

## Investigating the DNA methylation profile of e-cigarette use

Rebecca C Richmond, PhD<sup>1,2\*</sup>, Carlos Sillero Rejon, PhD<sup>1,3</sup>, Jasmine N Khouja, PhD<sup>1,4</sup>, Claire Prince, BSc<sup>1,2</sup>, Alexander Board, MSc<sup>5</sup>, Gemma Sharp, PhD<sup>1,2</sup>, Matthew Suderman, PhD<sup>1,2</sup>, Caroline L Relton, PhD<sup>1,2</sup>, Marcus Munafò, PhD<sup>1,4</sup>, Suzanne H Gage, PhD<sup>1,6</sup>

<sup>1</sup> MRC Integrative Epidemiology Unit at the University of Bristol

<sup>2</sup> Population Health Sciences, Bristol Medical School, University of Bristol, UK

<sup>3</sup> National Institute for Health Research Applied Research Collaboration West (NIHR ARC West), University Hospitals Bristol NHS Foundation Trust, Bristol, UK

<sup>4</sup> School of Psychological Science, University of Bristol, UK

<sup>5</sup> Department of Experimental Psychology, Medical Sciences Division, University of Oxford, UK

<sup>6</sup> Department of Psychological Sciences, University of Liverpool, UK

**\*Corresponding author:** Dr Rebecca Richmond

Address: Oakfield House, Oakfield Grove, Bristol, UK, BS8 2BN

Email: [rebecca.richmond@bristol.ac.uk](mailto:rebecca.richmond@bristol.ac.uk)

Telephone: +44(0) 117 331 4043

**Running title:** DNA methylation profile of e-cigarette use

## Abstract

### **Rationale and objectives**

Little evidence exists on the health effects of e-cigarette use. DNA methylation may serve as a biomarker for exposure and could be predictive of future health risk. We aimed to investigate the DNA methylation profile of e-cigarette use.

### **Methods**

Among 117 smokers, 117 non-smokers and 116 non-smoking vapers, we evaluated associations between e-cigarette use and epigenome-wide methylation from saliva. We tested associations between e-cigarette use and methylation scores known to predict smoking and smoking-related disease. We assessed the ability of a methylation score for predicting e-cigarette use and for discriminating lung cancer.

### **Measurements and Main Results**

7 CpGs were identified in relation to e-cigarette use at  $p < 1 \times 10^{-5}$  and none at  $p < 5.91 \times 10^{-8}$ . 13 CpGs were associated with smoking at  $p < 1 \times 10^{-5}$  and one at  $p < 5.91 \times 10^{-8}$ . CpGs associated with e-cigarette use were largely distinct from those associated with smoking. There was strong enrichment of known smoking-related CpGs in the smokers but not the vapers. A methylation score for e-cigarette use showed poor prediction internally (AUC 0.55, 0.41-0.69) and externally (AUC 0.57, 0.36-0.74) compared with a smoking score (AUCs 0.80) and was less able to discriminate lung squamous cell carcinoma from adjacent normal tissue (AUC 0.64, 0.52-0.76 versus AUC 0.73, 0.61-0.85).

### **Conclusions**

The DNA methylation profile for e-cigarette use is largely distinct from that of cigarette smoking, did not replicate in independent samples, and was unable to discriminate lung cancer

from normal tissue. The extent to which methylation related to long-term e-cigarette use translates into chronic effects requires further investigation.

## Key Messages

### **What is the key question?**

Is there a DNA methylation signature of e-cigarette use and is it distinct from that of smoking?

### **What is the bottom line?**

Smoke exposure is known to lead to widespread changes in DNA methylation which have been identified in different populations and samples, persist for many years after smoking cessation, and may act as a biomarker for smoking-related disease risk and mortality. Whether a similar methylation profile exists in relation to e-cigarette use has not been widely investigated.

### **Why read on?**

We obtained saliva samples from 116 e-cigarette users and compared their DNA methylation profile with 117 smokers and 117 non-smokers. The e-cigarette users in this study had a minimal smoking history, and so we were able to distinguish the effects of e-cigarette use from those of smoke exposure. Overall, we found that the methylation profile associated with e-cigarette use is less pronounced and distinct from that associated with cigarette smoking.

**Key words:** e-cigarettes, DNA methylation, smoking, SEE-Cigs, ALSPAC

## Introduction

Electronic cigarettes (e-cigarettes) have the potential to reduce smoking-related harm. Although little evidence currently exists on long-term effects, their lack of tar and very low levels of other dangerous substances<sup>1</sup> suggest they are considerably less harmful than cigarettes.<sup>2</sup> They have been shown to be an efficacious<sup>3</sup> and cost-effective<sup>4</sup> smoking cessation aid. While it will take years to fully estimate the impact of e-cigarette use on diseases including cancer, we can investigate whether it is associated with any biomarkers that may predict future health risk.<sup>5</sup>

Recent studies have found a reduction in harmful biomarkers among e-cigarette users compared with smokers, with some biomarkers showing levels similar to non-smokers.<sup>5-7</sup> However, only a few biomarkers have been investigated and all with relatively short half-lives,<sup>8,9</sup> meaning their utility for predicting long-term effects of e-cigarette use may be limited. DNA methylation is a type of epigenetic modification involving the addition of methyl groups to the DNA which influences how the underlying DNA sequence is interpreted and expressed. Pronounced differences in DNA methylation have been found between cigarette smokers and non-smokers,<sup>10</sup> have been replicated in different populations<sup>11,12</sup> and tissues,<sup>13</sup> have been shown to persist for several years post-cessation,<sup>10</sup> and have been shown to distinguish tumour from normal samples<sup>14</sup>. DNA methylation profiles can act as a biomarker of tobacco smoke exposure, can be used to categorise participants in relation to smoking status, and are predictive of disease and mortality.<sup>15,16</sup>

It has been proposed that the products of tobacco combustion might be responsible for the change in methylation rather than tobacco itself, since no distinctive methylation patterns associated with smokeless tobacco (“snuff”) have been identified.<sup>25</sup> However, this previous study

only investigated peripheral blood and not tissue-specific methylation changes at the site of exposure (i.e. oral samples). Support for the investigation of methylation in saliva has recently been provided given the overlap in methylation signals related to smoke exposure in both blood and saliva<sup>17</sup>, with one study demonstrating a stronger signal in saliva compared with matched blood samples<sup>27</sup>.

