup>15</sup>. Reproducibility of melanoma risk factors in the NOWAC questionnaire was good (kappa/intraclass correlation coefficient 0.49–0.77) and independent of age and education<sup>31</sup>.

**Statistical methods.** Conditional logistic regression was used to study the association between future status of melanoma as the outcome and white blood cell composition as continuous exposure, accounting for time to diagnosis and potential confounders: hair colour, nevi, and UVR exposure. To minimize technical variation and capture unmeasured confounding, we constructed surrogate variables using the *sva* package in R<sup>32–34</sup>. The surrogate variables were constructed as orthogonal decompositions of the residuals after projecting melanoma status on the DNA methylation data matrix<sup>35</sup>. We used conditional logistic regression, with control for the matching variables (age at blood sample and time in freezer) to assess the associations between future melanoma as the outcome and DNA methylation, adjusting for lifetime history of sunburns (as an indicator of severe UVR

exposure)<sup>15</sup>, hair colour (the best measure of skin sensitivity to UVR exposure in the NOWAC cohort)<sup>17,35</sup>, and surrogate variables as potential confounders.

To control for multiple testing, we used the false discovery rate (FDR) procedure of Benjamini and Hochberg<sup>36</sup>. The genes annotated to the top 2000 CpG sites in our main model were included in an enrichment analysis, using the *Enricher* web interface<sup>37</sup>.

An EWAS was run using logistic regression without any covariate adjustments in the GSE120878 dataset, and the log p-values included as covariates in an adaptive multiple testing FDR method called AdaPT<sup>38</sup> when correcting our main model for multiple testing. This method is based on the covariate modulated FDR (cmFDR) proposed by Ferkingstad et al.<sup>39</sup> which weight the FDR significance by information from the new data set. The combined AdaPT analysis was restricted to CpG sites with a nominal p-value  $< 10^{-10}$  ( $N_{\text{CpG}} = 2176$ ) in the GSE120878 EWAS and the respective log p-values from these sites were included as side information in AdaPT, for each CpG, respectively.

A prediction model was trained on the same CpG sites ( $p\text{-value} < 10^{-10}$ ,  $N_{\text{CpG}} = 2176$ ) from the GSE120878 data, using a regression and decision tree algorithm<sup>40</sup> similar to that of Onwuka et al.<sup>41</sup>. The prediction model was applied on the NOWAC data set of incident melanoma cases and controls.

Lastly, we performed an EWAS including only the melanoma cases in linear regressions with log transformed tumour thickness as the outcome, adjusting for lifetime history of sunburns and hair colour.

**Institutional review board statement.** The Medical Ethical Committees of North Norway has approved the NOWAC study, the storage of human biological material, as well as this sub-study (2016/976/REK Nord). All methods in this study were performed in accordance with the relevant ethical guidelines and regulations.

**Informed consent statement.** All participants gave written informed consent.

## Results

Baseline characteristics of the cases and controls are presented in Table 1. Having higher education, being a non-smoker, having blond/yellow/red hair, freckling when sunbathing, and large asymmetric nevi on the legs were more common in the melanoma cases than in the controls. Compared to controls, melanoma cases reported more UVR exposure: lower proportion living in the region with low ambient UVR, the lower proportion experiencing no sunburns, the higher proportions in the highest categories of sunbathing vacations, and lower proportions in the two lowest categories of indoor tanning and cumulative UVR exposure (Table 1). Mean age at melanoma diagnosis was 60.2 years (range 49–70) and mean time from blood sampling to diagnosis was 4.4 years (range 0–9.7 years) (Table 1). Estimated cell-type proportions were similar in cases and controls ( $0.09 \leq p \leq 0.94$ ) (Supplementary Table S1). None of the white blood cell fractions were significantly differently distributed between the melanoma cases and controls after adjustment for time to diagnosis ( $0.14 \leq p_{\text{adj}} \leq 0.56$ ) (Supplementary Table S2).

We did not find any CpGs significantly associated with melanoma risk in the genome-wide analyses ( $0.85 \leq p_{\text{adj}} \leq 0.99$ ). The top 10 CpG sites are listed in Table 2. The estimated odds ratios (ORs) genome wide were in equal proportions in both directions, indicating no global loss of methylation. The pathway enrichment analysis of the top 2000 CpGs did not identify any pathways previously reported for melanoma (Supplementary Table S3).

