Assessing the role of genome-wide DNA methylation between smoking and risk of lung cancer using repeated measurements: the HUNT study

Yi-Qian Sun,1,2,3* Rebecca C Richmond,4 Matthew Suderman,4 Josine L Min,4 Thomas Battram,4 Arnar Flatberg,1,5 Vidar Beisvag,1,5 Therese Haugdahl Nøst,6,7 Florence Guida,8 Lin Jiang,9 Sissel Gyrid Freim Wahl,1,2 Arnulf Langhammer,10 Frank Skorpen,1 Rosie M Walker,11 Andrew D Bretherick,12 Yanni Zeng,13 Yue Chen,14 Mattias Johansson,8 Torkjel M Sandanger,6 Caroline L Relton4 and Xiao-Mei Mai9

1Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway, 2Department of Pathology, Clinic of Laboratory Medicine, St Olav’s University Hospital, Trondheim, Norway, 3Center for Oral Health Services and Research Mid-Norway (TkMidt), Trondheim, Norway, 4MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK, 5Central Administration, St Olav’s University Hospital, Trondheim, Norway, 6Department of Community Medicine, Faculty of Health Sciences, Arctic University of Norway, Tromsø, Norway, 7K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway, 8Genetic Epidemiology Group, International Agency for Research on Cancer, Lyon, France, 9Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway, 10HUNT Research Centre, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway, 11Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK, 12MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK, 13Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, China and 14School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

*Corresponding author. NTNU Department of Clinical and Molecular Medicine, NO-7491 Trondheim, Norway. E-mail: yi-qian.sun@ntnu.no

Received 17 October 2020; Accepted 18 February 2021; editorial decision 9 February 2021

Abstract

Background: It is unclear if smoking-related DNA methylation represents a causal pathway between smoking and risk of lung cancer. We sought to identify novel smoking-related DNA methylation sites in blood, with repeated measurements, and to appraise the putative role of DNA methylation in the pathway between smoking and lung cancer development.
Methods: We derived a nested case-control study from the Trøndelag Health Study (HUNT), including 140 incident patients who developed lung cancer during 2009–13 and 140 controls. We profiled 850 K DNA methylation sites (Illumina Infinium EPIC array) in DNA extracted from blood that was collected in HUNT2 (1995–97) and HUNT3 (2006–08) for the same individuals. Epigenome-wide association studies (EWAS) were performed for a detailed smoking phenotype and for lung cancer. Two-step Mendelian randomization (MR) analyses were performed to assess the potential causal effect of smoking on DNA methylation as well as of DNA methylation (13 sites as putative mediators) on risk of lung cancer.

Results: The EWAS for smoking in HUNT2 identified associations at 76 DNA methylation sites ($P < 5 \times 10^{-8}$), including 16 novel sites. Smoking was associated with DNA hypomethylation in a dose-response relationship among 83% of the 76 sites, which was confirmed by analyses using repeated measurements from blood that was collected at 11 years apart for the same individuals. Two-step MR analyses showed evidence for a causal effect of smoking on DNA methylation but no evidence for a causal link between DNA methylation and the risk of lung cancer.

Conclusions: DNA methylation modifications in blood did not seem to represent a causal pathway linking smoking and the lung cancer risk.

Key words: Causal inference, EWAS, Mendelian randomization

Introduction

Lung cancer has been the most common cancer type for several decades worldwide, and it kills the largest number of people with a 5-year survival rate of 10% globally. Clinical diagnostics are challenging when nodules $\leq 8 \text{ mm}$ are found in the lungs of patients, as such nodules may not be due to a malignant disease. Moreover, it is difficult and not without risk to obtain tissue samples from such nodules, and usually these patients are followed up with computed tomography surveillance over time. As a supplement to current standard procedures, it is important to identify biomarkers that are associated with the risk even before cancerous changes arise. In line with this, recent research has shed light on the involvement of epigenetic modifications in cancer development. Among the epigenetic modifications, DNA methylation involving the addition of a methyl group to the carbon-5 of a cytosine residue, which occurs predominantly at CpG sites (regions of DNA where a cytosine nucleotide is followed by a guanine nucleotide along DNA’s 5’ to 3’ direction) is of particular interest as a molecular mechanism underlying cancer risk.

DNA methylation in blood is highly sensitive to lifestyle influences such as smoking and emerging evidence suggests that it may also reflect changes in the target tissue such as in the lung. Recently, Fasanelli et al. reported that hypomethylation of smoking-related genes in blood was associated with future onset of lung cancer. Since tobacco smoking is a causal risk factor of lung cancer, it is possible that DNA methylation changes lie on the causal pathway between smoke exposure and lung cancer risk.
There have been some previous attempts to determine if DNA methylation mediates the influence of lifestyle factors on diseases.\(^5,14\) Fasanelli et al. suggested that hypomethylation in smoking-related genes AHRR and F2RL3 mediated the effect of tobacco on lung cancer risk with large magnitude.\(^5\) This study, however, used observational methods that often have limitations such as confounding and reverse causation and thus make causal inference difficult. A Mendelian randomization (MR) approach can be applied in this context, as it has developed to evaluate causal relationships by using genetic variants as instrumental variables for the exposure of interest.\(^15,16\) Genetic variants at a given locus may influence methylation pattern across an extended genomic region.\(^17\) These variants are defined as methylation quantitative trait loci (mQTLs), and can be used as a proxy for methylation levels in an MR analysis.\(^18–20\)

In this study, we performed epigenome-wide association studies (EWAS) for smoking and lung cancer with repeatedly measured DNA methylation obtained from pre-diagnostic blood samples. The DNA methylation was assayed using the Infinium MethylationEPIC BeadChip (Illumina Inc., CA, USA), which can detect >850 K methylation sites. This supersedes the Illumina Infinium HumanMethylation450 array which has been used in previous EWAS for smoking and lung cancer.\(^5,14,21,22\) We also performed two-step MR analyses\(^20\) to appraise the putative causal role of DNA methylation in the pathway between smoking and lung cancer development.

