



**HAL**  
open science

## **Alcohol consumption and DNA methylation: an epigenome-wide association study within the French E3N cohort**

Dzevka Dragic, Fanny Artaud, Mojgan Karimi, Thérèse Truong, Laura Baglietto, Jean-François Deleuze, Caroline Diorio, Gianluca Severi

### ► To cite this version:

Dzevka Dragic, Fanny Artaud, Mojgan Karimi, Thérèse Truong, Laura Baglietto, et al.. Alcohol consumption and DNA methylation: an epigenome-wide association study within the French E3N cohort. *Clinical Epigenetics*, 2025, 17 (1), pp.118. <10.1186/s13148-025-01893-1>. <hal-05159990>

**HAL Id: hal-05159990**

**<https://hal.science/hal-05159990v1>**

Submitted on 12 Jul 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

RESEARCH

Open Access



# Alcohol consumption and DNA methylation: an epigenome-wide association study within the French E3N cohort

Dzevka Dragic<sup>1,2,3</sup>, Fanny Artaud<sup>3</sup>, Mojgan Karimi<sup>3</sup>, Thérèse Truong<sup>3</sup>, Laura Baglietto<sup>3,4</sup>, Jean-François Deleuze<sup>5</sup>, Caroline Diorio<sup>1,2,6</sup> and Gianluca Severi<sup>3,7\*</sup>

## Abstract

**Background** Alcohol consumption can have harmful effects on health, depending on the quantity and frequency. Understanding the underlying molecular mechanisms is essential to grasp its health consequences. The study aimed to assess the association between alcohol consumption and blood DNA methylation, an epigenetic mechanism that controls gene expression.

**Methods** The epigenome-wide association study (EWAS) included 1,538 women from a case-cohort study within the French E3N cohort. Weighted linear mixed-effects models were used to assess the associations between self-reported alcohol consumption (in g/day in 1993) and DNA methylation at 715,986 CpGs measured with the HumanMethylationEPIC Beadchip. Women were cancer-free at blood collection in 1995–1999.

**Results** Of the 715,986 sites analyzed, 19,255 were associated with alcohol consumption (FDR < 0.05). Over-representation analysis highlighted enrichment of genes involved in cancer, the nervous system and aging. Of these 19,255 sites, 1,528 were replicated in an independent case–control study, with 85 also identified in other EWAS. Notably, at least six studies reported sites in *SLC7A11*, *ANP32B*, *MCM2*, *HNRNPA1*, *SNORD30*, and *TRA2B* genes.

**Conclusions** Several potential methylation markers for alcohol consumption, documented prior to blood sampling, have been identified. The link between these sites and chronic diseases should be investigated to understand the molecular mechanisms underlying the harmful effects of alcohol consumption on health.

**Keywords** Alcohol consumption, DNA methylation, Epigenetic biomarkers, Case-cohort design, Epigenome-wide association study

<sup>†</sup>Caroline Diorio and Gianluca Severi contributed equally to this work as co-last authors.

\*Correspondence:

Gianluca Severi  
gianluca.severi@inserm.fr

<sup>1</sup> Department of Social and Preventive Medicine, Faculty of Medicine, Université Laval, Quebec, QC, Canada

<sup>2</sup> Oncology Division, Cancer Research Center, CHU de Québec Research Center, Quebec, QC, Canada

<sup>3</sup> Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, U1018, INSERM, Hôpital Paul Brousse, Bat 15/16, 16 Avenue Paul Vaillant Couturier, 94807 Villejuif Cedex, France

<sup>4</sup> Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

<sup>5</sup> Centre National de Recherche en Génomique Humaine (CNRGH), Université Paris-Saclay, CEA, 91057 Evry, France

<sup>6</sup> Deschênes-Fabia Center for Breast Diseases, Saint-Sacrement Hospital, Quebec, QC, Canada

<sup>7</sup> Department of Statistics, Computer Science and Applications "G. Parenti" (DISIA), University of Florence, Florence, Italy



## Background

Alcohol consumption is a widespread behavior in the world, driven by cultural and social influences, as well as country-specific policies. Depending on the quantity and frequency of consumption, this practice can have numerous harmful effects on health. Indeed, epidemiological studies have suggested that alcohol consumption is linked to an increased risk of chronic diseases such as cancer, cardiovascular disease, and metabolic disorders [1]. Yet, the underlying biological mechanisms are not completely well-defined. Investigating the link between alcohol consumption and chronic diseases at the epigenetic level is a promising approach to discovering potential targets.

DNA methylation (DNAm) is a potentially reversible epigenetic mechanism that plays a role in gene regulation by modifying gene expression without altering the DNA sequence [2]. The DNAm reaction is the addition of a methyl group to the fifth carbon of a cytosine, mostly in Cytosine-phosphate-Guanine (CpG) dinucleotides. Apart from its hereditary nature, a large number of studies have suggested that environmental and lifestyle factors, including alcohol consumption, may have an impact on DNAm [3]. Alterations in DNAm caused by alcohol consumption can lead to changes in gene expression and thus contribute to the development of chronic diseases. Understanding the molecular mechanisms underlying this association is essential for gaining insight into the health consequences that result from alcohol consumption. To study the association between alcohol consumption and DNAm, the ideal study design would have repeated measurements of both variables available. However, obtaining multiple measurements of alcohol consumption prior to blood collection is not common, and measuring DNAm in multiple blood samples is rare and often very costly.

We published a systematic literature review in 2022 on the association between alcohol consumption and epigenome-wide DNAm in blood [4]. Eleven cross-sectional studies published between 2015 and 2020 were included [5–15], and an additional study was identified after the databases were searched [16]. While numerous CpGs have been linked to alcohol consumption, only a limited fraction has been consistently identified across multiple studies. Furthermore, the literature lacks longitudinal studies and investigations by alcohol subtype.

The aim of this study was to assess the association between alcohol consumption and blood epigenome-wide DNAm in a case-cohort design, using both total alcohol intake and alcohol subtypes. As we do not have repeated measurements to conduct longitudinal analysis, we used the prospective aspect of the cohort to take into account the temporality of our variables of interest,

which provides added value compared to cross-sectional studies.

## Methods

### Study population

The French E3N-Generations is a prospective familial cohort originating from the E3N cohort of 98,995 women affiliated to the MGEN, the health insurance provider for the national education system [17]. The present study focuses only on the E3N original cohort. E3N women, who were aged between 40 and 65 at recruitment in 1990, are still being followed up after 30 years through self-reported questionnaires every 2–3 years. The average response rate for the 12 questionnaires was over 80%. Additionally, blood samples were collected from 23,674 volunteers between 1995 and 1999 and stored in liquid nitrogen as aliquots of buffy coat, plasma, serum, and erythrocytes. The study received ethical approval from the French National Commission for Data Protection and Privacy (Commission Nationale Informatique et Liberté) and all participants gave written informed consent. The study identifier at ClinicalTrials.gov is NCT03285230.

A subset of participants with an available blood sample was selected for a case-cohort study. From the full cohort of participants with a blood sample available and no prevalent cancer at blood collection ( $n=21,653$  including 1,640 incident breast cancer cases diagnosed after blood collection), a sub-cohort of 734 women including 33 with incident breast cancer was randomly selected, stratified on age at blood draw (in 5-years categories: [44–50]; [50–55]; [55–60]; [60–74]) and teaching status in 1992 (teacher/non-teacher/missing). Among the 1,640 incident breast cancer cases, 1,059 were included (of which 33 were part of the sub-cohort) after excluding participants without available buffy coat samples or those who already had DNAm data. The case-cohort study included 1,760 women, 1,059 incident breast cancer cases and 701 non-cases. The workflow of the case-cohort selection is presented in supplementary material (Additional File 1: Fig. S1).

### Alcohol consumption

Information on alcohol consumption was collected in 1993 using a 238-item self-completed food frequency questionnaire on daily consumption [18]. Participants were asked to indicate the frequency of consumption (by ticking 'never/less than once a month' or indicating the number of times per month or per week) and quantity consumed (by indicating the number of glasses and specifying the type of glass using photos) for several types of alcohol such as beer, cider, wine/champagne, muscat/porto/vermouth, whisky/gin/vodka, aniseed-flavored alcohol, cocktail/punch/kir, and digestif. Based on the

data collected, the total alcohol consumption was calculated in grams per day [19]. We considered three subtypes of alcohol in milliliters per day: wine/champagne, cider/beer, and spirits.