In the combined AdaPT analysis, after adjusting our findings with the log p-value from the GSE120878 EWAS, seven CpG sites from the NOWAC study had an FDR adjusted p-value below 0.15 (Table 3). The distribution of the ORs for these seven CpGs were shifted towards higher risk, with 5/7 CpGs having an OR above 1, as compared to the entire set of OR in our main EWAS, where the OR was in equal proportions in both directions.

The prediction model trained on the GSE120878 data set, did not predict melanoma status well: while 48% were predicted true positive, 49% were predicted as false positive, with the true negative only 1% and the false negative 2%. We did not find any significant CpG associations in the EWAS analysis of DNAm and tumour thickness in the melanoma cases ( $0.86 \leq p_{\text{adj}} \leq 0.99$ ).

## Discussion

We compared DNAm profiles of incident melanoma cases to healthy controls, to identify potential biomarkers for melanoma risk. We did not identify any genome wide significant CpGs related to melanoma risk. However, by combining different data sources, weighing the FDR adjustment, we identified seven potentially differentially methylated CpG sites associated to incident melanoma, all previously associated to melanoma in a case-control study<sup>21</sup>.

Two of the top 10 genes identified in our EWAS (Table 2) have previously been associated to melanoma; *RSF1*<sup>42</sup> and *NTN4*<sup>43</sup>. However, they have been associated with more advanced stages in melanoma from case only studies and cell lines, and not with melanoma risk. We observed an equal number of effect sizes in both directions while the proportion of hypomethylation was larger in the GSE120878 study (50% vs ~57.6%), indicating a global loss of methylation in melanoma biopsies, which was not observed in the pre-diagnostic samples. This indicates that the loss of methylation observed in cancers may be a consequence of the disease, and not its cause. The log-odds was consistently over ten times larger in the study including samples from prevalent cases as compared to this pre-diagnostic study, which is also to be expected, given the differences in sample tissues in the two studies.

None of the top genes found in our primary analysis were associated to melanoma risk in the largest GWAS of melanoma to date<sup>14</sup>, which included almost 37,000 melanoma cases and ten times as many controls. Of the CpGs associated with melanoma risk in Table 3, two are associated to genes *MIR196B* and *SH3RF3*, which have been observed differentially expressed in sun exposed skin, as compared to non-exposed skin<sup>45</sup>. Given the prominent role of UVR exposure in melanoma risk, this is a potentially interesting finding that should be

	Melanoma cases, n = 183	Controls, n = 183
Age at recruitment (years), mean (SD)	48.0 (6.55)	47.9 (6.53)
<b>Birth cohort, n (%)</b>		
1943–1947	73 (39.9)	73 (39.9)
1948–1952	70 (38.3)	70 (38.3)
1953–1957	40 (21.9)	40 (21.9)
<b>Recruitment year, n (%)</b>		
1991–1992	80 (43.7)	81 (44.3)
1996–1997	29 (15.8)	29 (15.8)
Age at blood sample (years), mean (SD)	55.7 (4.24)	55.7 (4.24)
Time in freezer (years), mean (SD)	11.23 (0.97)	11.21 (0.95)
<b>Education (years), n (%)</b>		
≤ 10	43 (23.5)	54 (29.5)
11–13	57 (31.1)	55 (30.1)
≥ 14	78 (42.6)	67 (36.6)
Missing	5 (2.7)	7 (3.8)
<b>Smoking, n (%)</b>		
Non-smoker	77 (42.1)	61 (33.3)
Former	71 (38.8)	73 (39.9)
Current	33 (18.0)	46 (25.1)
Missing	2 (1.1)	3 (1.6)
<b>Hair color, n (%)</b>		
Black/dark brown	23 (12.6)	33 (18.0)
Brown	53 (29.0)	75 (41.0)
Blond/yellow	82 (44.8)	60 (32.8)
Red	15 (8.2)	5 (2.7)
Missing	10 (5.5)	10 (5.5)
<b>Freckling when sunbathing, n (%)</b>		
Yes	78 (42.6)	67 (36.6)
No	103 (56.3)	111 (60.7)
Missing	2 (1.1)	5 (2.7)
<b>No. of large asymmetric nevi on legs, n (%)</b>		
0	136 (74.3)	147 (80.3)
1	11 (6.0)	11 (6.0)
≥ 2	20 (10.9)	8 (4.4)
Missing	16 (8.7)	17 (9.3)
<b>Residential ambient UVR</b>		
Low (northern Norway)	23 (12.6)	33 (18.0)
Medium–low (central Norway)	27 (14.8)	22 (12.0)
Medium (south-west Norway)	37 (20.2)	26 (14.2)
Highest (south-east Norway)	96 (52.5)	102 (55.7)
Missing	0	0
<b>Lifetime no. of sunburns<sup>a</sup></b>		
0	15 (8.2)	24 (13.1)
1–20	39 (21.3)	42 (23.0)
21–30	34 (18.6)	24 (13.1)
31–47	45 (24.6)	45 (24.6)
48–58	26 (14.2)	21 (11.5)
59+	20 (10.9)	20 (10.9)
Missing	4 (2.2)	7 (3.8)
<b>Lifetime no. of sunbathing vacations<sup>a</sup></b>		
0	14 (7.7)	14 (7.7)
1–29	45 (24.6)	44 (24.0)
30–62	32 (17.5)	44 (24.0)
63–104	39 (21.3)	41 (22.4)
105–143	21 (11.5)	16 (8.7)
144+	26 (14.2)	17 (9.3)
Continued		