**Methods**

All participants gave their informed consent for participation in HUNT. The current study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REK 2015/78). Ethical approval for Generation Scotland was obtained from the Tayside Committee on Medical Research Ethics (on behalf of the National Health Service).

**Study design and population**

The Trøndelag Health Study (the HUNT Study) is one of the largest population-based health surveys conducted in Norway.\(^23\) The HUNT Study invited all inhabitants aged 20 years or older in the northern area of Trøndelag in four waves: HUNT1 (1984–86), HUNT2 (1995–97), HUNT3 (2006–08) and HUNT4 (2017–19). A nested case-control study was designed within HUNT2 and HUNT3, including 140 incident cases who developed lung cancer during 2009–13 and 140 age-(±3 years) and sex-matched controls. The study design and selection criteria for cases and controls are described in Figure 1. Incident lung cancer cases were ascertained based on the linkage of data between HUNT and the Cancer Registry of Norway. Pre-diagnostic blood samples were collected in HUNT2 and HUNT3 from both the cases and the controls and stored at

![Figure 1](https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyab044/6174543)

**Selection criteria for incident lung cancer cases**

- Participated in both HUNT2 and HUNT3
- With available blood and DNA samples in both HUNT2 and HUNT3
- Never diagnosed with cancer in HUNT2 or HUNT3
- Diagnosis of lung cancer between 2009 and 2013

**Selection criteria for controls**

- Participated in both HUNT2 and HUNT3
- With available blood and DNA samples in both HUNT2 and HUNT3
- Never diagnosed with cancer or long-term chronic diseases in HUNT2 or HUNT3
- Alive during the follow-up period by end of 2013
–80°C for later use. Among the incident cases, the mean years from blood collection to lung cancer diagnosis were 15.0 (range: 11.8–18.0) in HUNT2 and 3.8 (range: 1.0–6.7) in HUNT3.

Genotype and lifestyle variables

Information on genotypes and lifestyle factors was extracted from the HUNT databank.24 Information on smoking was collected in both HUNT2 and HUNT3. A smoking phenotype (seven levels) was generated taking into account the smoking status and pack-years (pyrs): 0: never smokers; 1: former ≤10.0 pyrs; 2: former 10.1–20.0 pyrs; 3: former ≥20.1 pyrs; 4: current ≤10.0 pyrs; 5: current 10.1–20.0 pyrs; and 6: current ≥20.1 pyrs. A variable for change in smoking status between HUNT2 and HUNT3 [0: decrease (current to former smokers); 1: no change (never to never and former to former); and 2: increase (never to former, never to current, former to current and current to current)] was generated based on status of never, former and current smokers in the HUNT2 and HUNT3. Current to current was classified as an increase in smoking status as exposure to tobacco smoke had been accumulated.

Genome-wide DNA methylation analysis, quality control and normalization

Genome-wide DNA methylation was analysed in a total of 560 pre-diagnostic blood samples that were collected from 280 study subjects on two occasions when they participated in HUNT2 and HUNT3. About 500 ng DNA isolated from peripheral blood cells was subject to bisulphite conversion, in HUNT2 and HUNT3 samples (about 11 years apart) in relation to the smoking phenotype in HUNT2 among the controls. A less computationally intensive strategy with cluster-robust standard errors (LMRSE) was performed.27 We also explored the possible effect of change in smoking status between HUNT2 and HUNT3 (categorized as decrease, no change or increase) on change in DNA methylation (beta-value of DNA methylation in HUNT2) among the controls.

Third, EWAS for lung cancer was performed among the lung cancer cases vs controls with DNA methylation as the exposure measured in HUNT2 and HUNT3, respectively, and the P-value cut-off was set at 5 × 10⁻⁸.

Fourth, the smoking-related DNA methylation sites that overlapped between the EWAS for smoking and the EWAS for lung cancer in the HUNT2 samples were individually evaluated as potential mediators between the smoking phenotype and lung cancer, using mediation analysis. Multiple mediators were then considered simultaneously, and a weighted methylation score was calculated.

Fifth, two-step MR analyses were performed. A first step was applied to evaluate the causal effect of smoking on DNA methylation. We used a smoking genetic score including three single nucleotide polymorphisms (SNPs) as an instrumental variable for the smoking phenotype: rs6265 (BDNF) associated with smoking initiation, rs1051730 (CHRNA3) with smoking quantity and rs3025343 (DBH) with smoking cessation.28 One-sample MR using the two-stage least square (2SLS) method was applied to investigate a causal relationship between smoking and DNA methylation at the sites identified in the EWAS for smoking. A second-step MR was performed to evaluate the putative description of the statistical analyses is given in the online Supplementary Material. Different sets of data that were used for specific statistical analyses are described in Supplementary Table S1, available as Supplementary data at IJE online.

First, we carried out an EWAS for the smoking phenotype (the seven levels) in blood samples collected from the controls in HUNT2. Linear regressions were performed with DNA methylation beta-values as the outcome and smoking phenotype as the exposure. Covariates were included in the linear regression models to adjust for the effects of sex, age and estimated cell counts. Surrogate variable analysis (SVA)25 was used to generate 12 variables that were also included as covariates in the EWAS models to adjust for batch and other technical artefacts. The P-value cut-off was set at epigenome-wide level (5 × 10⁻⁸). EWAS for smoking, performed with R package meffil (version 1.1.0).26 Second, to confirm the associations identified from the EWAS for smoking, we performed an analysis using repeatedly measured DNA methylation data from both the HUNT2 and the HUNT3 samples (about 11 years apart) including three single nucleotide polymorphisms (SNPs) as an instrumental variable for the smoking phenotype: rs6265 (BDNF) associated with smoking initiation, rs1051730 (CHRNA3) with smoking quantity and rs3025343 (DBH) with smoking cessation.28 One-sample MR using the two-stage least square (2SLS) method was applied to evaluate the causal effect of smoking on DNA methylation. We used a smoking genetic score including three single nucleotide polymorphisms (SNPs) as an instrumental variable for the smoking phenotype: rs6265 (BDNF) associated with smoking initiation, rs1051730 (CHRNA3) with smoking quantity and rs3025343 (DBH) with smoking cessation.28 One-sample MR using the two-stage least square (2SLS) method was applied to investigate a causal relationship between smoking and DNA methylation at the sites identified in the EWAS for smoking. A second-step MR was performed to evaluate the putative
causal association between DNA methylation and the risk of lung cancer. We applied a two-sample MR in order to leverage power from large genome-wide association studies (GWAS). Instruments for the DNA methylation sites detected as putative mediators with the mediation analyses were extracted from an mQTL (both cis and trans) GWAS in a subset of Generation Scotland (n = 5101). Summary statistics of lung cancer GWAS were derived from McKay et al. with sample size 85 716 (cases 29 266 vs controls 56 450). The inverse-variance weighted (IVW) method or Wald ratio method (when only one mQTL as instrumental variable) was used to calculate the causal estimates.