Assessing the DNA methylation profile of e-cigarette users could inform our understanding of the potential biological impact associated with their use and the relative health risks compared to cigarettes.<sup>18</sup> We propose that saliva is the most appropriate sample type to conduct an investigation into the potential impact of e-cigarette use on DNA methylation, as it represents the location of direct exposure-tissue contact and is a non-invasive sample which can be easily obtained.

The SEE-Cigs study (Studying the Epigenetics of E-cigarette Use) was set up to recruit a sample of vapers, smokers and non-smokers, from whom we obtained detailed questionnaire data on smoking and vaping habits, demographic characteristics, as well as saliva samples from which DNA methylation was assessed. In the present study, we explored whether e-cigarette use is associated with DNA methylation and evaluated the degree of similarity between DNA methylation profiles in e-cigarette users (vapers) and cigarette smokers (compared with non-smokers) in saliva samples. We investigated associations between e-cigarette use and previously-developed DNA methylation scores used to predict smoking-related disease and mortality,<sup>12 16</sup> and biological ageing.<sup>19 20</sup> We also generated a novel DNA methylation score for predicting e-cigarette use and investigated this in lung tumour and adjacent normal tissue, in order to make

inferences about the potential importance of e-cigarette-related DNA methylation in lung cancer development.

## Methods

The analysis plan was pre-registered<sup>21</sup> and is summarized in **Supplementary Figure 1**.

### Study Design

The SEE-Cigs study (Studying the Epigenetics of E-cigarette Use) recruited vapers, smokers and non-smokers from the United Kingdom general population. It was important that vapers did not have a long previous smoking history, given the persistence of DNA methylation marks associated with smoke exposure many years after cessation<sup>10 22</sup> Vapers were therefore defined as having used e-cigarettes at least weekly for the past 6 months and having smoked <100 times in their lifetime; smokers as having smoked at least weekly for the past 6 months and having used an e-cigarette <100 times in their lifetime; never smokers as having smoked and/or used an e-cigarette <100 times in their lifetime. We aimed to recruit 120 participants per group (vapers, smokers, non-smokers) in order to provide >80% power to detect a 4.5% mean difference in DNA methylation at  $p < 0.05$  and >80% power to detect an 11% mean difference at  $p < 1 \times 10^{-6}$ . Further eligibility criteria and recruitment details are described in the **Online Supplement**.

### Questionnaire

We asked participants questions about their smoking and vaping behavior, as well as socio-demographic and behavioural factors including age, gender, height and weight, ethnicity, education, occupation, household smoking, and recreational drug use. Participants were allocated

to three participant groups (smokers, vapers and never-smokers), after which additional group-specific questions were asked (see **Online Supplement**).

## DNA methylation

Oragene™ kits ([www.dnagenotek.com/US/products/OGR500.html](http://www.dnagenotek.com/US/products/OGR500.html)) were used for collecting saliva samples. Genome-wide methylation status of over 850,000 CpGs was measured using the Illumina HumanMethylationEPIC array according to standard protocol. Details of sample acquisition, processing, data quality control and normalization are provided in the **Online Supplement**.

## Statistical analysis

### Epigenome-wide association study (EWAS)

Multivariable linear regression was used to assess the differences in DNA methylation at each measured CpG site between (1) vapers vs. non-smokers, (2) smokers vs. non-smokers, and (3) smokers vs. vapers, with adjustment for age, biological sex, body mass index (BMI), educational attainment, household smoking, recreational drug use and 20 surrogate variables, using *meffil*.<sup>23</sup> We investigated CpGs which reached a Bonferroni-significance threshold of  $p < 5.91 \times 10^{-8}$  (0.05/846,244 CpGs tested), as well as a less stringent threshold of  $p < 1 \times 10^{-5}$ . From these EWAS results, we identified differentially methylated regions (DMRs) using the *dmrff* R package.<sup>24</sup> DMRs were defined as regions containing at least two CpGs within 500 bp, each with EWAS meta-analysis P-values  $< 0.05$  and methylation changes in a consistent direction, and where the regional p-value surpassed Bonferroni correction. For the EWAS of vapers vs. non-smokers, three additional models were run: i) with adjustment for estimated cell type

composition (see **Online Supplement**), ii) with adjustment for previous smoking history (number of cigarettes) and iii) restricted to individuals of white ethnicity.

### Enrichment and annotation

From the EWAS results of (1) vapers vs. non-smokers, and (2) smokers vs. non-smokers, we investigated evidence for enrichment of associations among 2,623 and 1,501 smoking-related DNA methylation sites identified in previous large-scale studies of blood (Joehanes et al, 2016)<sup>10</sup> and buccal samples (Teschendorff et al, 2015),<sup>14</sup> using a Wilcoxon rank sum test.

While we excluded potential participants who reported drug dependence, given the widespread use of e-cigarettes for inhaling cannabinoids<sup>25</sup>, we assessed whether any of the 15 CpG sites identified in an EWAS of cannabis<sup>26</sup> were associated with e-cigarette use after Bonferroni correction. We also investigated whether there was any evidence for replication of 14 CpGs related to e-cigarette use in a previous EWAS,<sup>27</sup> and assessed the extent to which the CpGs identified in our EWAS had been previously reported in relation to other traits in two publicly available repositories.<sup>28 29</sup> We explored the potential functions of the top 50 CpGs identified in each EWAS via GO and KEGG enrichment analysis using the *missMethyl* R package.<sup>30</sup>

### DNA methylation scores for smoking and epigenetic ageing

DNA methylation scores can be derived by summing methylation values at relevant CpG sites previously identified in relation to a relevant exposure, weighted by the effect sizes observed in independent epigenetic studies. Five DNA methylation scores of smoke exposure (Joehanes, McCartney, Teschendorff, Lu and *AHRR*)<sup>10 14 16 31 32</sup> and four DNA methylation scores of ageing (intrinsic epigenetic age acceleration (IEAA), extrinsic epigenetic age



acceleration (EEAA), PhenoAge and GrimAge)<sup>31 33-35</sup> were quantified. Details of the scores are provided in the **Online Supplement**.