	Melanoma cases, n = 183	Controls, n = 183
Missing	6 (3.3)	7 (3.8)
<b>Lifetime no. of indoor tanning sessions<sup>a</sup></b>		
0	46 (25.1)	49 (26.8)
1–14	30 (16.4)	32 (17.5)
15–24	50 (27.3)	38 (20.8)
25–120	25 (13.7)	32 (17.5)
121–418	14 (7.7)	18 (9.8)
419+	17 (9.3)	14 (7.7)
Missing	1 (0.6)	0 (0)
<b>Cumulative UVR<sup>b</sup></b>		
0	8 (4.4)	10 (5.5)
1	15 (8.2)	18 (9.8)
2	26 (14.2)	20 (10.9)
3	15 (8.2)	31 (16.9)
4	31 (16.9)	26 (14.2)
5	34 (18.6)	25 (13.7)
6	24 (13.1)	24 (13.1)
7	15 (8.2)	15 (8.2)
8	8 (4.4)	7 (3.8)
9	3 (1.6)	5 (2.7)
10	4 (2.2)	2 (1.1)
Age at diagnosis (years), mean (SD)	60.2 (4.89)	N/A
Time from blood sampling to diagnosis (years), mean (SD)	4.44 (2.53)	N/A
Tumor thickness (mm), median (IQR)	0.9 (0.5,1.5)	N/A
<b>Tumor thickness (mm), n (%)</b>		
T1 ( $\leq 1$ mm)	109 (59.6)	N/A
T2 (1.01–2.0)	32 (17.5)	N/A
T3 (2.01–4.0)	22 (12.0)	N/A
T4 ( $> 4$ )	6 (3.3)	N/A
Missing	14 (7.7)	N/A

**Table 1.** Characteristics of the cases and their matched controls. Matched by age at blood sample and time in freezer. <sup>a</sup>Categorized in six categories: non-exposed and quartiles, with the upper quartile further divided into two equally sized groups. <sup>b</sup>Sunbathing vacations and indoor tanning.

CpG ID	Position	Gene	OR	95% CI	p-value	FDR adj. p-value
cg09396032	16q24.1	GSE1	0.000349	(0.00002, 0.0054)	1.672e–06	0.85
cg05900127	6q23.2	TBPL1	42.52	(10.61, 170.45)	8.709e–06	0.85
cg19014186	17q23.2	NACA2	7.03	(3.33, 14.84)	1.661e–05	0.85
cg27080643	16p13.3	SLX4	8.94	(3.87, 20.65)	1.709e–05	0.85
cg24287218	6p21.32	HLA-DPA1	10.38	(4.20, 25.66)	2.066e–05	0.85
cg10964388	12q22	NTN4	16.28	(5.50, 48.20)	2.248e–05	0.85
cg17459743	11q14.1	RSF1	0.13	(0.06, 0.29)	2.268e–05	0.85
cg21150053	5p14.1	CDH9	9.78	(4.02, 23.78)	2.330e–05	0.85
cg03347745	14q23.3	N/A	8.94	(3.79, 21.06)	2.538e–05	0.85
cg02804047	6p21.32	C6orf10	0.12	(0.05, 0.28)	2.709e–05	0.85