Results

Characteristics of study participants

Characteristics of the lung cancer cases and controls whose DNA methylation was measured in HUNT2 and HUNT3 and passed QC are presented in Supplementary Table S2, available as Supplementary data at IJE online. There were more men than women (55% vs 45%). The mean age was similar in cases and controls (56.4 vs 55.6 years in HUNT2 and 67.8 vs 66.8 years in HUNT3). About 90% of the lung cancer cases were former or current smokers whereas about half of the controls were never smokers in HUNT2 and HUNT3.

Identification of DNA methylation sites associated with smoking

The EWAS for smoking in blood samples collected in HUNT2 was performed in 128 of the 137 controls, due to missing data on the smoking phenotype. We identified 76 (P < 5 × 10^{-8}) DNA methylation sites (Table 1 and Figure 2; Supplementary Figure S2, available as Supplementary data at IJE online). The range of the effect sizes (difference in DNA methylation beta-value per one level increase in smoking phenotype) was from -0.052 to 0.030. Smoking was inversely associated with DNA methylation for 63 (83%) of the 76 sites, among which cg05575921 had the strongest association (P = 3.0 × 10^{-38}). Top DNA methylation sites around or within genes (5’-UTR or gene body) such as AHRR, F2RL3, RARA, MGAT3, GPR15 and PRSS23, were identified as being associated with smoking. Box plots showed a dose-response association between the smoking phenotype and DNA hypomethylation for most of the 12 top sites (P-values < 5.5 × 10^{-15}, Figure 3).

Among the identified 76 sites, 35 sites were exclusive to the MethylationEPIC BeadChip compared with the HumanMethylation450 BeadChip (Table 1). Nineteen of the 35 EPIC BeadChip specific sites confirmed previous smoking loci, such as F2RL3, AHRR, MGAT3, GPR15, PRSS23, ELMSAN1 and RARA etc. Sixteen DNA methylation sites are novel signals (Table 1), and three of them were annotated to the following genes: NBR1 (cg13849276, P = 8.7 × 10^{-11}), SLAMP7 (cg00045592, P = 2.8 × 10^{-10}) and HERC2 (cg13258799, P = 2.5 × 10^{-8}). The remaining 13 signals were not annotated.

Confirmation of EWAS for smoking

Among the 76 sites, 75 sites (i.e. except cg23079012) were confirmed after Bonferroni correction (actual P-value × 76 < 0.05) using repeated measurements with LMRSE (n = 124). The results of LMRSE highly correlated with those of a computationally intensive linear mixed effects model (LMEM) with random intercept for randomly selected 1000 DNA methylation sites (Supplementary Figure S3, available as Supplementary data at IJE online: correlation R = 0.97, P < 2.2 × 10^{-16}). Estimates from the EWAS for smoking and the LMRSE analysis showed a strong correlation for the 76 sites (R = 0.99, P < 2.2 × 10^{-16}, Supplementary Figure S4, available as Supplementary data at IJE online).

Change in smoking status between HUNT2 and HUNT3 was available for the 128 controls who were categorized as 16 with decrease (current to former), 88 with no change (59 never to never, 29 former to former), and 24 with increase (3 never to former, 1 never to current, 6 former to current, 14 current to current) in smoking status. Of the 76 DNA methylation sites, five sites were associated with smoking change (Bonferroni corrected P < 0.05) and showed a dose-response relationship (Table 2 and Figure 4). Among the five sites, cg18110140 is a novel site.

Identification of DNA methylation sites associated with lung cancer

When the smoking phenotype was included in the EWAS model for lung cancer, no DNA methylation sites survived adjustment for multiple tests (P < 5 × 10^{-8}) in either HUNT2 (139 cases vs 137 controls) or HUNT3 (131 cases vs 135 controls). When smoking was not included in the EWAS model to study DNA methylation sites as potential mediators linking smoking and lung cancer, associations at 50 and 18 DNA methylation sites survived adjustment for multiple tests in HUNT2 (Table 3; Supplementary Figure S5, available as Supplementary data at IJE online) and HUNT3, respectively. Of these, 30 sites from HUNT2 and all the 18 sites from HUNT3 overlapped with the 76 smoking-related sites and 17 sites overlapped between HUNT2 and HUNT3.
Table 1  Epigenome-wide associations ($P < 5 \times 10^{-8}$) for smoking in blood samples collected in HUNT2 in controls ($n = 128$)