### DNA methylation

DNA was extracted from the buffy coat of 1,760 women using Chemagic technology for over 76% of the samples, and Puregene Qiagen for the remaining samples. Differences due to the DNA extraction kit were not expected, as blood sample processing and sample storage have minimal impact on DNAm [20]. The Infinium Human-MethylationEPIC Beadchip (Illumina, California, USA) was used by the French National Centre for Research on Human Genomics (CNRGH) in Evry, France, to measure methylation levels of > 850,000 CpGs in enhancer regions, gene bodies, promoters, and CpG islands. Samples were allocated on chips at random using the OSAT R package (version 4.1.2) and ensuring that the distribution was balanced on the following variables: case/non-case status, time from blood sampling to breast cancer diagnosis (in years: [0–4]; [4–8]; [8–12]; [12–16]; [16–20]), and birth generation (< 1935; [1935–1940]; [1940–1945]; ≥ 1945). Quality control and data preprocessing were completed using R software version 4.2.3. Raw methylation data (IDAT files) were imported using the Minfi package. Quality control was conducted using Minfi and shinyMethyl packages. Samples with more than ten percent of failed probes (detection *P* value > 0.01, *n* = 8) or a mismatch between predicted sex and recorded sex (*n* = 1) were excluded. Methylation data were normalized using preprocessIllumina from Minfi package and corrected for the probe design type bias using regression on correlated probes (RCP) method of the ENmix package [21]. Then, we also excluded participants with missing data for alcohol consumption (*n* = 203) or missing data for one of the covariates (smoking *n* = 5 and body mass index *n* = 6). In total, 1,538 women (926 cases and 612 non-cases) were included in the epigenome-wide association study (EWAS). Probes that failed in several samples (mean detection *P* values > 0.01 across the 1,538 samples, *n* = 6,373), probes overlapping with single-nucleotide polymorphisms (SNP, *n* = 91,591), cross-reactive probes (*n* = 42,558) [22], and probes located on the X (*n* = 19,090) and Y (*n* = 537) chromosomes were filtered out. Out of 865,859 probes assessed by the EPIC Beadchip, 715,986 autosomal CpGs remained for the analyses. DNAm levels are frequently expressed using  $\beta$ -values, calculated from intensity measurements for unmethylated (U) or methylated (M) CpGs as  $\beta = M / (U + M + 100)$ . These values are easily interpreted, with a  $\beta$ -value of 0 indicating no methylation of the CpG, while a  $\beta$ -value of 1 indicates complete methylation of the CpG. In our

analyses, we used M-values, calculated as  $M = \log_2[\beta / (1 - \beta)]$ , as better performance in detection rate and true positive rate has been shown for highly methylated and unmethylated CpGs compared with  $\beta$ -values [23]. Before calculating M-values,  $\beta$ -values equal to zero were imputed by the minimum  $\beta$ -value (there were no missing data and no  $\beta$ -value equal to one).

### Statistical analyses

Weighted linear mixed-effects models were used to assess the associations between alcohol consumption (in grams per day) and each CpG (*n* = 715,986), with DNAm as the dependent variable and alcohol consumption as the independent variable. The batch was set as a random effect. Models were adjusted for age and parity at recruitment, body mass index in 1992 (< 18.5; [18.5–25]; ≥ 25 kg/m<sup>2</sup>), smoking status in 1992 (never smokers; current smokers; former smokers), menopausal status in 1992, and cellular composition (CD8+ T cells, CD4+ T cells, natural killer cells, B cells, monocytes, and neutrophils) estimated with the estimateCellCounts2 function from FlowSorted.Blood.EPIC package [24]. Additionally, models were weighted to account for the case-cohort study design (specifically, weights of 1.77 [ $w = 1 / (926 / 1,640)$ ] were assigned to cases and 32.70 [ $w = 1 / (612 / (21,653 - 1,640))$ ] to non-cases), as recommended in a guide to reframing case-cohort studies to improve usability and flexibility [25]. This guide recommends, when assessing research questions not related to the outcome, including both cases and non-cases to increase statistical power. To ensure that the measured association is “case-independent”, weights should be applied. Specifically, selected cases would be representative of all cases reported at the time the case-cohort was designed, with a weight of 1/(number of cases included in the study/total number of cases in the entire cohort). Non-cases would be representative of all individuals who had not yet developed breast cancer at the time the case-cohort was designed, with a weight of 1/(number of non-cases included in the study/total number of non-cases in the entire cohort). *P* values were corrected for multiple testing using the false discovery rate (FDR) method from Benjamini-Hochberg [26] and the family wise error rate (FWER) with Bonferroni (0.05/715,986) [27]. CpGs with a FDR < 0.05 were defined as differentially methylated positions (DMP).

To assess if the education level (< High school diploma; High school diploma to 2-y university; > 2-y university diploma) was a confounding factor, we compared the results from the models with and without adjustment for education. After excluding missing data for education (*n* = 52), 1,486 participants (902 cases and 584 non-cases) were included in this sensitivity analysis.

To assess whether the inclusion of cancers influenced the findings, we conducted sensitivity analyses on non-cases only (no weight included in the model) and on all participants after removing incident breast cancer cases occurring within the first five years after blood collection (weights of 1.91 [ $w = 1/(574/1,096)$ ] were assigned to cases and 32.7 to non-cases).

We also investigated associations with alcohol in categories (non-drinkers; moderate drinkers 0.1–19.9 g/day; heavy drinkers  $\geq 20$  g/day) and alcohol type in ml/day (wine/champagne, cider/beer or spirits).

To identify differentially methylated regions (DMR), we applied *dmrff* function on the EWAS summary statistics ( $n = 715,986$  CpGs) [28]. A candidate region is composed of at least two CpGs with EWAS  $P$  value  $< 0.05$ , estimated associations in the same direction, and that are within 1,000 bp of each other. A region was considered differentially methylated if it contained more than one CpG and Bonferroni adjusted  $P$  value  $< 0.05$ . All analyses were conducted using R version 4.2.3.

### Pathway analysis

Pathway analysis was performed using the *GOMeth* function from the *missMethyl* package [29]. *IlluminaHumanMethylationEPICanno.ilm10b4.hg19* annotation package was used to map each CpG to genes. A gene is considered differentially methylated if there is at least one DMP identified by the EWAS on alcohol consumption. The *GOMeth* method performs over-representation analysis that takes into account the probe number per gene and the CpGs that are annotated to multiple genes [30]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) collection pathways were used to identify any pathway enriched for CpGs that achieved significance at the epigenome-wide threshold ( $FDR < 0.05$ ).  $FDR$  was used to correct for multiple testing.

### Validation of results

#### Replication analysis

We selected the CpGs with a  $FDR < 0.05$  and conducted a replication analysis in an independent case–control study nested in the E3N cohort. Details about the case–control study have been published previously [31]. A total of 354 women were included in the analysis, 177 breast cancer cases and 177 controls. There was no overlap with participants in the case-cohort study. Quality control and DNAm data preprocessing were done identically as for the case-cohort study. The same statistical analysis was applied only on the subset of CpGs that passed epigenome-wide significance in the case-cohort study. Models were weighted to consider the case–control study design (weights calculated for cases as 0.16 and for non-cases as 1.84), as described previously [32].

### Comparison with literature results

In 2022, we conducted a systematic review of observational studies, published before December 2020, assessing the association between alcohol consumption and DNAm in blood [4]. Briefly, eleven cross-sectional studies with a genome-wide approach were included. The EWAS identified between 0 and 5,458 DMPs, with a total of 6,873 unique DMPs. An additional cross-sectional EWAS study was published in 2021 [16]. We combined the results and obtained a list of 8,856 DMPs, including 1,268 DMPs identified by at least two studies. After the exclusion of CpG specific to Illumina 450K Beadchip ( $n = 445$ ), the reference list contained 8,411 CpGs, with 1,197 CpGs identified by at least two studies. Regarding DMR, four studies identified between 33 and 221 DMRs associated with alcohol consumption. Available information on DMRs was reported in our review paper, but for one study they only reported the region with the smallest  $P$  value out of the 40 DMRs identified. We combined the results with the 536 DMRs identified by the study published in 2021, resulting in 869 unique DMRs with 63 overlaps.

We investigated whether cell composition was a confounding factor for the overlapping CpG sites between our EWAS and the literature review results. For the selected CpGs, we compared the results of the EWAS on alcohol consumption (in g/day) with and without adjustment for cell composition.

## Results

### Characteristics of the study participants

The characteristics of the 1,538 participants based on their alcohol consumption (non-drinkers, moderate drinkers 0.1–19.9 g/day, and heavy drinkers  $\geq 20$  g/day) are presented in Table 1. Overall median alcohol consumption was 8.3 g/day, with 157 non-drinkers (10.2%), 1,030 moderate drinkers (67.0%), and 351 heavy drinkers (22.8%). Consumption was self-reported in a questionnaire completed at a median time of four years prior to blood collection. Participants had a median age at recruitment (1990) of 48.7 years, and their median follow-up duration was 17.8 years. Compared with non-drinkers, moderate drinkers, and heavy drinkers were less overweight (respectively, 17.7% and 18.2% vs 21.0%), more likely to be current smokers (respectively, 12.2% and 19.7% vs 3.2%) or former smokers (respectively, 34.0% and 37.9% vs 21.7%), less likely to be in postmenopause (respectively, 45.4% and 49.9% vs 54.8%), more likely to be nulliparous (respectively, 13.9% and 14.0% vs 8.3%), and had a lower education level (respectively, 57.6% and 54.1% vs 61.2%,  $\leq 2$ -year university diploma).