**Table 2.** The top 10 CpG sites in the conditional logistic regression analyses of CpG sites for melanoma cases (n = 183) versus controls (n = 183). Adjusted for lifetime number of sunburns, hair colour, and the 10 first surrogate variables. OR odds ratio, CI confidence interval, FDR False discovery rate.

followed up. Further among the genes indicated, the *HOXA9-HOXA10* cluster has been observed differentially expressed in multiple cancers<sup>46</sup>, and upregulation of *HOXA9* is related to poor survival in melanoma cases<sup>47</sup>. Since the analysis was informed using findings from a previous case–control study on melanoma, all findings in this analysis have previously been associated with melanoma.

CpG ID	Position	Gene	OR	95% CI	p value	cmFDR adj. p-value
cg01904393	12q13.13	SLC4A8 <sup>a</sup>	4.90	(1.99,12.08)	3.69e-03	0.15
cg25961816	6q22.31	N/A	3.35	(1.65, 6.93)	5.34e-03	0.15
cg05250768	7p15.2	MIR196B/HOXA9-HOXA10 <sup>a</sup>	0.16	(0.06, 0.40)	1.12e-03	0.15
cg18919097	3q26.1	C3orf57	11.25	(2.98,42.45)	2.7e-03	0.15
cg03505866	14q24.1	KIAA0247/SUSD6	3.86	(1.78, 8.31)	4.0e-03	0.15
cg14966782	2q12.3	SH3RF3	2.75	(1.69,4.48)	6.2e-04	0.15
cg00974255	17q12	N/A	0.15	(0.06, 0.36)	5.8e-04	0.15

**Table 3.** The seven CpG sites rejected at FDR level of 0.15 with *AdaPT* (covariate modulated FDR; cmFDR). p value from the conditional logistic regression analyses and cmFDR adjusted with *AdaPT* using log p-values from the GSE120878 analysis. Adjusted for lifetime number of sunburns, hair colour, and the 10 first surrogate variables. OR odds ratio, CI confidence interval, FDR False discovery rate, cmFDR covariate modulated FDR. <sup>a</sup>GRCh37/hg19 (nearest gene).

None of the white blood cell fractions were significantly differently distributed between the melanoma cases and controls, even after adjusting for time to diagnosis. This indicated that the cell type composition would not be a confounder for disease status, and was thus not adjusted for in our analysis.

The discovery of pre-diagnostic biomarkers relies on a large number of samples with biological material stored in biobanks, since the future cancer status of each participant is unknown. Biobanks of the size needed for this type of incident sampling, are almost exclusively storing biological samples derived from blood. The use of blood leukocytes may explain some of the poor performance for the multi-CpG prediction for melanoma, which did not predict case status with high accuracy. Additionally, the tissue differences need to be kept in mind when comparing the results between pre-diagnostic blood samples and tissue specific cancer samples. Using results from cancer tissue to inform the FDR correction could help detect cancer like signals in blood samples early on in the disease. Circulating blood leukocytes are constantly in contact with all organ systems in the body, and exposed to the same environment, thus, weak signals from the environmental exposure can often be detected in blood leukocytes<sup>48</sup>. Additionally, pre-clinical tumors are likely to shed DNA fragments in the blood stream, which can influence the DNAm signature in the blood sample, and DNAm isolated from whole blood may then contain weak cancer specific signals.

Being nested in the NOWAC cohort, this study benefits from a large population-based cohort with well documented case information and prospective baseline information on major risk factors for melanoma, such as UVR exposure, but the pre-diagnostic biological material was limited to whole blood and with a limited sample size. The distribution of T categories was not even across the cases, as it can be in a selected clinical sample, but reflective of what is found in the general population (i.e. more T1 than T4 melanomas).