| DNAm sites   | Coefficient | P-value  | Chr | Position | Gene     | Gene region | Exclusively in EPIC Beadchip | Novel DNAm sites and loci |
|--------------|-------------|----------|-----|----------|----------|-------------|-------------------------------|---------------------------|
| cg05375921   | -0.052      | 2.97E-36 | 5   | 373378   | AHR      | Body        |                               |                           |
| cg03636183   | -0.027      | 2.07E-27 | 19  | 17000585 | F2R      | Body        |                               |                           |
| cg21566642   | -0.031      | 8.57E-26 | 2   | 233284661|          |             |                               |                           |
| cg17739917   | -0.020      | 4.04E-23 | 17  | 38477572 | RARA     | 5' UTR      |                               |                           |
| cg01940273   | -0.021      | 1.65E-21 | 2   | 233284934|          |             |                               |                           |
| cg21911711   | -0.013      | 9.32E-19 | 19  | 169986688| F2RL3    | TSS1500     |                               |                           |
| cg05086879   | -0.013      | 2.06E-18 | 22  | 398614900| MGAT3    | 5' UTR      |                               |                           |
| cg26703534   | -0.013      | 4.74E-18 | 5   | 377358   | AHR      | Body        |                               |                           |
| cg21161138   | -0.015      | 6.53E-18 | 5   | 399360   | AHR      | Body        |                               |                           |
| cg17087741   | -0.009      | 8.98E-18 | 2   | 233283010|          |             |                               |                           |
| cg04180924   | -0.007      | 5.35E-17 | 3   | 98272064 |          |             |                               |                           |
| cg14391737   | -0.009      | 1.47E-14 | 15  | 75350380 | AHR      | Body        |                               |                           |
| cg14466441   | -0.009      | 2.92E-14 | 6   | 11392193 |          |             |                               |                           |
| cg19859270   | -0.009      | 5.35E-13 | 17  | 38476024 | RARA     | 5' UTR      |                               |                           |
| cg14391737   | -0.012      | 9.17E-13 | 11  | 86513428 | AHR      | Body        |                               |                           |
| cg18110140   | -0.014      | 1.47E-13 | 15  | 75350380 | AHR      | Body        |                               |                           |
| cg14466441   | -0.009      | 2.92E-13 | 6   | 11392193 |          |             |                               |                           |
| cg16841366   | -0.017      | 3.73E-12 | 2   | 233286192| AHR      | Body        |                               |                           |
| cg22812571   | -0.017      | 3.86E-12 | 2   | 233286229|          |             |                               |                           |
| cg19572487   | -0.004      | 5.35E-15 | 3   | 98251294 | GPR15    | 1st Exon    |                               |                           |
| cg14391737   | -0.012      | 9.17E-13 | 11  | 86513428 | AHR      | Body        |                               |                           |
| cg18110140   | -0.014      | 1.47E-13 | 15  | 75350380 | AHR      | Body        |                               |                           |
| cg14466441   | -0.009      | 2.92E-13 | 6   | 11392193 |          |             |                               |                           |
| cg16841366   | -0.017      | 3.73E-12 | 2   | 233286192| AHR      | Body        |                               |                           |
| cg22812571   | -0.017      | 3.86E-12 | 2   | 233286229|          |             |                               |                           |
| cg19572487   | -0.004      | 5.35E-15 | 3   | 98251294 | GPR15    | 1st Exon    |                               |                           |
| cg14391737   | -0.012      | 9.17E-13 | 11  | 86513428 | AHR      | Body        |                               |                           |
| cg18110140   | -0.014      | 1.47E-13 | 15  | 75350380 | AHR      | Body        |                               |                           |
| cg14466441   | -0.009      | 2.92E-13 | 6   | 11392193 |          |             |                               |                           |
| cg16841366   | -0.017      | 3.73E-12 | 2   | 233286192| AHR      | Body        |                               |                           |
| cg22812571   | -0.017      | 3.86E-12 | 2   | 233286229|          |             |                               |                           |
| cg19572487   | -0.004      | 5.35E-15 | 3   | 98251294 | GPR15    | 1st Exon    |                               |                           |
| cg14391737   | -0.012      | 9.17E-13 | 11  | 86513428 | AHR      | Body        |                               |                           |
| cg18110140   | -0.014      | 1.47E-13 | 15  | 75350380 | AHR      | Body        |                               |                           |
|cg(Continued)
Mediation effects of DNA methylation on the pathway between smoking and risk of lung cancer

The 30 smoking- and lung cancer-overlapped DNA methylation sites from HUNT2 were tested as potential mediators between smoking and lung cancer, among which 14 sites were identified. The relative mediation effects of the 14 DNA methylation sites and the weighted mediation score based on the sum of the 14 sites are presented in Table 4. The indirect effect carried by the weighted mediation score accounted for 61% of total effect from smoking phenotype to lung cancer development.

Evaluation of potential causal association between smoking and DNA methylation

Summary statistics from the first-step MR between smoking and DNA methylation for the 76 DNA methylation sites are presented in Supplementary Table S3, available as Supplementary data at IJE online. Eleven sites showed statistical evidence for a causal association (P < 0.05). The genetic score explained 1.8% of the variance in smoking with an F statistic of 2.4. To further evaluate the extent to which the EWAS associations reflect causal effects, we plotted the MR estimates against the EWAS estimates for the 76 sites (Figure 5) and it showed a good correlation (R = 0.74, P = 2.1 × 10^{-14}).

Evaluation of putative causal association between DNA methylation and lung cancer risk

The second-step MR evaluated the effect of DNA methylation on risk of lung cancer (Table 5). The 14 putative DNA methylation mediators identified by mediation analysis are linked to genes GPR15, AHRR, MIR4505/ELMSAN1, RARA, F2RL3, PRSS23 and SLAMF7. We were not able to perform MR for cg24859433 as summary statistics for its mQTLs with lung cancer were not available.
Both cis and trans mQTLs (range 1 to 9 per DNA methylation site) were used as instrumental variables for DNA methylation. The mQTLs explained 0.6% to 6.8% of the variance in DNA methylation for the included 13 sites. None of the 13 DNA methylation sites demonstrated a causal effect on the risk of lung cancer (Bonferroni correction: actual $P$-value × 13 > 0.05 for all, Table 5). In addition, there was no clear correlation ($R = 0.083$, $P = 0.79$) between the estimates derived from the MR and EWAS for lung cancer for the 13 sites (Supplementary Figure S6, available as Supplementary data at IJE online). To reduce the possibility of pleiotropy of the instrumental variables, the second-step MR was also performed using cis-only mQTLs and it showed no causal evidence (Supplementary Table S4, available as Supplementary data at IJE online).