Multivariable linear regression was used to assess differences in DNA methylation scores between the three groups with adjustment for age, sex, BMI, educational attainment, household smoking and recreational drug use. Further analyses were restricted to individuals of white ethnicity only, with adjustment for DNA methylation-derived smoking pack-years in the GrimAge model<sup>31</sup>, and with adjustment for self-reported smoking history when evaluating DNA methylation scores in relation to e-cigarette use.

#### DNA methylation score for e-cigarette use

We generated DNA methylation scores for e-cigarette use and smoking within SEE-Cigs and then assessed the discriminative performance of these scores for predicting e-cigarette use and smoking within SEE-Cigs and in an independent dataset, the Avon Longitudinal Study of Parents and Children (ALSPAC)<sup>36-38</sup>. We also investigated whether the DNA methylation score for e-cigarette use was able to discriminate tumour from normal tissue in lung to the same extent as the DNA methylation score for smoking, using data from publicly available DNA methylation data in the Cancer Genome Atlas (TCGA)<sup>2</sup>. Details of the scores generated and the data used are provided in the **Online Supplement**.

## Results

### Descriptive characteristics

**Figure 1** shows the participant flow for the SEE-Cigs study. The final sample consisted of 117 smokers, 117 non-smokers and 116 vapers with DNA methylation data (see **Online**

**Supplement**). Descriptive characteristics are displayed in **Table 1**. Compared with non-smokers, vapers were more likely to have higher BMI, be male, have lower educational attainment and be more exposed to household smoke, while smokers were more likely to be male, have lower educational attainment, be more exposed to household smoke and to use drugs recreationally. Smokers were slightly older on average than non-smokers and vapers. The majority of participants were of white ethnicity, with a slightly higher proportion of non-white individuals in the group of non-smokers. Smokers had smoked for a median of 1.20 (IQR 0.38, 3.15) pack-years, while both non-smokers and vapers had a minimal smoking history. There were no differences in cell type proportions of the saliva samples obtained from the participants. Most vapers used e-cigarettes containing nicotine and vaped daily.

## EWAS

DNA methylation at 7 CpGs was associated with e-cigarette use (vs non-smoking) at  $p < 1 \times 10^{-5}$  and none at  $p < 5.91 \times 10^{-8}$  (**Table 2**). The top three CpGs were located in protein-coding genes for a ribonuclease P/MRP subunit (*RPP14*), an insulin-like growth factor receptor (*IGF1R*) and a gamma-aminobutyric acid (GABA) A receptor (*GABRP*). After accounting for cell composition, smoking history and ethnicity, associations at the 7 CpGs remained but were weakened, with the exception of signals at cg12435725 (*RPP14*), which strengthened with adjustment for smoking history ( $p = 9.02 \times 10^{-8}$ ), and cg02066693, which was stronger among individuals of white ethnicity ( $p = 6.27 \times 10^{-7}$ ) (**Supplementary Table 1**). DNA methylation at 13 CpGs was associated with smoking (vs non-smoking) at  $p < 1 \times 10^{-5}$  and one at  $p < 5.91 \times 10^{-8}$ : cg05575921 located in *AHRR* encoding the aryl hydrocarbon receptor repressor (**Supplementary Table 2**). DNA methylation at 14 CpGs was associated with e-cigarette use (vs smoking) at  $p < 1 \times 10^{-5}$ , with the top association also at *AHRR* (cg05575921) (**Supplementary Table 3**). DNA

methylation at cg05575921 was 8.2% (95% CI 5.7, 10.5) lower in smokers compared with non-smokers, and 7.1% (95% CI 4.6, 9.6) lower in smokers compared with vapers. In differentially methylated region (DMR) analysis, 31 DMRs were identified in vapers vs non-smokers, 39 in smokers vs. non-smokers and 81 in vapers vs smokers. The top DMR identified between vapers and non-smokers comprised 9 CpGs in *MUC4*, with higher methylation in vapers compared with non-smokers (7.3% (95% CI 5.7, 9.0;  $p=4.13 \times 10^{-18}$ ) (**Supplementary Table 4**).

Apart from associations at *AHRR* in the models involving smoking, there was limited overlap in the top CpGs identified in the three EWAS (**Figure 2, Supplementary Figure 2**). 9 DMRs were found in common between at least two of the EWAS models (**Figure 2**). One DMR was hypermethylated (chr20:*BLCAP;NNAT*) and two hypomethylated (chr20:*SLC2A10* and chr3:*THRB*) in non-smokers compared with vapers and smokers. Two DMRs were hypermethylated (chr10 and chr3:*CACNAID*) and two were hypomethylated (chr17:*BRCA1;NBR2* and chr6:*PRRT1;PPT2*) in smokers compared with non-smokers and vapers. Two DMRs were hypermethylated in vapers compared with smokers and non-smokers (chr10:*ANXA11;LINC00857* and chr17:*HSPB9;KAT2A*).

### Enrichment and annotation

There was strong enrichment of known smoking-related CpGs in the EWAS of smokers vs. non-smokers (Wilcoxon test  $p=1.05 \times 10^{-14}$  for Joehanes et al. CpGs<sup>10</sup> and  $p=1.10 \times 10^{-9}$  for Teschendorff et al. CpGs<sup>14</sup>) (**Supplementary Figure 2**), unlike in the EWAS of vapers vs. non-smokers (Joehanes  $p=0.67$ , Teschendorff  $p=0.28$ ) (**Supplementary Figure 3**). One smoking-related CpG was found to be similarly differentially methylated in vapers in *GABRP* (cg14872828;  $p=2.77 \times 10^{-6}$ ).

One CpG site previously associated with cannabis was found to be associated with e-cigarette use after Bonferroni correction (cg04180046;  $p=0.0029$ ). This CpG has also been previously associated with smoke exposure<sup>10</sup>, and there was no difference in methylation at this site between vapers and smokers in the present study ( $p=0.865$ ) (**Supplementary Table 5**). We also found little evidence of associations between 14 CpGs previously found in relation to e-cigarette use by Song et al.<sup>27</sup> in any of the EWAS after Bonferroni correction ( $p>0.01$ ) (**Supplementary Table 6**).