NOWAC is a female only cohort, while GSE120878 included both sexes. Previous cancer studies have included only one sex, either only females<sup>14</sup> or males<sup>49</sup>. Studies addressing association between UV exposure and melanoma, found no interaction between sun exposure and sex<sup>50</sup>. The GSE120878 data set was balanced with respect to sex ratio, and the p-value for any sex differences between the groups not significant<sup>21</sup>.

The lack of genotype information in the cohort is a limitation. Multiple genetic markers have been found to increase melanoma risk<sup>44</sup>, most notable variants in the CDKN2A gene<sup>51</sup>, however, the consent in NOWAC did not open for genotyping of the participants.

We find that the use of covariate modulated FRD methods, like *AdaPT*, is a good way of combining our results with public data from a different source.

## Conclusion

No epigenome-wide significant associations to melanoma risk was found, but 7 CpGs identified by combining data and previous knowledge was suggestive of melanoma risk. Future melanoma status was not well predicted in this study, however, using a more targeted tissue, such as skin biopsies could have resulted in more informative epigenetic markers for melanoma risk.

## Data availability

The DNA methylation data generated and/or analysed in the current study can be accessed upon reasonable request to the originating cohort. Access will be conditional on adherence to both local and national ethical and security policy. R codes used for the analyses presented in the paper are available upon request.

Received: 18 November 2021; Accepted: 16 August 2022

Published online: 20 August 2022

## References

- Whiteman, D. C., Green, A. C. & Olsen, C. M. The growing burden of invasive melanoma: Projections of incidence rates and numbers of new cases in six susceptible populations through 2031. *J. Investig. Dermatol.* **136**, 1161–1171. <https://doi.org/10.1016/j.jid.2016.01.035> (2016).
- Berwick, M. *et al.* Melanoma epidemiology and prevention. *Cancer Treat. Res.* **167**, 17–49. [https://doi.org/10.1007/978-3-319-22539-5\\_2](https://doi.org/10.1007/978-3-319-22539-5_2) (2016).