**Discussion**

**Main findings**

In this study, we identified 76 DNA methylation sites associated with smoking, using the Illumina Infinium MethylationEPIC BeadChip, among which 16 sites were...
Figure 3 Associations between smoking (horizontal axis) and DNA methylation (vertical axis) for the top 12 smoking-related DNA methylation sites. X-axis represents seven levels of smoking phenotype: 0: never smokers; 1: former ≤10.0 pack-years (pyrs); 2: former 10.1–20.0 pyrs; 3: former >20.1 pyrs; 4: current ≤10.0 pyrs; 5: current 10.1–20.0 pyrs; 6: current >20.1 pyrs). sva: surrogate variable analysis.
novel and not captured on the older HumanMethylation450 array. Our results showed that smoking appeared to be a causal factor for DNA methylation modifications in the blood. There was no evidence for a causal effect of smoking-related DNA methylation on the risk of lung cancer.

Comparison with previous studies
To our knowledge, this is the first study to use the MethyltionEPIC BeadChip to identify smoking-related DNA methylation sites in the blood. Of the 76 sites, we replicated 41 sites that were previously identified with 450K. New probes on 850K further confirmed some previously identified smoking-related genes. Although it is difficult to compare effect sizes in our study with those in previous work, due to different definitions of smoking phenotype, the genes associated with our top DNA methylation sites are consistent with those frequently found in previous 450K studies, such as AHRR, F2RL3 and PRSS23.

Of the 76 DNA methylation sites, 35 were exclusive to the MethyltionEPIC BeadChip. Of the 35 sites, 19

Table 2: Associations (Bonferroni corrected \(P < 0.05\)) between change in smoking status and change in DNA methylation among smoking-related DNA methylation sites\(^a\)

| DNAm sites | Coefficient\(^b\) | 95% CI | Chromosome | Position | Bonferroni corrected \(P\)-value |
|------------|------------------|--------|------------|----------|-----------------------------|
| cg05575921 | -0.037           | -0.053 | 5          | 373378   | 1.20E-03                    |
| cg26703534 | -0.020           | -0.029 | 5          | 377358   | 1.47E-03                    |
| cg17087741 | -0.012           | -0.017 | 2          | 233283010| 2.80E-03                    |
| cg18110140 | -0.015           | -0.023 | 15         | 75350380 | 2.06E-02                    |
| cg10765427 | -0.012           | -0.017 | 19         | 17005225 | 8.50E-04                    |

DNAm, DNA methylation.
\(^a\)Smoking-related DNA methylation sites: the 76 sites were identified in EWAS for smoking in the cross-sectional analysis.
\(^b\)Coefficient: difference in the change of DNA methylation beta-value per level increase in smoking change.

Figure 4: Associations between change in smoking status (horizontal axis) and change in DNA methylation (vertical axis) between HUNT2 and HUNT3 for the five smoking-related DNA methylation sites (Bonferroni corrected \(P < 0.05\)). Horizontal axis stands for change in smoking status [0: decrease (from current to former smokers); 1: no change (never to never; former to former); and 2: increase (never to former; never to current; former to current; current to current)]. HUNT: the Trøndelag Health Study; sva: surrogate variable analysis.
### Table 3

Epigenome-wide associations ($P < 5 \times 10^{-8}$) for lung cancer in 139 cases vs 137 controls in the HUNT2 study ($n = 276$)

| DNAm sites | Coefficient | SE      | P-value     | OR | Chromosome | Position | Gene     | Smoking-related DNAm sites |
|------------|-------------|---------|-------------|----|------------|----------|----------|---------------------------|
| cg05575921 | -11.854     | 1.603   | 1.43E-13    | 0.89 | 5          | 373378   | AHRR     | Yes                       |
| cg21911711 | -33.453     | 4.815   | 3.72E-12    | 0.7 | 19         | 16998666 | F2RL3    | Yes                       |
| cg03636183 | -18.879     | 2.728   | 4.46E-12    | 0.83 | 19         | 17000585 | F2RL3    | Yes                       |
| cg21566642 | -17.334     | 3.512   | 5.21E-12    | 0.84 | 2          | 232384661 |         |                           |
| cg01940273 | -23.919     | 3.474   | 5.67E-12    | 0.79 | 2          | 232384934 |         |                           |
| cg17739917 | -24.414     | 3.601   | 1.20E-11    | 0.78 | 17         | 38477572 | RARA     | Yes                       |
| cg21161138 | -25.790     | 3.892   | 3.44E-11    | 0.77 | 5          | 399360   | AHRR     | Yes                       |
| cg24859433 | -39.290     | 5.970   | 6.66E-11    | 0.68 | 6          | 30720203 |         |                           |
| cg19572487 | -28.832     | 4.494   | 1.40E-10    | 0.75 | 17         | 38476024 | RARA     | Yes                       |
| cg03086879 | -27.038     | 4.216   | 1.43E-10    | 0.76 | 22         | 39861490 | MGAT3    | Yes                       |
| cg14391737 | -18.074     | 2.837   | 1.30E-09    | 0.84 | 2          | 233286192 |        |                           |

DNAm, DNA methylation; OR, odds ratio; SE, standard error.

*a*Coefficient when DNA methylation beta-value changes from 0 to 1; smoking was not adjusted for in the model.

*b*Odds ratio of lung cancer per 1% increase of DNA methylation at the site.
confirmed previous smoking loci and 16 were novel signals. Three of the novel sites were annotated to the following genes: **NBR1**, **SLAMF7** and **HERC2**. The protein encoded by **NBR1** functions as a specific autophagy receptor and is associated with bilateral breast and ovarian cancers. **SLAMF7** encodes a self-ligand receptor of the signalling lymphocytic activation molecule (SLAM) family. Activated SLAM receptors are involved in the regulation of both innate and adaptive immune response. **HERC2** encodes a group of large proteins that are involved in neurodevelopment, DNA damage repair and immune response. In line with our findings from the blood samples, DNA hypomethylation was also identified at cg05575921 (**MGAT3**) and cg12956751 (**ALPP**) in saliva of current smokers in a previous study using the MethylationEPIC BeadChip, and hypomethylation at cg24797066 was observed to be related to smoking in bronchoalveolar lavage cells.

Our study suggested that smoking had a causal effect on DNA methylation in the blood, which is consistent with the findings from a recent study. Although our genetic instrument for smoking was weak, the correlation of estimates derived from the MR and EWAS analyses was moderately high. Our results did not support a causal effect of smoking-related DNA methylation in **AHRR**, **F2RL3** and **PRSS2** on the risk of lung cancer, which confirmed and extended the results from a recent MR study.