Three of the seven CpGs associated with e-cigarette use have been identified in previous EWAS for smoking, Down Syndrome, systemic corticosteroid, prostate cancer, gestational age and fetal vs. adult liver (**Supplementary Tables 7 and 8**). We found limited enrichment for KEGG pathways or GO terms (FDR  $p>0.05$ ) (**Supplementary Tables 9-12**). In relation to e-cigarette use, response to ethanol/alcohol, positive regulation of insulin secretion and GABA transport were the top GO terms. Butanoate metabolism, synaptic vesicle cycle and GABAergic synapse were the top KEGG pathways.

#### DNA methylation scores for smoking

All of the DNA methylation scores for smoking were correlated with reported pack-years smoked ( $|r|=0.21-0.52$ ,  $p<0.0001$ ), except the Teschendorff score<sup>14</sup> ( $r=0.07$ ,  $p=0.19$ ) (**Supplementary Table 13**). All of the scores differed between the smokers vs. non-smokers (0.40-0.90 SD), and vapers vs. smokers (0.37-0.86 SD). While neither the Joehanes<sup>10</sup> nor the Teschendorff score<sup>14</sup> differed between vapers vs. non-smokers, higher levels of the other three scores (Lu,<sup>31</sup> McCartney,<sup>16</sup> *AHRR*) were observed in vapers vs. non-smokers (0.21-0.30 SD) (**Table 3**). These associations attenuated when reported smoking history was included in the

model (**Supplementary Table 14**). Associations were similar in analyses restricted to individuals of white ethnicity (**Supplementary Table 15**).

#### DNA methylation scores for epigenetic ageing

Epigenetic age estimates were strongly correlated with chronological age ( $r = 0.49-0.70$ ,  $p < 0.0001$ ) (**Supplementary Table 16**). There was little difference in estimates of epigenetic age acceleration (EAA) between the three groups, except for GrimAge, where smokers had higher EAA relative to both vapers (2.5, 95% CI = 1.4-3.6 years) and non-smokers (2.6, 1.4-3.7 years) (**Table 3**). Associations persisted in analyses adjusting for a DNA methylation score for smoking pack-years<sup>31</sup> (**Supplementary Table 17**) and when restricted to individuals of white ethnicity (**Supplementary Table 15**).

#### Using DNA methylation as a biomarker of e-cigarette use and lung cancer

We generated DNA methylation scores for predicting e-cigarette use (**Supplementary Table 18**). These performed poorly at discriminating e-cigarette users from non-smokers in both the internal (AUC 0.55, 0.41-0.69) and external validation sets (AUC 0.57, 0.36-0.74). This was in contrast to the high discriminative performance of the smoking scores (internal AUC 0.80, 0.69-0.91 and external AUC 0.80, 0.72-0.88; **Supplementary Figure 4**). The DNA methylation scores for both smoking and e-cigarette use performed poorly at discriminating tumour from adjacent normal tissue in lung adenocarcinoma (LUAD) cases (AUC 0.57, 0.40-0.73 and 0.58, 0.42-0.74). The e-cigarette score also performed poorly at discriminating tumour from adjacent normal tissue in lung squamous cell carcinoma (LUSC) cases (0.64, 0.52-0.76). Moderate discrimination was observed from the smoking score in LUSC cases (0.73, 0.61-0.85) (**Supplementary Figure 5**).

## Discussion

Using saliva samples from 117 smokers, 117 non-smokers and 116 vapers with a limited smoking history, we found that the DNA methylation signals of e-cigarette use were weak and largely distinct from those established in relation to cigarette smoking. In particular, a lack of enrichment for previous smoking-related CpG sites was observed in the vapers compared with the smokers. We found modest evidence for differential methylation related to e-cigarette use at a number of sites or regions proximal to genes related to metabolism (*IGF1R*, *SLC2A10*, *THRB*), cancer (*IGF1R*, *BLCAP*, *LINC00857*) and the nervous system (*NNAT*, *ANXA11*, *GABRP*). Vapers also exhibited differential methylation relative to non-smokers in *MUC4* encoding Mucin 4, an integral membrane glycoprotein present in mucus which has previously been found to be upregulated in vapers.<sup>39</sup> Ethanol/alcohol, positive regulation of insulin secretion, GABA transport and butanoate metabolism were among the most enriched pathways, which may reflect biological responses to e-cigarette constituents (ethyl alcohol, ethyl butyrate and nicotine).

Two previous studies (comprising 32 and 45 participants, respectively) have found associations between e-cigarette use and DNA methylation levels which overlapped with smoking-related signals.<sup>27 40</sup> This is in contrast to the present study where DNA methylation scores for smoking were similar between vapers and non-smokers. We were also unable to replicate DNA methylation differences for 14 CpGs previously related to e-cigarette use<sup>27</sup>. Previous findings may therefore represent false positives given previously small sample sizes or residual confounding by smoking status. Alternatively, the lack of replication could be due to tissue differences, since the previous study assessed methylation in bronchoalveolar samples.

We also investigated DNA methylation scores for biological ageing, and found that, while these were again similar between vapers and non-smokers, higher levels of a biological ageing score (GrimAge) were observed in smokers. This is of interest since epigenetic age is predictive of age-related disease and mortality independent of chronological age,<sup>41-43</sup> and so may serve as a predictor of higher mortality rates among smokers.

Finally, a DNA methylation score generated to index e-cigarette use poorly discriminated vapers from non-smokers in SEE-Cigs and in an independent dataset (ALSPAC). This was in contrast to a DNA methylation score generated to index smoking, which discriminated smokers from non-smokers in both datasets. The smoking score also showed better discrimination of tumour and adjacent normal tissue in lung squamous cell cases compared with the e-cigarette score. These findings are supported by a previous study which found that a DNA methylation profile for smoking derived from oral samples was able to discriminate tumour and adjacent normal samples with high accuracy.<sup>14</sup> This indicates that DNA methylation scores for smoking may incorporate more epigenetic aberrations relevant to tumourigenesis than an equivalent score for e-cigarette use. However, the smoking DNA methylation score generated in the present study was lower in lung squamous cell carcinoma relative to normal tissue, the inverse of what was expected due to higher levels being observed in smokers.<sup>14</sup> While this is surprising, we have previously found differential methylation in tumour tissue is in the opposite direction to that observed in relation to smoking for *AHRR* (cg05575921)<sup>44</sup> (**Supplementary Figure 5**), the CpG contributing most weight to the DNA methylation score for smoking (**Supplementary Table 18**).