**Table 1** Characteristics of the case-cohort study by alcohol consumption (n = 1,538)

Characteristics <sup>a</sup>	Overall	Alcohol consumption		
		0 g/day	0.1–19.9 g/day	≥ 20 g/day
n	1538	157	1030	351
Age in 1990 <sup>b</sup> (years)	48.71 [43.86, 53.75]	49.80 [44.67, 54.91]	48.13 [43.73, 53.26]	49.68 [44.36, 54.20]
Follow-up duration (years)	17.81 [12.60, 24.29]	18.98 [13.78, 24.22]	17.94 [12.47, 24.31]	17.16 [12.39, 24.14]
Time between self-reported alcohol consumption and blood collection (years)	4.00 [3.18, 4.92]	4.36 [3.60, 5.18]	3.99 [3.17, 4.91]	3.89 [3.16, 4.82]
Alcohol consumption in 1993 (g/day)	8.29 [2.12, 18.56]	0.00 [0.00, 0.00]	6.50 [2.44, 11.29]	29.86 [23.87, 38.78]
Wine/champagne (ml/day)	53.95 [7.39, 148.92]	0.00 [0.00, 0.00]	39.41 [9.85, 86.43]	260.00 [214.29, 353.82]
Cider/beer (ml/day)	0.00 [0.00, 23.57]	0.00 [0.00, 0.00]	0.00 [0.00, 21.35]	8.21 [0.00, 60.34]
Spirits (ml/day)	8.34 [1.31, 20.69]	0.00 [0.00, 0.00]	8.34 [2.63, 18.39]	20.95 [7.95, 46.11]
Body mass index in 1992 (kg/m <sup>2</sup> )	22.15 [20.70, 24.16]	21.72 [20.31, 24.46]	22.06 [20.81, 24.03]	22.31 [20.68, 24.34]
< 18.5	51 (3.3)	8 (5.1)	33 (3.2)	10 (2.8)
[18.5–25[	1208 (78.5)	116 (73.9)	815 (79.1)	277 (78.9)
≥ 25	279 (18.1)	33 (21.0)	182 (17.7)	64 (18.2)
Smoking status in 1992				
Never	821 (53.4)	118 (75.2)	554 (53.8)	149 (42.5)
Current	200 (13.0)	5 (3.2)	126 (12.2)	69 (19.7)
Former	517 (33.6)	34 (21.7)	350 (34.0)	133 (37.9)
Birth generation				
< 1935	286 (18.6)	38 (24.2)	174 (16.9)	74 (21.1)
[1935–1940[	329 (21.4)	38 (24.2)	210 (20.4)	81 (23.1)
[1940–1945[	410 (26.7)	36 (22.9)	284 (27.6)	90 (25.6)
≥ 1945	513 (33.4)	45 (28.7)	362 (35.1)	106 (30.2)
Postmenopausal in 1992	729 (47.4)	86 (54.8)	468 (45.4)	175 (49.9)
Nulliparous <sup>d</sup>	205 (13.3)	13 (8.3)	143 (13.9)	49 (14.0)
Education level				
< High school diploma	155 (10.1)	15 (9.6)	114 (11.1)	26 (7.4)
High school diploma to 2-y university	724 (47.1)	81 (51.6)	479 (46.5)	164 (46.7)
> 2-y university diploma	607 (39.5)	54 (34.4)	402 (39.0)	151 (43.0)
Missing values	52 (3.4)	7 (4.5)	35 (3.4)	10 (2.8)
Breast cancer cases	926 (60.2)	92 (58.6)	608 (59.0)	226 (64.4)
Non-cases	612 (39.8)	65 (41.4)	422 (41.0)	125 (35.6)

<sup>a</sup> Median [interquartile range] for continuous variables, n (%) for categorical variables

<sup>b</sup> Age and parity were measured at recruitment in 1990

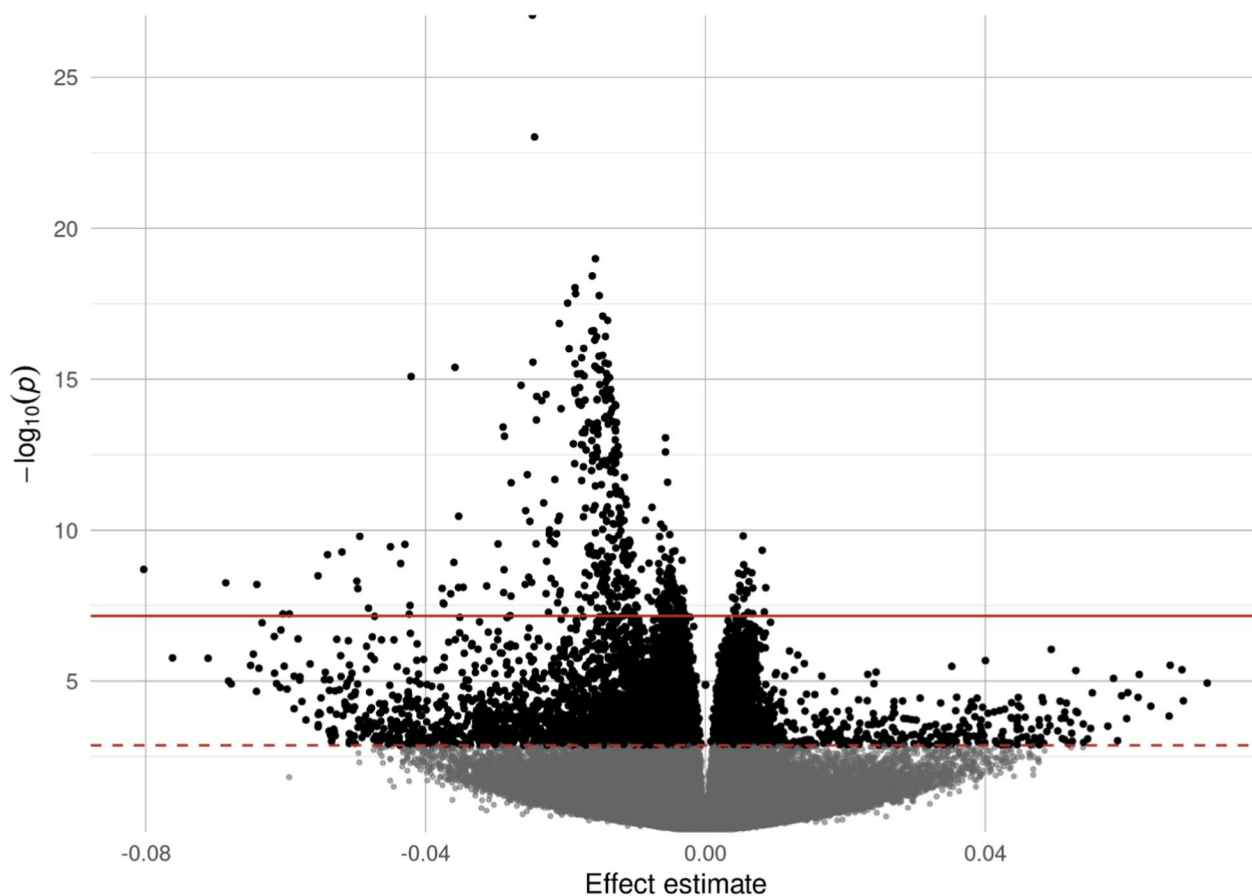
**EWAS of alcohol consumption**

Of the 715,986 CpGs analyzed, 453 CpGs were associated with alcohol consumption (in g/day) at Bonferroni threshold ( $p < 0.05/715,986$ ) and 19,255 CpGs at  $FDR < 0.05$ . The 19,255 CpGs mapped to 9,039 unique genes. As alcohol consumption increased, 12,433 CpGs were less methylated, and 6,822 CpGs had higher methylation levels. The results are presented graphically using a Volcano plot (Fig. 1) and a Manhattan plot (Fig. 2). The top twenty CpGs with the lowest *P* values are listed in Table 2, while the full list of 19,255 CpGs is reported in supplementary material (Additional File 2: Table S1). For all the top twenty CpGs,

increased alcohol consumption was associated with hypomethylation.

Using the 1,486 participants with complete data on education level (<High school diploma; High school diploma to 2-y university; >2-y university diploma), we compared the EWAS results with and without education adjustment and similar results were observed (Additional File 1: Fig. S2).

The sensitivity analysis on non-cases only (n=612) and on all participants after removing incident cancers occurring within the first five years after blood collection (n=1,186) showed similar results to those of the EWAS performed on all participants for each CpG (Additional



**Fig. 1** Volcano plot of the EWAS. The Volcano plot depicts effect estimate (mean change in methylation level, M-values, for each CpG site for one gram per day of alcohol consumption increase) on the x-axis and  $-\log_{10}(P)$  on the y-axis. Single CpG sites are depicted as dots. The plain line corresponds to the Bonferroni threshold ( $P$  values  $< 6.98 \times 10^{-8}$ ), and the dashed line corresponds to the False Discovery Rate threshold ( $FDR < 0.05$ )

File 1: Figure S3, S4). This supports that the cancer did not impact the findings.