3. Shain, A. H. *et al.* The genetic evolution of melanoma from precursor lesions. *N. Engl. J. Med.* **373**, 1926–1936. <https://doi.org/10.1056/NEJMoa1502583> (2015).
4. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421. <https://doi.org/10.1038/nature12477> (2013).
5. Arozarena, I. & Wellbrock, C. Overcoming resistance to BRAF inhibitors. *Ann. Transl. Med.* **5**, 387. <https://doi.org/10.21037/atm.2017.06.09> (2017).
6. Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251–263. <https://doi.org/10.1016/j.cell.2012.06.024> (2012).
7. Bauer, J. *et al.* BRAF mutations in cutaneous melanoma are independently associated with age, anatomic site of the primary tumor, and the degree of solar elastosis at the primary tumor site. *Pigment Cell Melanoma Res.* **24**, 345–351. <https://doi.org/10.1111/j.1755-148X.2011.00837.x> (2011).
8. Jones, P. A. DNA methylation and cancer. *Oncogene* **21**, 5358–5360. <https://doi.org/10.1038/sj.onc.1205597> (2002).
9. Van Laar, R., Lincoln, M. & Van Laar, B. Development and validation of a plasma-based melanoma biomarker suitable for clinical use. *Br. J. Cancer* **118**, 857–866. <https://doi.org/10.1038/bjc.2017.477> (2018).
10. Guo, Y., Long, J. & Lei, S. Promoter methylation as biomarkers for diagnosis of melanoma: A systematic review and meta-analysis. *J. Cell Physiol.* **234**, 7356–7367. <https://doi.org/10.1002/jcp.27495> (2019).
11. Micevic, G., Theodosakis, N. & Bosenberg, M. Aberrant DNA methylation in melanoma: Biomarker and therapeutic opportunities. *Clin. Epigenet.* **9**, 34. <https://doi.org/10.1186/s13148-017-0332-8> (2017).
12. Shitara, D. *et al.* Nevus-associated melanomas: Clinicopathologic features. *Am. J. Clin. Pathol.* **142**, 485–491. <https://doi.org/10.1309/AJCP4L5CJGKTJVDD> (2014).
13. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* **13**, 86. <https://doi.org/10.1186/1471-2105-13-86> (2012).
14. Lund, E. *et al.* Cohort profile: The Norwegian women and cancer study—NOWAC-Kvinner og kreft. *Int. J. Epidemiol.* **37**, 36–41. <https://doi.org/10.1093/ije/dym137> (2008).
15. Ghiasvand, R. *et al.* Indoor tanning and melanoma risk: Long-term evidence from a prospective population-based cohort study. *Am. J. Epidemiol.* **185**, 147–156. <https://doi.org/10.1093/aje/kww148> (2017).
16. Ghiasvand, R., Weiderpass, E., Green, A. C., Lund, E. & Veierod, M. B. Sunscreen use and subsequent melanoma risk: A population-based cohort study. *J. Clin. Oncol.* **34**, 3976–3983. <https://doi.org/10.1200/JCO.2016.67.5934> (2016).
17. Veierod, M. B. *et al.* A prospective study of pigmentation, sun exposure, and risk of cutaneous malignant melanoma in women. *J. Natl. Cancer Inst.* **95**, 1530–1538. <https://doi.org/10.1093/jnci/djg075> (2003).
18. Fasanelli, F. *et al.* Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nat. Commun.* **6**, 10192. <https://doi.org/10.1038/ncomms10192> (2015).
19. van Veldhoven, K. *et al.* Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. *Clin. Epigenet.* **7**, 67. <https://doi.org/10.1186/s13148-015-0104-2> (2015).
20. Guida, F. *et al.* Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum. Mol. Genet.* **24**, 2349–2359. <https://doi.org/10.1093/hmg/ddu751> (2015).
21. Conway, K. *et al.* Identification of a robust methylation classifier for cutaneous melanoma diagnosis. *J. Investig. Dermatol.* **139**, 1349–1361. <https://doi.org/10.1016/j.jid.2018.11.024> (2019).
22. Lund, E. *et al.* External validity in a population-based national prospective study—The Norwegian Women and Cancer Study (NOWAC). *Cancer Causes Control* **14**, 1001–1008. <https://doi.org/10.1023/b:caco.0000007982.18311.2e> (2003).
23. Dumeaux, V. *et al.* Deciphering normal blood gene expression variation—The NOWAC postgenome study. *PLoS Genet.* **6**, e1000873. <https://doi.org/10.1371/journal.pgen.1000873> (2010).
24. Ghiasvand, R. *et al.* Patient age and risk of recurrence of primary melanoma at high risk of spread. *Br. J. Dermatol.* **184**, 566–568. <https://doi.org/10.1111/bjd.19601> (2021).
25. Page, C. M. *et al.* Lifetime ultraviolet radiation exposure and DNA methylation in blood leukocytes: The Norwegian women and cancer study. *Sci. Rep.* **10**, 4521. <https://doi.org/10.1038/s41598-020-61430-3> (2020).
26. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369. <https://doi.org/10.1093/bioinformatics/btu049> (2014).
27. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genom.* **14**, 293. <https://doi.org/10.1186/1471-2164-14-293> (2013).
28. Reinius, L. E. *et al.* Differential DNA methylation in purified human blood cells: Implications for cell lineage and studies on disease susceptibility. *PLoS One* **7**, e41361. <https://doi.org/10.1371/journal.pone.0041361> (2012).
29. Edvardsen, K. *et al.* Vitamin D-effective solar UV radiation, dietary vitamin D and breast cancer risk. *Int. J. Cancer* **128**, 1425–1433. <https://doi.org/10.1002/ijc.25463> (2011).
30. Ghiasvand, R., Lund, E., Edvardsen, K., Weiderpass, E. & Veierod, M. B. Prevalence and trends of sunscreen use and sunburn among Norwegian women. *Br. J. Dermatol.* **172**, 475–483. <https://doi.org/10.1111/bjd.13434> (2015).
31. Veierod, M. B., Parr, C. L., Lund, E. & Hjartaker, A. Reproducibility of self-reported melanoma risk factors in a large cohort study of Norwegian women. *Melanoma Res.* **18**, 1–9. <https://doi.org/10.1097/CMR.0b013e3282f20d2> (2008).
32. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883. <https://doi.org/10.1093/bioinformatics/bts034> (2012).
33. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735. <https://doi.org/10.1371/journal.pgen.0030161> (2007).
34. R Core Team. R: A language and environment for statistical computing. (2013).
35. Veierod, M. B., Adami, H. O., Lund, E., Armstrong, B. K. & Weiderpass, E. Sun and solarium exposure and melanoma risk: Effects of age, pigmentation characteristics, and nevi. *Cancer Epidemiol. Biomark. Prev.* **19**, 111–120. <https://doi.org/10.1158/1055-9965.EPI-09-0567> (2010).
36. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate—A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x> (1995).
37. Kuleshov, M. V. *et al.* Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–97. <https://doi.org/10.1093/nar/gkw377> (2016).
38. Lei, L. H. & Fithian, W. AdaPT: An interactive procedure for multiple testing with side information. *J. R. Stat. Soc. B* **80**, 649–679. <https://doi.org/10.1111/rssb.12274> (2018).
39. Ferkingstad, E., Frigessi, A., Rue, H., Thorleifsson, G. & Kong, A. Unsupervised Empirical Bayesian Multiple Testing with External Covariates. *Ann Appl Stat* **2**, 714–735. <https://doi.org/10.1214/08-Aoas158> (2008).
40. Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. *Classification and Regression Trees*. (CRC Press, 1984).
41. Onwuka, J. U. *et al.* A panel of DNA methylation signature from peripheral blood may predict colorectal cancer susceptibility. *BMC Cancer* **20**, 692. <https://doi.org/10.1186/s12885-020-07194-5> (2020).
42. He, J., Fu, L. & Li, Q. Rsf1 regulates malignant melanoma cell viability and chemoresistance via NFκB/Bcl2 signaling. *Mol. Med. Rep.* **20**, 3487–3498. <https://doi.org/10.3892/mmr.2019.10610> (2019).
43. Jayachandran, A. *et al.* Identifying and targeting determinants of melanoma cellular invasion. *Oncotarget* **7**, 41186–41202. <https://doi.org/10.18632/oncotarget.9227> (2016).