Major strengths relate to the design of the study, including the recruitment individuals with a limited smoking history and the assessment of DNA methylation levels in an easily accessible and exposure-relevant tissue in order to investigate epigenetic profiles of e-cigarette use. Limitations include the representativeness of our study sample, with demographic characteristics different to the general population due to the strict inclusion criteria. The young age of the study sample (mean age = 21 years) and limited smoking and vaping history could hamper the detection of DNA methylation signals. This may explain why so few CpG sites were identified at epigenome-wide significance in relation to smoking than anticipated based on previous studies of oral samples<sup>14</sup>. Nonetheless, we found evidence for enrichment of smoking-related CpG sites among the smokers, suggesting that the signals were present but below the strict Bonferroni-correction threshold. Similar enrichment was not found among the vapers, indicating that the methylation signature of e-cigarette use is distinct from that of smoking, and that vapers in the present study were accurate in their reporting of having a limited smoking history, despite this not being biochemically verified.

Additional research in cohort studies is required to investigate DNA methylation changes among ex-smokers quitting with different methods, including e-cigarettes. It would be of particular interest to investigate dual users in comparison with exclusive smokers and vapers to assess whether methylation signals are more prominent in this group, who make up the majority of e-cigarette users. While findings from this study suggest that e-cigarettes may have distinct health effects from cigarettes, we cannot provide robust conclusions regarding the safety of e-cigarettes. Furthermore, although the DNA methylation changes identified in relation to both smoking and e-cigarette use may be predictive of future disease risk, the causal consequences of these DNA methylation changes on health outcomes are currently uncertain.<sup>44</sup>



Our study was powered to detect an 11% mean difference in DNA methylation at  $p < 10^{-6}$ . This is below that typically required for obtaining epigenome-wide significant findings, although is adequate for ensuring replication of previously-associated CpGs, such as that at *AHRR* (cg05575921) in relation to smoke exposure, which has been found to exhibit up to 40% higher DNA methylation in current versus never smokers in both blood and saliva samples<sup>17</sup>. Therefore, while we can be confident in our finding that differential methylation at this site is specific to smokers and not vapers, we were less powered to detect novel CpGs in relation to e-cigarette use. While it appears that the DNA methylation profile of vapers is less pronounced than that of smokers, the DNA methylation changes associated with e-cigarette use may be commensurate in scale with other lifestyle exposures and replication of the signals identified in relation to e-cigarette use in larger studies is warranted.

## Conclusions

Findings from this study suggest that e-cigarette use does not impact saliva DNA methylation in the same way as cigarette smoking. Unlike for smoking, the DNA methylation profile for e-cigarette use did not replicate in independent samples and was not able to discriminate cancer from normal tissue. Further studies are required to detect a robust methylation signature for long-term e-cigarette use. The extent to which differential methylation related to e-cigarette use translates into chronic effects and relevant health outcomes should also be investigated.

## Acknowledgements

**SEE-Cigs** – We are extremely grateful to all the participants who took part in this study.

**ALSPAC** - We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

### **Author contributions:**

Study conception and design: R.C.R., G.S., M.S., C.L.R., M.M., S.H.G.

Data acquisition, analysis and interpretation of data: R.C.R., C.S.R., J.N.K., C.P., A.B., S.H.G.

Drafting the manuscript: R.C.R., C.S.R., S.H.G.

Critically revising the manuscript: J.N.K., C.P., A.B., G.S., M.S., C.L.R., M.M.

Final approval: All authors

### **Funding:**

This work was supported by a Cancer Research UK Population Research Committee project grant (C57854/A22171). Further support was provided by the UK Medical Research Council (MC\_UU\_00011/5 and MC\_UU\_00011/7), which funds a Unit at the University of Bristol where R.C.R., C.S.R., J.K., C.P., G.S., M.S., C.L.R and M.M. work, and the CRUK-funded Integrative Cancer Epidemiology Programme (C18281/A19169). R.C.R is a de Pass Vice Chancellor's Research Fellow at the University of Bristol. C.S.R's time is supported by the National Institute for Health Research Applied Research Collaboration West (NIHR ARC West) at University Hospitals Bristol NHS Foundation Trust. The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This

publication is the work of the authors who will serve as guarantors for the contents of this paper.

A comprehensive list of grants funding is available on the ALSPAC website

(<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>).

**Competing interests:** None to declare.

## References

1. Goniewicz ML, Knysak J, Gawron M, et al. Levels of selected carcinogens and toxicants in vapour from electronic cigarettes. *Tob Control* 2014;23(2):133-9. doi: 10.1136/tobaccocontrol-2012-050859 [published Online First: 2013/03/08]
2. McNeill A, Brose LS, Calder R, et al. E-cigarettes: an evidence update. *Public Health England* 2015;3
3. Hajek P, Phillips-Waller A, Przulj D, et al. A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy. *N Engl J Med* 2019;380(7):629-37. doi: 10.1056/NEJMoa1808779 [published Online First: 2019/01/31]
4. Li J, Hajek P, Pesola F, et al. Cost-effectiveness of e-cigarettes compared with nicotine replacement therapy in stop smoking services in England (TEC study): a randomized controlled trial. *Addiction* 2020;115(3):507-17. doi: 10.1111/add.14829 [published Online First: 2019/10/10]
5. McNeill A, Brose LS, Calder R, et al. Evidence review of e-cigarettes and heated tobacco products 2018. *A report commissioned by Public Health England London: Public Health England* 2018;6
6. Shahab L, Goniewicz ML, Blount BC, et al. Nicotine, Carcinogen, and Toxin Exposure in Long-Term E-Cigarette and Nicotine Replacement Therapy Users: A Cross-sectional Study. *Ann Intern Med* 2017;166(6):390-400. doi: 10.7326/M16-1107 [published Online First: 2017/02/07]
7. Goniewicz ML, Gawron M, Smith DM, et al. Exposure to Nicotine and Selected Toxicants in Cigarette Smokers Who Switched to Electronic Cigarettes: A Longitudinal Within-Subjects Observational Study. *Nicotine Tob Res* 2017;19(2):160-67. doi: 10.1093/ntr/ntw160 [published Online First: 2016/09/11]
8. Sandberg A, Skold CM, Grunewald J, et al. Assessing recent smoking status by measuring exhaled carbon monoxide levels. *PLoS One* 2011;6(12):e28864. doi: 10.1371/journal.pone.0028864 [published Online First: 2011/12/24]
9. Goniewicz ML, Havel CM, Peng MW, et al. Elimination kinetics of the tobacco-specific biomarker and lung carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol. *Cancer Epidemiol Biomarkers Prev* 2009;18(12):3421-5. doi: 10.1158/1055-9965.EPI-09-0874 [published Online First: 2009/12/05]
10. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* 2016;9(5):436-47. doi: 10.1161/CIRCGENETICS.116.001506 [published Online First: 2016/10/21]
11. Philibert RA, Beach SR, Brody GH. Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. *Epigenetics* 2012;7(11):1331-8. doi: 10.4161/epi.22520 [published Online First: 2012/10/17]
12. Elliott HR, Tillin T, McArdle WL, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics* 2014;6(1):4. doi: 10.1186/1868-7083-6-4 [published Online First: 2014/02/04]
13. Monick MM, Beach SR, Plume J, et al. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet B Neuropsychiatr Genet* 2012;159B(2):141-51. doi: 10.1002/ajmg.b.32021 [published Online First: 2012/01/11]
14. Teschendorff AE, Yang Z, Wong A, et al. Correlation of Smoking-Associated DNA Methylation Changes in Buccal Cells With DNA Methylation Changes in Epithelial Cancer. *JAMA Oncol* 2015;1(4):476-85. doi: 10.1001/jamaoncol.2015.1053 [published Online First: 2015/07/17]