When using alcohol consumption in categories (non-drinkers; moderate drinkers 0.1–19.9 g/day; heavy drinkers  $\geq 20$  g/day), 13,101 CpGs and 13,172 CpGs were associated with heavy drinking and moderate drinking, respectively, at  $FDR < 0.05$ . The overlap between the EWAS results (Fig. 3), with alcohol consumption as a continuous or categorical variable, is relatively small. There were 3,512 common CpGs between continuous alcohol consumption and heavy drinking, and 438 common CpGs between continuous alcohol consumption and moderate drinking.

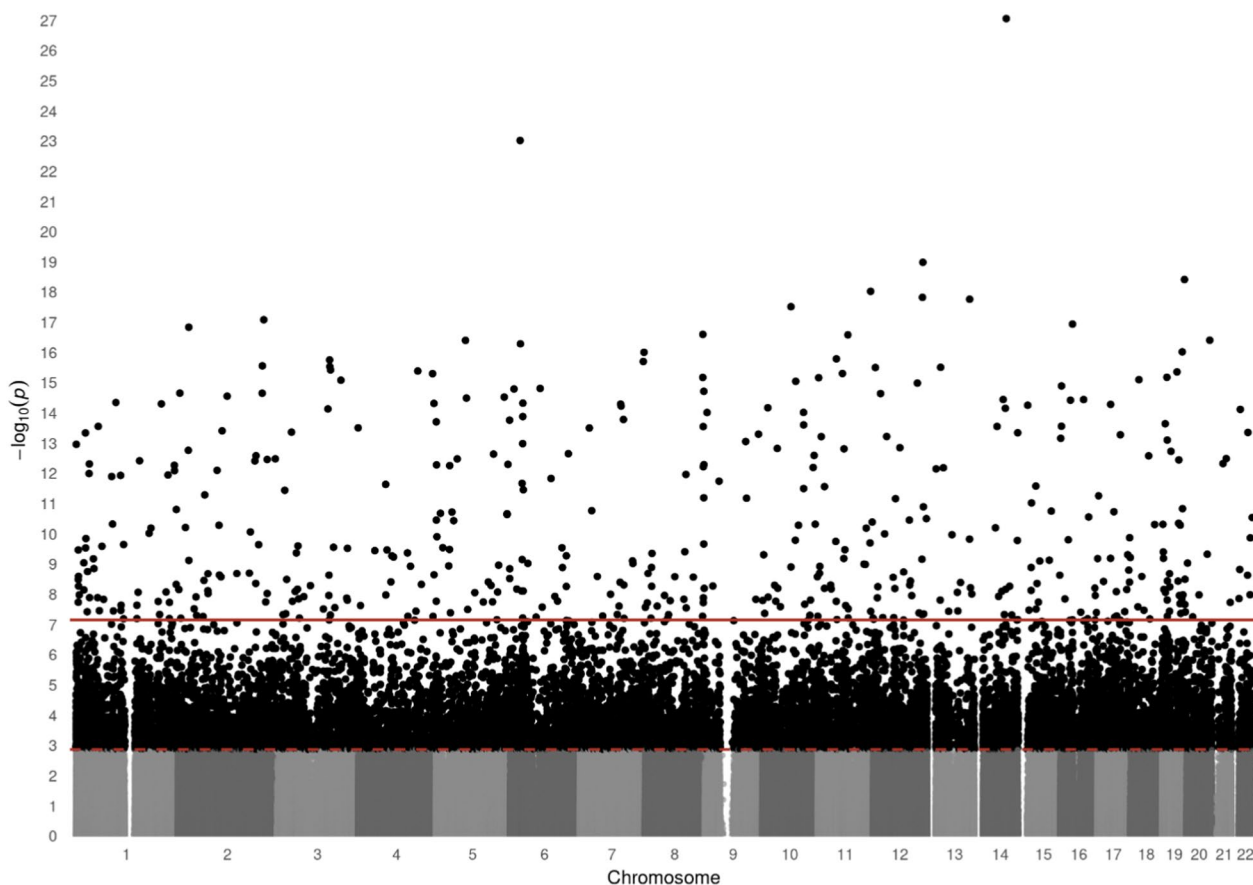
When assessing the associations by alcohol subtype, 19,632 CpGs, 13,131 CpGs, and 13,387 CpGs were associated with wine/champagne, cider/beer, and spirits, respectively, at  $FDR < 0.05$ . The overlaps between the DMPs reported in the different EWAS are in Fig. 4. There were 12,884 common DMPs between total alcohol consumption and wine/champagne, 1,045 common DMPs

between total alcohol consumption and cider/beer, and 2,108 common DMPs between total alcohol consumption and spirits. Thirty CpG sites were identified in the four EWAS and are reported in Table 3. Additional information ( $FDR$  for each EWAS, estimate and  $p$  value from the replication analysis) is provided in supplementary material (Additional File 1: Table S2).

Among 102,980 candidate regions associated with alcohol consumption in our EWAS, 90 were DMRs. Table 4 contains the 90 DMRs and supplementary material (Additional File 2: Table S3) details information on the CpGs included in the different DMR.

#### Pathway analysis

The 9,039 genes identified in the EWAS (from the 19,255 DMPs) were significantly over-represented in 26 pathways (Table 5). These pathways highlight genes involved in cancer, the nervous system and aging. Results for pathway analyses (alcohol in continuous and in categories, alcohol subtypes, DMPs replicated in the case-control



**Fig. 2** Manhattan plot of the EWAS on alcohol consumption. The Manhattan plot depicts the position on chromosome on the x-axis and  $-\log_{10}(P)$  value on the y-axis. Single CpG sites are depicted as dots. The plain line corresponds to the Bonferroni threshold ( $P$  values  $< 6.98 \times 10^{-8}$ ), and the dashed line corresponds to the False Discovery Rate threshold ( $FDR < 0.05$ )

study) are in supplementary material (Additional File 2: Table S4).

### Validation of the results

#### Replication analysis

The characteristics of the 354 women included in the replication case–control study are presented by alcohol consumption in supplementary material (Additional File 1: Table S5). Among the 19,255 CpGs identified in our EWAS, 1,528 CpGs were replicated in the case–control study ( $P$  value  $< 0.05$  with same direction of association). The 1,528 CpGs mapped to 1,163 unique genes. The top 20 replicated CpGs are listed in Table 6 and the full list is in supplementary material (Additional File 2: Table S6).

#### Comparison with literature results

Of 8,411 DMPs identified in the literature to be associated with alcohol consumption, 591 CpGs were also identified in our EWAS, and 85 were replicated in the case–control study. When we restricted to the 1,197 CpGs identified in at least two studies from the literature

review, 146 CpGs overlapped with our EWAS, including 23 replicated CpGs. Among the 146 CpGs, one was identified by eight studies (cg06690548), six by six studies (cg00422488, cg02578470, cg02583484, cg05512483, cg12825509, and cg18917643), 12 by five studies, 35 by four studies, and 92 by three studies. The list of CpGs ( $n=19$ ) identified by at least five studies is presented in Table 7, while the full list of the 8,411 DMPs identified in the literature is reported in supplementary material (Additional File 2: Table S7) with indication of the studies that identified each specific DMP.

To assess if the cellular composition acts as a confounding factor on alcohol-associated CpG sites, we compared the results with and without adjustment for cell composition for the 591 CpGs that overlapped with the previous findings in the literature and similar results were observed (Additional File 1: Fig. S5).

Six of the 90 DMRs identified in our analysis overlapped with the literature results. These DMRs are shown in Table 8 and are located in *CPT1A*, *UBE2E1*, *DUSP15/TLL9*, *FKBP11*, *CCDC106* and *MATR3* genes.