The 13 DNA methylation sites that were tested for causal relationship with lung cancer risk in our study included seven sites (three novel) from the EPIC BeadChip and six from the 450 K array, whereas the aforementioned MR study included 16 DNA methylation sites from the 450 K array among which only cg05575921 overlapped with ours. DNA methylation at cg05575921 in **AHRR** has been

| DNAm sites              | Total effect |        | Indirect effect |        | Relative indirect effect |        |
|-------------------------|--------------|--------|----------------|--------|--------------------------|--------|
|                         | Coefficient  | 95% CI | Coefficient    | 95% CI |                         | 95% CI |
| cg19859270              | 0.74         | 0.48   | 0.98           | 0.19   | 0.05                     | 0.33   |
| cg05575921              | 0.71         | 0.47   | 0.95           | 0.36   | 0.00                     | 0.74   |
| cg25845814†             | 0.72         | 0.48   | 0.93           | 0.15   | -0.01                    | 0.35   |
| cg24859433              | 0.72         | 0.47   | 0.90           | 0.18   | 0.06                     | 0.37   |
| cg15342087              | 0.71         | 0.48   | 0.90           | 0.15   | 0.00                     | 0.28   |
| cg26768182‡             | 0.72         | 0.49   | 0.93           | 0.16   | 0.00                     | 0.37   |
| cg19572487              | 0.74         | 0.51   | 1.00           | 0.22   | 0.11                     | 0.35   |
| cg24797066‡             | 0.71         | 0.49   | 0.94           | 0.13   | -0.03                    | 0.26   |
| cg21911711†             | 0.72         | 0.48   | 1.00           | 0.21   | 0.05                     | 0.42   |
| cg00475490†             | 0.73         | 0.48   | 1.02           | 0.19   | 0.06                     | 0.37   |
| cg00045592‡             | 0.70         | 0.47   | 0.96           | 0.14   | 0.03                     | 0.31   |
| cg03229539              | 0.70         | 0.48   | 0.93           | 0.14   | 0.00                     | 0.31   |
| cg14391737‡             | 0.70         | 0.49   | 0.93           | 0.13   | -0.04                    | 0.29   |
| cg21161138              | 0.69         | 0.47   | 0.92           | 0.16   | 0.04                     | 0.38   |
| Weighted mediation score| 0.74         | 0.46   | 1.00           | 0.45   | 0.12                     | 0.76   |

DNA methylation.

†14 DNA methylation sites were identified as mediators individually with the counterfactual framework.

‡Indirect effect divided by total effect.

Exclusively in EPIC Beadchip.

The sum of methylation beta-value at each of 14 DNA methylation sites weighted by effect size with lung cancer.

**Figure 5** Correlation between Mendelian randomization and epigenome-wide association study estimates for smoking-DNA methylation associations for the smoking-related 76 DNA methylation sites. The 11 sites with $P<0.05$ in MR analysis are labelled.
## Table 5 Second-step Mendelian randomization of DNA methylation and risk of lung cancer

| DNAm sites as putative mediators | Chr | Gene | Instrumental variables (mQTLs) | mQTLs chromosome | Variance explained (%) | OR\(^a\) | 95% CI | P-value |
|----------------------------------|-----|------|--------------------------------|------------------|------------------------|---------|--------|---------|
| cg19859270\(^c\)                 | 3   | GPR15| rs4540316; rs1529047; rs1864203; rs6855577; rs6826969; rs74995805; rs150605105 | 7; 3; 9; 4; 4; 2 | 2.9 | 0.93 | 0.84 | 1.02 | 0.13 |
| cg05575921                       | 5   | AHRR | rs11956656                      | 5                | 0.9 | 0.96 | 0.80 | 1.16 | 0.70 |
| cg25845814\(^b\)                 | 14  | MIR4505; ELMSAN1                | rs1323124; rs111686083; rs139016638; rs951574; rs80186749; rs74959723; rs112116518; rs3756764; rs141965025 | 1; 14; 2; 16; 2; 2; 1; 5; 20 | 3.7 | 1.04 | 0.93 | 1.16 | 0.53 |
| cg15342087\(^d\)                 | 6   |      | rs11190127                      | 10               | 0.7 | 1.08 | 0.88 | 1.32 | 0.45 |
| cg26768182\(^b\)                 | 9   |      | rs78581928                      | 9                | 0.9 | 1.23 | 1.05 | 1.46 | 0.01 |
| cg19572487\(^d\)                 | 17  | RARA | rs17032705                      | 4                | 0.8 | 0.97 | 0.80 | 1.18 | 0.79 |
| cg24797066\(^b\)                 | 20  |      | rs602598                        | 20               | 0.6 | 1.00 | 0.80 | 1.25 | 0.99 |
| cg21911711\(^b\)                 | 19  | F2RL3 | rs56298289; rs79977579; rs3848656 | 19; 12; 19     | 6.8 | 0.98 | 0.91 | 1.06 | 0.64 |
| cg00475490\(^b\)                 | 11  | PRSS23 | rs2279046                      | 11               | 3.3 | 1.05 | 0.96 | 1.16 | 0.25 |
| cg00045392\(^b\)                 | 1   | SLAMF7 | rs3766573; rs352684             | 1; 1             | 2.4 | 0.96 | 0.86 | 1.08 | 0.52 |
| cg03529539\(^b\)                 | 2   |      | rs13023370                      | 2                | 4.0 | 1.01 | 0.93 | 1.10 | 0.85 |
| cg14391737\(^b\)                 | 11  | PRSS23 | rs7607726; rs7606236; rs62010937; rs9424468; rs147426883 | 2; 2; 15; 1; 19 | 2.1 | 0.99 | 0.88 | 1.12 | 0.91 |
| cg21161113\(^b\)                 | 5   | AHRR | rs62289477; rs117666260; rs79694935; rs2466287; rs79991330 | 3; 8; 4; 5; 10 | 2.2 | 1.05 | 0.95 | 1.17 | 0.31 |

**Bold type:** sentinel mQTL (\(\text{ cis}\) mQTL with smallest \(P\)-value in Generation Scotland).

**Chr, chromosome; DNAm, DNA methylation; mQTL, methylation quantitative trait locus.**

\(^a\)Per 1-unit increase of DNA methylation M-value.