15. Bojesen SE, Timpson N, Relton C, et al. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax* 2017;72(7):646-53. doi: 10.1136/thoraxjnl-2016-208789 [published Online First: 2017/01/20]
16. McCartney DL, Hillary RF, Stevenson AJ, et al. Epigenetic prediction of complex traits and death. *Genome Biol* 2018;19(1):136. doi: 10.1186/s13059-018-1514-1 [published Online First: 2018/09/28]
17. Philibert R, Dogan M, Beach SRH, et al. AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. *Am J Med Genet B Neuropsychiatr Genet* 2020;183(1):51-60. doi: 10.1002/ajmg.b.32760 [published Online First: 2019/08/29]
18. Freudenheim JL, Shields PG, Song MA, et al. DNA Methylation and Smoking: Implications for Understanding Effects of Electronic Cigarettes. *Current Epidemiology Reports* 2019;6(2):148-61.
19. Bell CG, Lowe R, Adams PD, et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol* 2019;20(1):249. doi: 10.1186/s13059-019-1824-y [published Online First: 2019/11/27]
20. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* 2018;19(6):371-84. doi: 10.1038/s41576-018-0004-3 [published Online First: 2018/04/13]
21. Gage SH, Richmond RC, Board AG, et al. DNA methylation in e-cigarette users versus cigarette smokers and never-smokers: Studying the Epigenetics of E-Cigarette Use (SEE-Cig). *osfio/aenbh* 2017
22. Guida F, Sandanger TM, Castagne R, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum Mol Genet* 2015;24(8):2349-59. doi: 10.1093/hmg/ddu751 [published Online First: 2015/01/04]
23. Min JL, Hemani G, Davey Smith G, et al. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* 2018;34(23):3983-89. doi: 10.1093/bioinformatics/bty476 [published Online First: 2018/06/23]
24. Suderman M, Staley JR, French R, et al. dmrff: identifying differentially methylated regions efficiency with power and control. *bioRxiv* 2018;508556
25. Giroud C, de Cesare M, Berthet A, et al. E-Cigarettes: A Review of New Trends in Cannabis Use. *Int J Environ Res Public Health* 2015;12(8):9988-10008. doi: 10.3390/ijerph120809988 [published Online First: 2015/08/27]
26. Osborne AJ, Pearson JF, Noble AJ, et al. Genome-wide DNA methylation analysis of heavy cannabis exposure in a New Zealand longitudinal cohort. *Transl Psychiatry* 2020;10(1):114. doi: 10.1038/s41398-020-0800-3 [published Online First: 2020/04/24]
27. Song MA, Freudenheim JL, Brasky TM, et al. Biomarkers of Exposure and Effect in the Lungs of Smokers, Nonsmokers, and Electronic Cigarette Users. *Cancer Epidemiol Biomarkers Prev* 2020;29(2):443-51. doi: 10.1158/1055-9965.EPI-19-1245 [published Online First: 2019/12/19]
28. Staley JR, Battram T, Hemani G, et al. <http://www.ewascatalog.org/> 2019
29. Li M, Zho D, Gao R, et al. <http://bigd.big.ac.cn/ewas/index>. 2019
30. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* 2016;32(2):286-8. doi: 10.1093/bioinformatics/btv560 [published Online First: 2015/10/02]
31. Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)* 2019;11(2):303-27. doi: 10.18632/aging.101684 [published Online First: 2019/01/23]

32. Philibert RA, Beach SR, Lei MK, et al. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clin Epigenetics* 2013;5(1):19. doi: 10.1186/1868-7083-5-19 [published Online First: 2013/10/15]
33. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013;14(10):R115. doi: 10.1186/gb-2013-14-10-r115 [published Online First: 2013/10/22]
34. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 2013;49(2):359-67. doi: 10.1016/j.molcel.2012.10.016 [published Online First: 2012/11/28]
35. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* 2018;10(4):573-91. doi: 10.18632/aging.101414 [published Online First: 2018/04/21]
36. Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013;42(1):111-27. doi: 10.1093/ije/dys064 [published Online First: 2012/04/18]
37. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* 2013;42(1):97-110. doi: 10.1093/ije/dys066 [published Online First: 2012/04/18]
38. Northstone K, Lewcock M, Groom A, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Res* 2019;4:51. doi: 10.12688/wellcomeopenres.15132.1 [published Online First: 2019/04/26]
39. Ghosh A, Coakley RC, Mascenik T, et al. Chronic E-Cigarette Exposure Alters the Human Bronchial Epithelial Proteome. *Am J Respir Crit Care Med* 2018;198(1):67-76. doi: 10.1164/rccm.201710-2033OC [published Online First: 2018/02/27]
40. Caliri AW, Caceres A, Tommasi S, et al. Hypomethylation of LINE-1 repeat elements and global loss of DNA hydroxymethylation in vapers and smokers. *Epigenetics* 2020:1-14. doi: 10.1080/15592294.2020.1724401 [published Online First: 2020/01/31]
41. Marioni RE, Shah S, McRae AF, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* 2015;16:25. doi: 10.1186/s13059-015-0584-6 [published Online First: 2015/01/31]
42. Chen BH, Marioni RE, Colicino E, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* 2016;8(9):1844-65. doi: 10.18632/aging.101020 [published Online First: 2016/10/01]
43. Perna L, Zhang Y, Mons U, et al. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics* 2016;8:64. doi: 10.1186/s13148-016-0228-z [published Online First: 2016/06/09]
44. Battram T, Richmond RC, Baglietto L, et al. Appraising the causal relevance of DNA methylation for risk of lung cancer. *Int J Epidemiol* 2019;48(5):1493-504. doi: 10.1093/ije/dyz190 [published Online First: 2019/09/25]