44. Landi, M. T. *et al.* Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. *Nat. Genet.* **52**, 494–504. <https://doi.org/10.1038/s41588-020-0611-8> (2020).
45. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419. <https://doi.org/10.1126/science.1260419> (2015).
46. Paco, A., de Bessa Garcia, S. A. & Freitas, R. Methylation in HOX clusters and its applications in cancer therapy. *Cells* <https://doi.org/10.3390/cells9071613> (2020).
47. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* <https://doi.org/10.1126/science.aan2507> (2017).
48. Terry, M. B., Delgado-Cruzata, L., Vin-Raviv, N., Wu, H. C. & Santella, R. M. DNA methylation in white blood cells: Association with risk factors in epidemiologic studies. *Epigenetics* **6**, 828–837. <https://doi.org/10.4161/epi.6.7.16500> (2011).
49. Rimm, E. B. *et al.* A prospective study of alcohol consumption and the risk of coronary-disease in men. *Am. J. Epidemiol.* **134**, 785–786 (1991).
50. Liu, F. C. *et al.* Ultraviolet radiation and risk of cutaneous melanoma and squamous cell carcinoma in males and females in the Norwegian Offshore Petroleum Workers cohort. *Am. J. Ind. Med.* **64**, 496–510. <https://doi.org/10.1002/ajim.23240> (2021).
51. Begg, C. B. *et al.* Lifetime risk of melanoma in CDKN2A mutation carriers in a population-based sample. *J. Natl. Cancer Inst.* **97**, 1507–1515. <https://doi.org/10.1093/jnci/dji312> (2005).

## Acknowledgements

We thank all the participants in the NOWAC cohort, who donated blood and answered the questionnaires.

## Author contributions

Conception and design: T.M.S., A.F., M.T., M.B.V. Acquisition of data: T.M.S., T.H.N., M.T., M.B.V. Analysis of data: C.M.P., V.D. Interpretation of data: All authors. Drafted the work: C.M.P., M.B.V. Funding: A.F., M.T., M.B.V. Revising the work: All authors. All authors read and approved the final version of the manuscript.

## Funding

CMP was funded by a grant from the South-East Regional Health Authority (#2076086). The DNA methylation arrays were funded by grants from the Norwegian Research Council (#248804) and UNIFOR Frimed. The NOWAC post-genome cohort study was funded by the ERC advanced grant Transcriptomic in Cancer Epidemiology (ERC-2008-AdG-232997).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-18585-y>.

**Correspondence** and requests for materials should be addressed to C.M.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022