\(^b\)Exclusively in EPIC Beadchip.

\(^c\)\(P < 1 \times 10^{-5}\) for association with the mQTLs as no mQTLs found if smaller \(P\)-value was set.
found to be most strongly influenced by smoking in the current and previous studies.\textsuperscript{5,10,14,23} However, there was no clear evidence for a causal link between DNA methylation at cg05575921 and the risk of lung cancer in our second-step MR analysis nor in the referred MR study.\textsuperscript{14} This is in contrast to previous findings by Fasanelli et al., who reported that hypomethylation of DNA methylation sites in AHRR and F2RL3 may mediate the effect of tobacco smoking on lung cancer risk, based on observational mediation analyses.\textsuperscript{5} Our results indicate this might have been due to residual confounding in the previous mediation analysis. We also identified several other potential mediating DNA methylation sites near or in genes such as RARA, GPR15, SLAMF7 and MIR4505/ELMSAN1. Among these genes, SLAMF7 is a novel signal identified by the EPIC array in our study. Our second-step MR analysis, however, did not show evidence for a causal effect of cg19572487 in RARA, cg19859270 in GPR15 or cg25845814 in MIR4505/ELMSAN1 on the risk of lung cancer. Nor did we find that cg00045592 in SLAMF7 was causally associated with lung cancer risk.

Strengths and limitations

There are several strengths to our study. We used the latest Illumina HumanMethylation EPIC BeadChip to analyse DNA methylation, which covers over 850 K DNA methylation sites and thus provides a higher coverage compared with the previous arrays. Blood samples used to generate DNA methylation profiles were collected years before the diagnosis of lung cancer. In HUNT2 this was on average 15 years before diagnosis and therefore reverse causation was unlikely. The information on smoking status and pack-years was recorded years before the diagnosis, which reduced the recall bias. A detailed smoking phenotype was derived based on both smoking quantity in total and smoking status. By using the detailed smoking phenotype, a clear dose-response association of smoking with DNA methylation was demonstrated. To date, there have been few studies investigating the association between smoking and DNA methylation over time using repeated measurements.\textsuperscript{10} Our study showed that smoking-related DNA methylation was reliable: among the 76 DNA methylation sites identified from the EWAS, 75 sites were confirmed in the analysis using repeatedly measured DNA methylation data. In addition, we applied two-step MR analyses to evaluate if causal associations existed between smoking and DNA methylation as well as between DNA methylation and lung cancer risk. Our study confirmed and extended the findings of the previous studies assessing the above causal relationships respectively\textsuperscript{14,36} by including more and novel methylation sites identified with the EPIC BeadChip.

Our study also has limitations. We used the beta-values of DNA methylation for EWAS as they have intuitively biological interpretation. However, beta-values have severe heteroscedasticity outside the middle methylation range.\textsuperscript{37,38} The beta difference directly obtained from the beta-value linear regression model can give biased results when beta-values are not between 0.2 and 0.8.\textsuperscript{38} Our study may not have sufficient power to detect a small effect of DNA methylation on the risk of lung cancer. This power issue is reflected by the relatively wider 95% confidence intervals (CIs) in Table 5. Some of the null associations may be due to weak instrument bias, as the mQTLs explained only 0.6% to 6.8% of the variance in DNA methylation on the 13 CpG sites (the putative mediator sites). In two-sample MR, weak instrument bias inclines the association towards the null.\textsuperscript{19} Due to the small number of cases, we were not able to evaluate the causal effect of smoking-related DNA methylation in blood on the risk of specific histological types. Future studies are warranted to investigate the potential causal effect of DNA methylation in blood on risk of lung cancer histological types.

In conclusion, we identified 16 novel DNA methylation sites related to smoking, using the latest DNA methylation array. Smoking had a causal association with DNA methylation modifications. We did not find evidence for DNA methylation in blood being a causal factor for lung cancer risk. However, the newly identified smoking-related DNA methylation signals have the potential to be explored as additional markers for smoking, to improve the early prediction of lung cancer risk in future studies.

Data from the HUNT Study that are used in research projects will, when reasonably requested by others, be made available on request to the HUNT Data Access Committee [hunt@medisin.ntnu.no]. The HUNT data access information describes the policy regarding data availability [https://www.ntnu.edu/hunt/data].

**Supplementary Data**

Supplementary data are available at IJE online.

**Funding**

This work was supported by the Norwegian Cancer Society (project ID 182688-2016) and the Research Council of Norway ‘Gaveforsterkning’. Y.Q.S. was supported by a Researcher grant from the Liaison Committee for education, research and innovation in Central Norway (project ID 2018/42794). R.C.R. is a de Pass VC Research Fellow at the University of Bristol. T.B. is funded by a Wellcome Trust PhD studentship (203746). A.D.B. would like to acknowledge funding from the Wellcome PhD training fellowship for
clincians (204979Z/16/Z), the Edinburgh Clinical Academic Track (ECAT), R.C.R. and C.L.R. are supported by a Cancer Research UK programme grant (C18281/A19169) and R.C.R., J.M., M.S., C.L.R. and T.B. are all members of the MRC Integrative Epidemiology Unit at the University of Bristol supported by the UK Medical Research Council (MC_UU00011/5). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The corresponding authors had access to all the data in the study and had final responsibility for the decision to submit for publication. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Acknowledgements
The Trøndelag Health Study (HUNT) is a collaboration between the HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority and the Norwegian Institute of Public Health. The genotype quality control and imputation have been conducted by the K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology. The authors especially thank the HUNT databank and biobank for sample selection and DNA isolation from the blood. The DNA methylation and parts of bioinformatic analysis were provided in close collaboration with the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and Central Norway Regional Health Authority. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HRQ3006). Genotyping of the GS: SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, and was funded by the Medical Research Council UK and the Wellcome Trust (Welcome Trust Strategic Award ‘STratifying Resilience and Depression Longitudinally’ (STRADL) Reference 104036/Z/14/Z).