**Table 2** Top 20 CpG sites identified in the EWAS on alcohol consumption

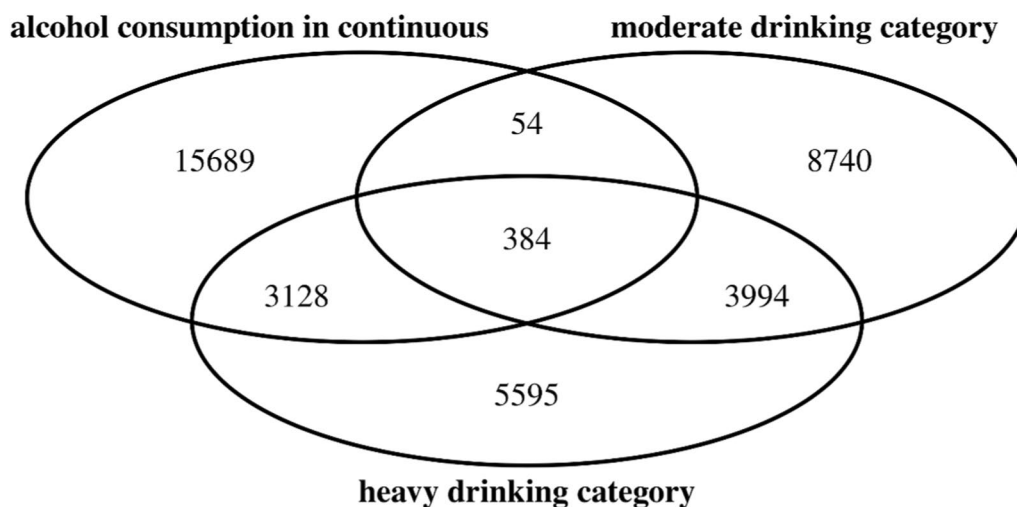
CpG	Chr	CpG island region <sup>a</sup>	Gene <sup>b</sup>	Gene region <sup>c</sup>	Estimate	Standard error	P values	FDR
cg13560408	14	Island	<i>C14orf174; TMED8</i>	TSS1500; Body	-0.02475	0.00222	8.72E-28	6.24E-22
cg17037491	6	S_Shore	<i>LRRC16A</i>	Body	-0.02444	0.00239	9.50E-24	3.40E-18
cg00483526	12	S_Shore	<i>CLIP1</i>	TSS1500	-0.01575	0.00171	1.01E-19	2.42E-14
cg00813890	19	N_Shore	<i>TTYH1</i>	Body	-0.01619	0.00179	3.77E-19	6.76E-14
cg14482569	11	S_Shore	<i>ZBTB44</i>	TSS1500	-0.01866	0.00208	9.25E-19	1.32E-13
cg13964378	13	Island	<i>BIVM; KDELC1</i>	TSS1500; 1stExon	-0.01520	0.00171	1.69E-18	1.73E-13
cg26285350	12	OpenSea	<i>KDM2B</i>	Body	-0.01858	0.00209	1.47E-18	1.73E-13
cg00128195	10	N_Shore	<i>TET1</i>	TSS200	-0.01972	0.00223	2.99E-18	2.67E-13
cg21656520	2	Island	<i>RPE</i>	TSS200	-0.01470	0.00169	8.05E-18	6.40E-13
cg01486387	16	Island	<i>KCTD13</i>	TSS200	-0.01398	0.00161	1.12E-17	8.03E-13
cg03431514	2	Island	<i>ZNF513</i>	TSS200	-0.02090	0.00242	1.42E-17	9.24E-13
cg17718679	11	Island	<i>KCNE3</i>	TSS200	-0.01628	0.00190	2.54E-17	1.40E-12
cg11268280	8	N_Shelf			-0.01599	0.00187	2.45E-17	1.40E-12
cg18798289	5	S_Shore			-0.01557	0.00183	3.87E-17	1.85E-12
cg03255221	20	Island	<i>SLMO2; SLMO2</i>	5'UTR; 1stExon	-0.01430	0.00168	3.83E-17	1.85E-12
cg15945016	6	Island	<i>HIST1H4A; HIST1H3A</i>	TSS1500; TSS200	-0.01583	0.00187	5.06E-17	2.26E-12
cg12412075	19	Island	<i>LIN7B</i>	Body	-0.01741	0.00207	9.42E-17	3.87E-12
cg08492145	7	S_Shore	<i>PTPRN2</i>	Body	-0.01950	0.00232	9.74E-17	3.87E-12
cg03034818	11	S_Shore	<i>PEX16</i>	TSS1500	-0.01472	0.00176	1.61E-16	6.05E-12
cg10957984	3	S_Shore	<i>RAB43; RAB43; RAB43; RAB43; ISY1-RAB43</i>	5'UTR; TSS1500; TSS200; 1stExon; Body	-0.01525	0.00183	1.75E-16	6.25E-12

Chr chromosome, CpG cytosine-phosphate-guanine

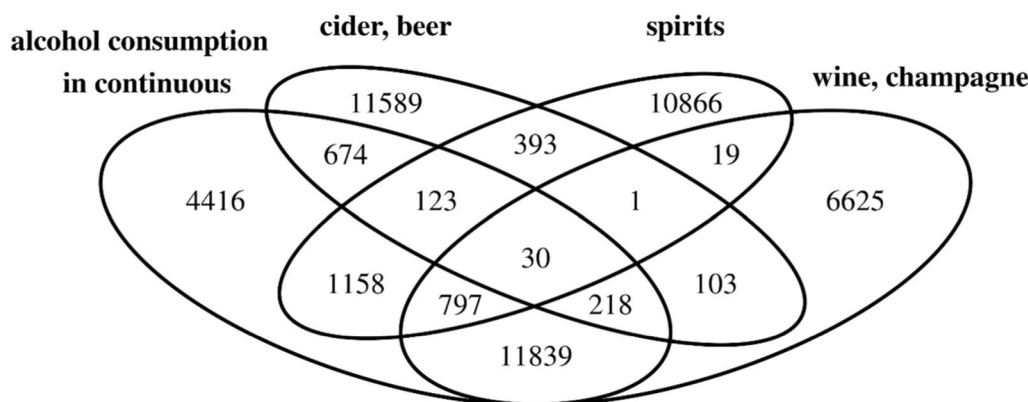
<sup>a</sup> CpG islands are regions of approximately one thousand base pairs long, enriched in CpG dinucleotides. In humans, 72% of gene promoters contain CpG islands. CpG can then be localized toward these islands: Island = inside the island; Shelf = 2–4 kb from the island; Shore = 0–2 kb from the island; N = upstream (5') from the island; S = downstream (3') from the island; OpenSea ≥ 4 kb from the islands

<sup>b</sup> Multiple lists of the same gene name indicate splicing variants

<sup>c</sup> Gene is divided into several regions: TSS200 = 0–200 bases upstream of the transcription start site (TSS); TSS1500 = 200–1500 bases upstream of the TSS; 5'UTR = in the 5' untranslated region, between the TSS and the ATG start site; Body = between the ATG and the stop codon, regardless of the presence of introns, exons, TSS or promoters; 3'UTR = between the stop codon and the poly A signal



**Fig. 3** Venn diagram of EWAS results, continuous and categorical variables for alcohol consumption. For alcohol in continuous, 19,255 CpG sites were associated (FDR < 0.05). For alcohol in categories, 13,172 CpG sites were associated with moderate drinking (0.1–19.9 g/day vs non-drinkers) and 13,101 CpG sites were associated with heavy drinking (≥ 20 g/day vs non-drinkers)



**Fig. 4 Venn diagram of EWAS results by alcohol subtype.** For alcohol in continuous, 19,255 CpG sites were associated (FDR < 0.05). In total, 19,632 CpGs, 13,131 CpGs, and 13,387 CpGs were associated with wine/champagne, cider/beer, and spirits, respectively, at FDR < 0.05

### Discussion

Our study identified 19,255 CpGs associated with alcohol consumption (12,433 hypomethylated and 6,822 hypermethylated with increased alcohol consumption) and the over-representation analysis of the 9,039 associated genes highlighted enrichment of genes involved in cancer, the nervous system and aging. The associations were also investigated with alcohol in categories (non-drinkers; moderate drinkers 0.1–19.9 g/day; heavy drinkers ≥ 20 g/day) and alcohol type in ml/day (wine/champagne, cider/beer or spirits). Overlaps in EWAS results were greater between the continuous variable and the subset of heavy drinkers within the categorical variable, and between total alcohol consumption and wine/champagne consumption. Of the 19,255 CpGs, 1,528 were replicated in the case–control study (1,260 hypomethylated and 268 hypermethylated). At the region level, we identified 90 DMRs (88 hypomethylated and two hypermethylated).

Among the 1,528 replicated CpGs, 85 were also found in other studies included in our systematic review of the literature. Of the 90 DMRs, six overlapped with DMRs reported in the literature.