## Figure legends

**Figure 1:** Participant flow chart for SEE-Cigs

**Figure 2:** Comparison of epigenome-wide associations studies

A) EWAS for e-cigarette use (vs. non-smoking) and smoking (vs. non-smoking) B) EWAS for e-cigarette use (vs. non-smoking) and e-cigarette use (vs. smoking) C) EWAS for smoking (vs. non-smoking) and e-cigarette use (vs. smoking)

## Tables

**Table 1:** Descriptive characteristics of participant groups in this study

	<b>Non-smokers (N = 117)</b>	<b>Smokers (N = 117)</b>	<b>Vapers (N = 116)</b>	<b>Overall P for difference</b>
<b>Socio-demographic factors</b>	<b>mean (SD)</b>	<b>mean (SD)</b>	<b>mean (SD)</b>	
Age (years)	20.6 (3.5)	22.8 (4.8)	20.9 (4.3)	0.044
BMI (kg/m <sup>2</sup> )	22.6 (3.5)	23.1 (4.2)	26.1 (7.0)	<0.001
	<b>median (IQR)</b>	<b>median (IQR)</b>	<b>median (IQR)</b>	
Smoking (pack-years)	0 (0, 0.0001)	1.20 (0.38, 3.15)	0.0013 (0.0006, 0.0027)	<0.001
	<b>% (n)</b>	<b>% (n)</b>	<b>% (n)</b>	
Sex (% male)	50.0 (60)	50 (60)	80 (73)	<0.001
Ethnicity (% white)	80 (96)	89.1 (106)	88.7 (102)	0.075
Education (% higher education)	86.7 (104)	79.0 (94)	60 (69)	<0.001
Occupation (% unemployed)	4.2 (5)	8.4 (10)	21 (18.3)	0.001
Household smoking (% yes)	40 (48)	67.5 (81)	67.2 (78)	<0.001
Recreational drug use (% yes)	5.8 (7)	26.3 (31)	9.3 (11)	<0.001
<b>Estimated cell type proportions (%)</b>	<b>mean (SD)</b>	<b>mean (SD)</b>	<b>mean (SD)</b>	
Buccal	48.3 (18.2)	49.4 (18.5)	45.4 (17.1)	0.201
Bcell	1.7 (1.2)	2.1 (1.8)	2.2 (2.1)	0.100
CD4T	0.2 (0.4)	0.2 (0.5)	0.2 (0.7)	0.808
CD8T	0.001 (0.005)	0.04 (0.44)	0.03 (0.24)	0.477
Granulocytes	47.3 (19.8)	45.4 (20.8)	49.1 (19.3)	0.357
Mono	0.4 (0.7)	0.5 (0.9)	0.6 (1.0)	0.307
NK	1.1 (1.4)	1.5 (2.6)	1.6 (2.3)	0.183
<b>Vaping characteristics</b>	<b>% (n)</b>	<b>% (n)</b>	<b>% (n)</b>	
E-cigarette device	N/A	N/A		N/A
Cigalike (1st generation)			1.7 (2)	
Tanks (2nd generation)			9.4 (11)	
Mods (3rd generation)			67.3 (76)	
Pods (4th generation)			15.0 (17)	
Other			6.2 (7)	
Frequency of use (% daily)			92.9 (105)	
Device contains nicotine (% yes)			96.4 (108)	



Increased nicotine concentration in the past (% yes)			14.1 (16)	
	<b>mean (SD)</b>	<b>mean (SD)</b>	<b>mean (SD)</b>	
Nicotine content (mg/ml)	N/A	N/A	5.6 (4.7)	N/A
Liquid per day (ml)			7.6 (7.6)	

**Table 2:** Differentially methylated CpG sites associated with e-cigarette use vs non-smoking

CpG site	Chromosome	Position	Gene Symbol	Beta	SE	P-value*
<b>cg12435725</b>	3	58293450	<i>RPP14</i>	- 0.060	0.012	6.43x10 <sup>-7</sup>
<b>cg02066693</b>	15	99366135	<i>IGF1R</i>	- 0.019	0.004	5.47x10 <sup>-6</sup>
<b>cg14872828</b>	5	170210761	<i>GABRP</i>	- 0.035	0.007	2.77x10 <sup>-6</sup>
<b>cg12734956</b>	12	112181430	<i>ACAD10</i>	- 0.013	0.003	5.89x10 <sup>-6</sup>
<b>cg02934082</b>	3	122705793	<i>SEMA5B</i>	- 0.036	0.008	5.92x10 <sup>-6</sup>
<b>cg00388391</b>	1	18459578	<i>IGSF21</i>	0.025	0.006	8.38x10 <sup>-6</sup>
<b>cg10440286</b>	22	37771664	<i>ELFN2</i>	0.044	0.010	9.31x10 <sup>-6</sup>