Author Contributions
Y.Q.S. and X.M.M. performed literature search. Y.Q.S., R.C.R., F.S., Y.C. and X.M.M. contributed to the study design. A.F. and V.B. performed DNA methylation analysis. R.M.W., A.D.B. and Y.Z. performed mQTL GWAS in Generation Scotland and provided the summary statistics. Y.Q.S, A.F., A.L. and X.M.M. were responsible for data collection. Y.Q.S., R.C.R. and X.M.M. conducted statistical analyses, interpreted results and wrote the initial draft of the manuscript. All authors participated in the data interpretation, contributed to statistical analyses and to the manuscript writing with important intellectual content and approved the final version of the manuscript.

Conflict of Interest
None declared.

References
1. GLOBOCAN 2012. Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. v1.0... Lyon, France: IARC, 2012.
2. Gould MK, Donington J, Lynch WR et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2013;143:e93S–e120S.
3. Relton CL, Davey Smith G. Is epidemiology ready for epigenetics? Int J Epidemiol 2012;41:5–9.
4. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell 2010;19:698–711.
5. Fasanelli F, Baglietto L, Ponzi E et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun 2015;6:10192.
6. Nebbioso A, Tambaro FP, Dell’Aversana C, Altucci L. Cancer epigenetics: Moving forward. PLoS Genet 2018;14:e1007362.
7. Herczeg Z, Ghantous A, Wild CP et al. Roadmap for investigating epigenome deregulation and environmental origins of cancer. Int J Cancer 2018;142:874–82.
8. Ambatipudi S, Caenin C, Hernandez-Vargas H et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics 2016;8:599–618.
9. Johanes R, Just AC, Marion RE et al. Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet 2016;9:436–47.
10. Wilson R, Wahl S, Pfeiffer L et al. The dynamics of smoking-related disturbed methylation: a two time-point study of methylation change in smokers, non-smokers and former smokers. BMC Genomics 2017;18:805.
11. Zeilinger S, Kuhnel B, Klopp N et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 2013;8:e63812.
12. Stueve TR, Li WQ, Shi J et al. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. Hum Mol Genet 2017;26:3014–27.
13. Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. BMC Cancer 2012;12:385.
14. Battram T, Richmond RC, Baglietto L et al. Appraising the causal relevance of DNA methylation for risk of lung cancer. Int J Epidemiol 2013;48:1493–504.
15. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014;23:R89–98.
16. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 2018;362:k601.
17. Bell JT, Pai AA, Pickrell JK et al. DNA methylation patterns associated with genetic and gene expression variation in HapMap cell lines. Genome Biol 2011;12:405. R10.
18. Hannon E, Gorrie-Stone TJ, Smart MC et al. Leveraging DNA methylation quantitative-trait loci to characterize the relationship between methylocnic variation, gene expression, and complex traits. Am J Hum Genet 2018;103:654–65.
19. Langdon R, Richmond R, Elliott HR et al. Identifying epigenetic biomarkers of established prognostic factors and survival in a
clinical cohort of individuals with oropharyngeal cancer. Clin Epigenet 2020;12:95.

20. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol 2012;41:161–76.

21. Baglietto L, Ponzi E, Haycock P et al. DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. Int J Cancer 2017;140:50–61.

22. Sandanger TM, Nøst TH, Guida F et al. DNA methylation and associated gene expression in blood prior to lung cancer diagnosis in the Norwegian Women and Cancer cohort. Sci Rep 2018;8:16714.

23. Krokstad S, Langhammer A, Hveem K et al. Cohort Profile: The HUNT study. Norway Int J Epidemiol 2013;42:968–77.

24. Sun YQ, Brumpton BM, Bonilla C et al. Serum 25-hydroxyvitamin D levels and risk of lung cancer and histologic types: a Mendelian randomisation analysis of the HUNT study. Eur Respir J 2018;51:18000329.

25. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet 2007;3:e161–35.

26. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient normalisation and analysis of very large DNA methylation datasets. Bioinformatics 2018;34:3983–89.

27. Staley JR, Suderman M, Simpkin AJ et al. Longitudinal analysis strategies for modelling epigenetic trajectories. Int J Epidemiol 2018;47:516–25.

28. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 2010;42:441–47.

29. Smith BH, Campbell A, Linksted P et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol 2013;42:689–700.

30. McKay JD, Hung RJ, Han Y et al.; SpiroMeta Consortium. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet 2017;49:1126–32.

31. Kirkin V, Lamark T, Sou YS et al. A role for NBR1 in autophagosomal degradation of ubiquitinylated substrates. Mol Cell 2009;33:505–16.

32. Chen J, Zhong MC, Guo H et al. SLAMF7 is critical for phagocytosis of haematopoietic tumour cells via Mac-1 integrin. Nature 2017;544:493–97.

33. Sánchez-Tena S, Cubillos-Rojas M, Schneider T, Rosa JL. Functional and pathological relevance of HERC family proteins: a decade later. Cell Mol Life Sci 2016;73:1955–68.

34. Barcelona V, Huang Y, Brown K et al. Novel DNA methylation sites associated with cigarette smoking among African Americans. Epigenetics 2019;14:383–91.

35. Ringh MV, Hagemann-Jensen M, Needhamsen M et al. Tobacco smoking induces changes in true DNA methylation, hydroxymethylation and gene expression in bronchoalveolar lavage cells. EBioMedicine 2019;46:290–304.

36. Li S, Wong EM, Bui M et al. Causal effect of smoking on DNA methylation in peripheral blood: a twin and family study. Clin Epigenet 2018;10:18.

37. Du P, Zhang X, Huang CC et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 2010;11:587.

38. Xie C, Leung YK, Chen A, Long DX, Hoyo C, Ho SM. Differential methylation values in differential methylation analysis. Bioinformatics 2019;35:1094–97.

39. Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. Int J Epidemiol 2016;45:908–15.