Several CpGs have been associated with alcohol consumption in the literature, and 19 CpGs were consistently identified in at least five studies (including our EWAS). Differences between studies may be explained by differences between countries in the types of alcohol consumed, the quantity of alcohol consumed (E3N women tended to consume less alcohol than in other studies), the difference in the women/men proportion, and the consideration of temporality between alcohol consumption and DNAm measurements (the studies in the review were cross-sectional). Although we still lack longitudinal analyses, several CpGs have been identified in multiple EWAS, including ours that has considered temporality (alcohol consumption was self-reported

with a median time of four years prior to blood collection), and have been replicated in a longitudinal analysis conducted by Dugué et al. [5]. Higher alcohol consumption was consistently associated with the hypomethylation of DNAm sites located in the solute carrier family 7 member 11 *SLC7A11* (cg06690548), the heterogeneous nuclear ribonucleoprotein A1 *HNRNPA1* (cg02583484), the acidic nuclear phosphoprotein 32 family member B *ANP32B* (cg00422488), the phosphoglycerate dehydrogenase *PHGDH* (cg16246545), the minichromosome maintenance complex component 2 *MCM2* (cg02578470), the carnitine palmitoyltransferase 1A *CPT1A* (cg17058475 and cg09737197), the ETS variant transcription factor 6 *ETV6* (cg18763536), the small nucleolar RNA C/D box 30 *SNORD30* (cg05512483); the LSM14A mRNA processing body assembly factor *LSM14A* (cg03339609); the cell division cycle 7 *CDC7* (cg03260624); the LIM domain binding 2 *LDB2* (cg04460609); the leucine rich repeat containing 8 VRAC subunit D *LRRC8D* (cg06983052) genes and another location outside known genes (cg14482712). These genes are involved in various cellular processes, such as cell growth and/or death for *SLC7A11* [33], *MCM2* [34], and *ANP32B* [35]; and amino acid transport for *SLC7A11* (cystine/glutamate antiporter [36]). Additionally, they play roles in metabolism, including amino acid metabolism for *PHGDH* [37] and lipid metabolism for *CPT1A* [38]. These genes are also involved in genetic information processing, such as transcription and messenger RNA biogenesis for *ANP32B* [39] and *HNRNPA1* [40], replication and repair for *MCM2* [41], and transcription factor for *ETV6* [42]. Moreover, they have implications in various human diseases, including neurodegenerative disease (*SLC7A11* [43, 44], *PHGDH* [45], *HNRNPA1* [46–48]), endocrine and metabolic disease (*CPT1A* [49, 50]), infectious diseases (*SLC7A11* [51], *HNRNPA1* [52], *ANP32B* [53],

**Table 3** Overlap of CpG sites (n = 30) associated with alcohol consumption in EWAS on total alcohol consumption and by alcohol subtype

CpG	Chr	CpG island region <sup>a</sup>	Gene <sup>b</sup>	Gene region <sup>c</sup>	Estimate alcohol	Estimate wine/ champagne	Estimate cider/ beer	Estimate spirits
cg10068417	13	N_Shore			-0.00284	-0.00034	0.00044	-0.00162
cg03947478	16	OpenSea	<i>DHX38</i>	Body	-0.00311	-0.00027	-0.00040	-0.00142
cg17900884	1	Island	<i>ICMT</i>	TSS200	-0.05540	-0.00442	-0.00745	-0.02031
cg21383891	17	Island	<i>SLFN5</i>	TSS1500	-0.02341	-0.00158	-0.00572	-0.00701
cg08096725	12	OpenSea			-0.00402	-0.00031	-0.00052	-0.00227
cg26022608	1	OpenSea	<i>SORT1;SORT1</i>	Body;5'UTR	-0.00436	-0.00033	-0.00067	-0.00209
cg15193215	6	OpenSea	<i>GMDS</i>	Body	-0.00475	-0.00039	-0.00068	-0.00217
cg00877958	16	S_Shelf			0.00563	0.00044	0.00062	0.00269
cg18003513	11	OpenSea	<i>ME3</i>	Body	-0.00401	-0.00033	-0.00050	-0.00181
cg20477849	12	OpenSea			0.00548	0.00041	0.00077	0.00285
cg17988326	1	OpenSea	<i>RIMS3</i>	3'UTR	0.00262	0.00019	0.00038	0.00137
cg12072973	1	Island	<i>PLEKHO1</i>	TSS1500	-0.00623	-0.00048	-0.00078	-0.00297
cg12309550	10	N_Shelf	<i>ZNF248</i>	Body	-0.02577	-0.00149	-0.00617	-0.01239
<b>cg06690548</b>	<b>4</b>	<b>OpenSea</b>	<b><i>SLC7A11</i></b>	<b>Body</b>	<b>-0.00558</b>	<b>-0.00042</b>	<b>-0.00070</b>	<b>-0.00249</b>
cg10852320	5	S_Shelf			-0.00353	-0.00028	-0.00047	-0.00162
cg01378578	1	Island	<i>PFKFB2</i>	5'UTR	-0.00565	-0.00040	-0.00078	-0.00332
cg07366967	14	OpenSea	<i>EXD2</i>	TSS1500	-0.00399	-0.00031	-0.00055	-0.00194
<b>cg15835363</b>	<b>15</b>	<b>OpenSea</b>	<b><i>STARD9</i></b>	<b>Body</b>	<b>-0.00541</b>	<b>-0.00043</b>	<b>-0.00052</b>	<b>-0.00332</b>
cg22862466	1	S_Shelf	<i>IBA57</i>	Body	-0.00319	-0.00024	-0.00042	-0.00159
cg24993742	2	OpenSea	<i>NCK2</i>	5'UTR	-0.01504	-0.00114	-0.00209	-0.00823
cg03976151	21	OpenSea			0.00710	0.00056	0.00098	0.00363
cg15072822	6	S_Shore			-0.00446	-0.00034	-0.00056	-0.00212
cg02513673	9	Island			-0.00238	-0.00018	-0.00029	-0.00109
cg10564845	10	OpenSea	<i>DOCK1</i>	Body	-0.01155	-0.00085	-0.00141	-0.00675
cg19856273	1	N_Shore	<i>C1orf190; POMGNT1</i>	TSS1500; Body	-0.00584	-0.00039	-0.00097	-0.00418
cg13615100	15	OpenSea	<i>TRPM1</i>	5'UTR	-0.00668	-0.00053	-0.00095	-0.00294
<b>cg02433086</b>	<b>10</b>	<b>N_Shelf</b>	<b><i>PIK3AP1</i></b>	<b>Body</b>	<b>-0.00297</b>	<b>-0.00022</b>	<b>-0.00041</b>	<b>-0.00190</b>
cg00099664	12	OpenSea			0.00724	0.00053	0.00080	0.00502
cg17907457	1	OpenSea	<i>HS2ST1</i>	Body	-0.00379	-0.00029	-0.00053	-0.00208
cg24532514	14	OpenSea			0.00542	0.00040	0.00076	0.00304

In bold, the three CpGs were also identified in the replication analysis

Chr chromosome, CpG cytosine-phosphate-guanine

<sup>a</sup> CpG islands are regions of approximately one thousand base pairs long, enriched in CpG dinucleotides. In humans, 72% of gene promoters contain CpG islands. CpG can then be localized toward these islands: Island = inside the island; Shelf = 2–4 kb from the island; Shore = 0–2 kb from the island; N = upstream (5') from the island; S = downstream (3') from the island; OpenSea ≥ 4 kb from the islands

<sup>b</sup> Multiple lists of the same gene name indicate splicing variants

<sup>c</sup> Gene is divided into several regions: TSS200 = 0–200 bases upstream of the transcription start site (TSS); TSS1500 = 200–1500 bases upstream of the TSS; 5'UTR = in the 5' untranslated region, between the TSS and the ATG start site; Body = between the ATG and the stop codon, regardless of the presence of introns, exons, TSS or promoters; 3'UTR = between the stop codon and the poly A signal

*MCM2* [54], *CPT1A* [55]), and cancer (*SLC7A11* [56–58], *HNRNPA1* [59, 60], *ANP32B* [61], *PHGDH* [62, 63], *MCM2* [64], *CPT1A* [65], *ETV6* [66–68]). Overall, alterations in these pathways could contribute to immune system dysregulation and alcohol-related diseases. Chronic ethanol exposure disrupts multiple metabolic pathways, including methionine metabolism and glutathione

synthesis [69], leading to profound effects on DNAm and immune system regulation. Exposure to ethanol disrupts methionine metabolism, resulting in reduced production of S-adenosylmethionine (SAME), a methyl group donor for DNA and histone methylation processes [70]. This dysregulated methylation pattern contributes to altered gene expression and cellular processes, such as