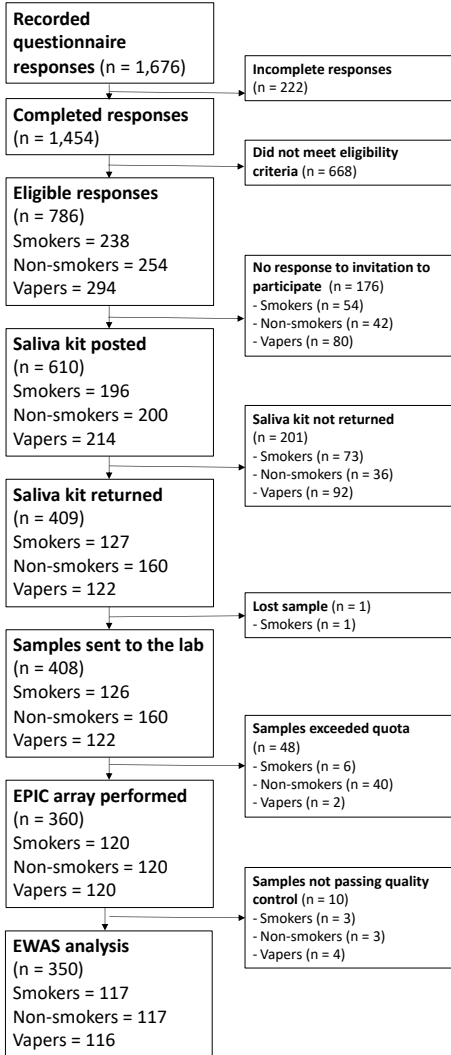
Model adjusted for age, sex, body mass index, educational attainment, household smoking, recreational drug use and 20 surrogate variables. n=111 vapers, n=117 non-smokers ; P-values <1x10<sup>-5</sup>

**Table 3:** Differences in DNA methylation scores between participant groups

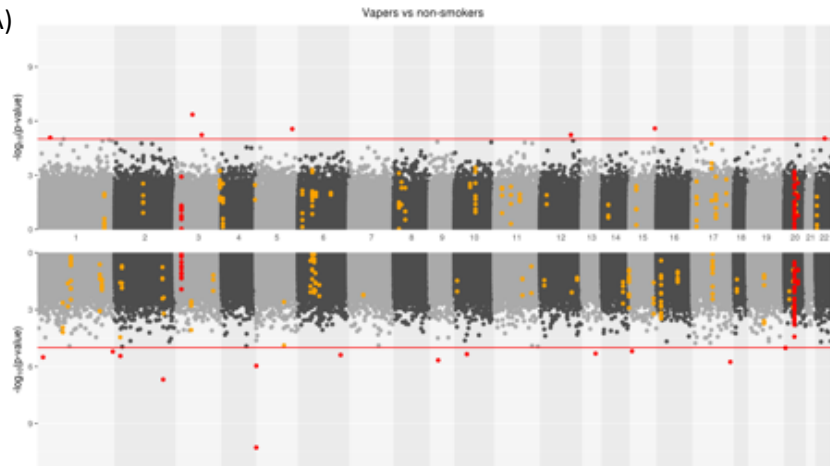
DNA methylation score		Smokers vs non-smokers (n= 231)		Vapers vs non-smokers (n = 228)		Vapers vs smokers (n = 225)	
		Coefficient (SE)	P-value	Coefficient (SE)	P-value	Coefficient (SE)	P-value
Smoking*	Joehanes <sup>10</sup>	0.62 (0.15)	4.39 x 10 <sup>-5</sup>	0.13 (0.15)	0.360	-0.59 (0.14)	3.25x10 <sup>-5</sup>
	Teschendorff <sup>14</sup>	0.40 (0.16)	0.010	0.12 (0.14)	0.864	-0.34 (0.15)	0.024
	Lu <sup>31</sup>	0.65 (0.11)	7.9 x 10 <sup>-9</sup>	0.30 (0.10)	0.002	-0.47 (0.13)	1.98x10 <sup>-4</sup>
	McCartney <sup>16</sup>	0.83 (0.13)	2.14 x 10 <sup>-10</sup>	0.30 (0.12)	0.012	-0.68 (0.13)	3.67x10 <sup>-7</sup>
	AHRR <sup>32</sup>	-0.90 (0.13)	1.12 x 10 <sup>-11</sup>	-0.23 (0.11)	0.28	0.82 (0.14)	1.62x10 <sup>-9</sup>
Epigenetic age**	IEAA <sup>33</sup>	-0.02 (0.68)	0.981	-0.08 (0.64)	0.907	0.41 (0.67)	0.538
	EEAA <sup>42</sup>	0.91 (0.73)	0.212	0.38 (0.69)	0.578	-0.68 (0.69)	0.325
	PhenoAge <sup>35</sup>	0.26 (0.81)	0.751	-0.20 (0.86)	0.816	0.11 (0.85)	0.897
	GrimAge <sup>31</sup>	2.57 (0.59)	1.36 x 10 <sup>-5</sup>	0.74 (0.55)	0.179	-241 (0.57)	2.09 x 10 <sup>-5</sup>

Adjusted for age, sex, body mass index, educational attainment, household smoking and recreational drug use

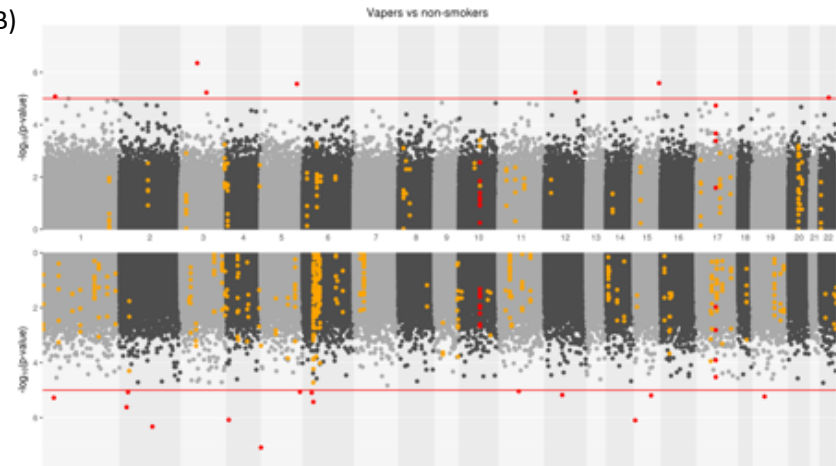
\*coefficients = SD units; \*\*coefficients = years



A)



B)



C)