**Table 4** Differentially methylated regions associated with alcohol consumption

Chr	Start	End	n	Estimate	Standard error	P values	Bonferroni	Gene
2	219,262,936	219,263,149	2	-0.00953	0.00127	4.98E-14	4.84E-08	<i>CTDSP1</i>
6	32,162,755	32,163,348	4	-0.00612	0.00084	2.52E-13	2.45E-07	<i>GPSM3;NOTCH4</i>
11	68,607,257	68,608,225	6	-0.00713	0.00100	1.10E-12	1.07E-06	<i>CPT1A</i>
7	37,472,663	37,473,895	4	-0.00546	0.00080	6.69E-12	6.50E-06	<i>ELMO1</i>
19	48,673,765	48,673,767	2	-0.00837	0.00129	8.25E-11	8.02E-05	<i>LIG1;C19orf68</i>
3	172,468,378	172,468,381	2	-0.00875	0.00136	1.29E-10	1.26E-04	<i>ECT2</i>
14	58,711,098	58,711,122	2	-0.00803	0.00127	2.60E-10	2.53E-04	<i>PSMA3</i>
9	34,457,129	34,458,458	5	-0.00609	0.00097	3.12E-10	3.03E-04	<i>C9orf25;DNAI1</i>
1	161,719,820	161,719,982	2	-0.00777	0.00124	3.44E-10	3.34E-04	<i>DUSP12</i>
1	24,127,177	24,127,242	3	-0.00702	0.00112	3.71E-10	3.61E-04	<i>GALE</i>
9	139,875,200	139,875,581	2	0.00575	0.00092	4.94E-10	4.80E-04	<i>PTGDS</i>
19	56,159,710	56,159,817	2	-0.00896	0.00145	6.08E-10	5.91E-04	<i>CCDC106</i>
1	93,300,695	93,302,167	6	-0.00575	0.00093	6.73E-10	6.54E-04	<i>SNORD21;RPL5;FAM69A</i>
15	91,162,876	91,163,732	2	-0.00658	0.00107	7.39E-10	7.18E-04	<i>CRTC3;CRTC3-AS1</i>
3	187,685,951	187,686,918	2	-0.00480	0.00078	7.42E-10	7.21E-04	
15	52,860,841	52,861,038	2	-0.00617	0.00101	9.06E-10	8.80E-04	<i>ARPP19</i>
11	8,008,592	8,008,752	4	-0.00448	0.00073	1.03E-09	1.00E-03	<i>EIF3F</i>
8	96,084,821	96,085,994	7	-0.00611	0.00100	1.08E-09	1.05E-03	<i>MIR3150B</i>
2	240,230,892	240,231,055	2	-0.00920	0.00151	1.09E-09	1.05E-03	<i>HDAC4</i>
12	56,508,205	56,510,114	7	-0.00481	0.00079	1.18E-09	1.15E-03	<i>RPL41</i>
13	73,355,953	73,355,978	2	-0.00783	0.00129	1.37E-09	1.33E-03	<i>DIS3;PIBF1</i>
9	130,689,862	130,692,341	8	-0.00465	0.00077	1.73E-09	1.68E-03	<i>PIP5KL1</i>
1	179,555,658	179,555,999	3	0.00758	0.00127	2.31E-09	2.25E-03	
4	25,862,037	25,863,659	6	-0.00457	0.00077	2.51E-09	2.44E-03	<i>SEL1L3</i>
20	30,458,515	30,459,158	7	-0.00423	0.00071	2.52E-09	2.45E-03	<i>DUSP15;TTL9</i>
6	133,563,764	133,564,066	4	-0.00531	0.00090	3.00E-09	2.92E-03	<i>EYA4</i>
1	11,120,546	11,120,633	2	-0.00821	0.00139	3.13E-09	3.04E-03	<i>SRM</i>
8	141,520,003	141,520,726	3	-0.00660	0.00112	3.36E-09	3.27E-03	<i>CHRAC1</i>
1	32,083,460	32,084,329	6	-0.00432	0.00073	3.73E-09	3.63E-03	<i>HCRTR1</i>
4	122,685,145	122,685,423	3	-0.00750	0.00127	3.86E-09	3.75E-03	<i>TMEM155;LOC100192379</i>
12	49,318,784	49,319,673	13	-0.00515	0.00088	4.59E-09	4.46E-03	<i>FKBP11</i>
5	138,634,374	138,634,507	2	-0.00825	0.00141	4.68E-09	4.55E-03	<i>MATR3</i>
3	23,851,358	23,851,386	2	-0.00833	0.00142	4.77E-09	4.64E-03	<i>UBE2E1</i>
17	55,938,871	55,938,940	2	-0.00806	0.00138	4.95E-09	4.81E-03	
19	44,124,409	44,124,666	2	-0.00816	0.00140	5.04E-09	4.90E-03	<i>ZNF428</i>
17	46,507,705	46,507,754	2	-0.00691	0.00118	5.54E-09	5.39E-03	<i>SKAP1</i>
3	14,990,466	14,991,006	2	-0.00813	0.00139	5.67E-09	5.51E-03	<i>NR2C2;FGD5-AS1</i>
13	103,426,299	103,427,172	9	-0.00404	0.00069	5.94E-09	5.77E-03	<i>C13orf27;TEX30</i>
12	127,544,951	127,545,259	3	-0.00638	0.00110	6.12E-09	5.94E-03	
3	180,707,907	180,708,442	2	-0.00765	0.00132	6.29E-09	6.11E-03	<i>DNAJC19</i>
1	111,747,324	111,748,053	3	-0.00525	0.00091	6.87E-09	6.68E-03	<i>DENND2D</i>
4	82,393,562	82,394,151	3	-0.00678	0.00117	7.33E-09	7.12E-03	<i>RASGEF1B</i>
9	37,593,003	37,594,221	5	-0.00549	0.00095	8.30E-09	8.07E-03	<i>TOMM5</i>
10	33,294,599	33,294,889	4	-0.00689	0.00120	8.30E-09	8.07E-03	
2	205,130,274	205,130,924	2	0.00635	0.00111	9.02E-09	8.77E-03	
3	140,770,617	140,770,683	2	-0.00819	0.00143	9.03E-09	8.78E-03	<i>SPSB4</i>
11	67,980,788	67,980,797	2	-0.00777	0.00135	9.17E-09	8.91E-03	<i>SUV420H1</i>
21	31,971,808	31,972,257	2	0.00699	0.00122	9.28E-09	9.02E-03	<i>KRTAP6-2</i>
6	36,165,010	36,165,021	2	-0.00761	0.00133	9.29E-09	9.03E-03	<i>BRPF3</i>

**Table 4** (continued)

Chr	Start	End	n	Estimate	Standard error	P values	Bonferroni	Gene
8	28,196,926	28,198,639	5	0.00340	0.00059	1.06E-08	1.03E-02	<i>PNOC</i>
17	80,561,982	80,563,734	7	-0.00453	0.00079	1.07E-08	1.04E-02	<i>FOXK2</i>
10	93,168,950	93,169,801	4	-0.00530	0.00093	1.08E-08	1.05E-02	<i>HECTD2;HECTD2-AS1;LOC100188947</i>
10	102,447,466	102,447,999	3	-0.00617	0.00108	1.13E-08	1.10E-02	
1	207,226,824	207,227,076	3	-0.00620	0.00109	1.22E-08	1.18E-02	<i>PFKFB2</i>
19	18,496,953	18,497,143	4	-0.00726	0.00129	1.65E-08	1.60E-02	<i>GDF15</i>
16	419,975	420,627	7	-0.00595	0.00105	1.66E-08	1.61E-02	<i>MRPL28</i>
16	80,574,999	80,575,548	3	-0.00595	0.00105	1.67E-08	1.62E-02	<i>DYNLRB2;LOC102724084</i>
19	55,790,889	55,790,967	2	-0.00955	0.00170	1.83E-08	1.78E-02	<i>HSPBP1</i>
2	198,061,578	198,063,653	6	-0.00404	0.00072	1.83E-08	1.78E-02	<i>ANKRD44</i>
2	232,275,105	232,275,714	2	-0.00719	0.00128	2.06E-08	2.00E-02	
1	151,738,470	151,738,555	2	-0.00760	0.00136	2.19E-08	2.12E-02	<i>OAZ3</i>
10	16,561,605	16,562,470	3	-0.00735	0.00131	2.22E-08	2.15E-2	<i>C1QL3</i>
12	128,796,292	128,796,424	2	-0.00762	0.00137	2.36E-08	2.30E-02	<i>TMEM132C</i>
5	132,948,256	132,948,515	3	-0.00592	0.00106	2.38E-08	2.31E-02	<i>FSTL4</i>
14	75,593,085	75,593,334	2	-0.00820	0.00147	2.42E-08	2.35E-02	<i>NEK9</i>
8	145,747,659	145,747,739	2	-0.00738	0.00132	2.47E-08	2.40E-02	<i>LRRC14</i>
1	229,543,254	229,543,603	3	-0.00663	0.00119	2.54E-08	2.47E-02	
16	89,048,369	89,048,433	2	-0.00780	0.00140	2.56E-08	2.48E-02	
2	114,256,394	114,256,406	2	-0.00795	0.00143	2.78E-08	2.70E-02	<i>FOXD4L1</i>
16	29,674,959	29,674,972	2	0.00788	0.00142	2.82E-08	2.74E-02	<i>SPN</i>
6	33,420,805	33,421,468	3	-0.00505	0.00091	2.97E-08	2.88E-02	<i>SYNGAP1;ZBTB9</i>
1	154,237,531	154,238,234	2	-0.00413	0.00075	3.01E-08	2.93E-02	<i>UBAP2L</i>
10	134,663,864	134,664,800	6	-0.00483	0.00087	3.11E-08	3.02E-02	<i>CFAP46</i>
11	73,471,141	73,471,500	2	-0.00613	0.00111	3.23E-08	3.14E-02	<i>RAB6A</i>
3	44,689,721	44,690,012	2	-0.00667	0.00121	3.24E-08	3.15E-02	<i>ZNF197;ZNF35</i>
12	111,619,455	111,620,199	2	-0.00481	0.00087	3.24E-08	3.15E-02	<i>CUX2</i>
20	62,693,971	62,694,000	3	-0.00869	0.00157	3.37E-08	3.28E-02	<i>TCEA2</i>
4	38,806,025	38,807,382	8	-0.00404	0.00073	3.54E-08	3.44E-02	<i>TLR1</i>
4	89,567,955	89,568,662	2	-0.00439	0.00080	3.62E-08	3.52E-02	<i>HERC3</i>
4	122,744,832	122,744,995	2	-0.00686	0.00125	3.84E-08	3.74E-02	<i>CCNA2</i>
19	45,901,567	45,901,649	2	-0.00898	0.00164	4.14E-08	4.02E-02	<i>PPP1R13L</i>
2	220,109,169	220,109,426	2	0.00643	0.00117	4.22E-08	4.10E-02	<i>GLB1L;STK16</i>
5	90,523,511	90,524,312	2	0.00663	0.00121	4.31E-08	4.19E-02	
17	29,297,941	29,298,832	5	-0.00435	0.00079	4.38E-08	4.25E-02	<i>RNF135</i>
13	107,569,698	107,570,209	4	-0.00533	0.00098	4.59E-08	4.46E-02	
7	18,067,533	18,067,535	2	-0.00811	0.00149	4.86E-08	4.72E-02	<i>PRPS1L1</i>
2	70,781,256	70,781,287	2	-0.00752	0.00138	4.97E-08	4.83E-02	<i>TGFA</i>
16	733,842	734,426	4	-0.00570	0.00105	5.02E-08	4.88E-02	<i>JMJD8</i>
7	36,192,835	36,193,115	3	-0.00597	0.00110	5.02E-08	4.88E-02	<i>EEPD1</i>
6	91,004,430	91,004,452	2	-0.00803	0.00147	5.04E-08	4.89E-02	<i>BACH2</i>

Chr chromosome, DMR differentially methylated region, n number of CpG sites in the DMR

cell growth, apoptosis, and stress response. Furthermore, ethanol-induced reduction in glutathione levels increases oxidative stress, resulting in DNA damage and further dysregulation of immune function [71]. The interaction between ethanol metabolism, DNAm, and immune system dysregulation highlights the complex mechanisms

underlying alcohol-related diseases, including neurodegenerative disorders, metabolic diseases, and cancer. A growing body of evidence supports the implication of these interactions in the pathogenesis and progression of alcohol-associated health conditions. It has been reported that PHGDH interacts with an RNA-binding protein to

**Table 5** Gene over-representation analysis using KEGG pathways collection

KEGG pathway id	Classification	Description	N	DE	P values	FDR
hsa05212	Cancer	Pancreatic cancer	73	54	1.05E-04	1.57E-02
hsa05221	Cancer	Acute myeloid leukemia	63	47	1.85E-04	1.57E-02
hsa04724	Nervous system	Glutamatergic synapse	114	80	1.89E-04	1.57E-02
hsa05220	Cancer	Chronic myeloid leukemia	74	54	2.29E-04	1.57E-02
hsa04015	Signal transduction	Rap1 signaling pathway	207	133	3.55E-04	1.57E-02
hsa05166	Infectious disease: viral	Human T-cell leukemia virus 1 infection	214	130	3.66E-04	1.57E-02
hsa05200	Cancer	Pathways in cancer	513	293	3.91E-04	1.57E-02
hsa04910	Endocrine system	Insulin signaling pathway	131	85	4.12E-04	1.57E-02
hsa04722	Nervous system	Neurotrophin signaling pathway	114	77	4.27E-04	1.57E-02
hsa04213	Aging	Longevity regulating pathway—multiple species	60	44	4.38E-04	1.57E-02
hsa05203	Cancer	Viral carcinogenesis	187	110	7.36E-04	2.23E-02
hsa04010	Signal transduction	MAPK signaling pathway	287	175	7.44E-04	2.23E-02
hsa04310	Signal transduction	Wnt signaling pathway	168	108	9.46E-04	2.61E-02
hsa04919	Endocrine system	Thyroid hormone signaling pathway	117	79	1.03E-03	2.63E-02
hsa05215	Cancer	Prostate cancer	94	64	1.11E-03	2.64E-02
hsa04750	Sensory system	Inflammatory mediator regulation of TRP channels	97	66	1.18E-03	2.64E-02
hsa04218	Cell growth and death	Cellular senescence	153	95	1.51E-03	2.95E-02
hsa04520	Cellular community	Adherens junction	92	63	1.56E-03	2.95E-02
hsa04211	Aging	Longevity regulating pathway	88	60	1.63E-03	2.95E-02
hsa05031	Substance dependence	Amphetamine addiction	66	46	1.64E-03	2.95E-02
hsa04148	Transport and catabolism	Efferocytosis	152	95	1.88E-03	3.20E-02
hsa04014	Signal transduction	Ras signaling pathway	227	137	1.96E-03	3.20E-02
hsa04726	Nervous system	Serotonergic synapse	108	67	2.12E-03	3.31E-02
hsa04510	Cellular community	Focal adhesion	194	123	3.29E-03	4.89E-02
hsa04725	Nervous system	Cholinergic synapse	113	74	3.49E-03	4.89E-02
hsa04210	Cell growth and death	Apoptosis	131	78	3.54E-03	4.89E-02

*N* number of genes in the KEGG term, *DE* number of genes that are differentially methylated

upregulate *SLC7A11* expression, leading to inhibition of ferroptosis and promoting malignant progression of bladder cancer [72]. The *SLC7A11* protein is a cystine/ glutamate antiporter responsible for importing cystine, one of the three essential amino acids for glutathione biosynthesis and antioxidant defense. Additionally, a study highlighted a positive correlation between *SLC7A11* and other amino acid transporters and enzymes involved in glutamine metabolism, suggesting a collaborative role in metabolic regulation [58]. An extensive review reported that increased expression of *SLC7A11* suppresses ferroptosis and promotes tumor development [56].

There are limitations to this study. The CpG cg02711608 in the *SLC1A5* gene associated with alcohol consumption in multiple studies was not included in our analysis because it contained a common SNP (rs76693964). The methylation levels of CpGs near SNPs can be the reflection of underlying genetic polymorphisms; hence, it has been suggested that they be removed prior to data analysis [22, 73]. There could also be residual confounding resulting from the unavailability of direct measures

of the cellular composition of the blood samples and information bias due to the self-reported nature of alcohol consumption (social desirability bias and recall bias). Nevertheless, the sensitivity analysis showed similar results for models with and without adjustment for cell composition. Moreover, the time gap between the assessment of alcohol consumption (1993) and blood sample collection (1995–1999) could influence the results if drinking habits changed significantly during this period. However, these temporal differences are also inherent in longitudinal studies and still represent an added value compared to cross-sectional studies. We could not exclude residual confounding from adjusting for smoking status only, which does not take into account the dose and duration of smoking. Additionally, these associations were estimated solely in a population of mostly Caucasian origin and may not be generalizable to other ethnic groups.

However, this study has four main strengths. First, the use of a case-cohort design offers several advantages in cohort studies. For example, it enables multiple

**Table 6** Top 20 replicated CpG sites in the case–control study

CpG	Chr	CpG island region <sup>a</sup>	Gene <sup>b</sup>	Gene region <sup>c</sup>	Estimate	Standard error	P values
cg12388309	20	Island	<i>PAK7</i>	TSS200	− 0.03532	0.00569	1.60E−09
cg15123817	13	Island	<i>PDS5B</i>	TSS200	− 0.04230	0.00694	3.04E−09
cg09892390	10	Island	<i>ARHGAP21</i>	TSS1500	− 0.03324	0.00553	4.75E−09
cg04477202	2	Island	<i>SPAST</i>	Body	− 0.01331	0.00238	4.74E−08
cg14295187	2	Island	<i>MDH1;WDPCP;WDPCP;MDH1</i>	TSS1500;1stExon; 5'UTR;TSS200	− 0.04906	0.00882	5.46E−08
cg02477200	10	Island	<i>HECTD2;HECTD2;HECTD2;HECTD2-AS1</i>	5'UTR;1stExon; Body;Body	− 0.02767	0.00523	2.19E−07
cg01738152	17	S_Shore	<i>FMNL1</i>	Body	− 0.02523	0.00491	4.61E−07
cg16021609	15	OpenSea	<i>PARP6</i>	5'UTR	− 0.00824	0.00165	9.26E−07
cg21912872	12	OpenSea	<i>DYRK2</i>	3'UTR	− 0.00500	0.00101	1.08E−06
cg10403753	19	OpenSea	<i>PINLYP;PINLYP</i>	Body;TSS200	− 0.04482	0.00906	1.21E−06
cg27302585	20	OpenSea	<i>DHX35</i>	Body	− 0.00484	0.00100	1.80E−06
cg24705741	12	Island			− 0.02292	0.00478	2.45E−06
cg25537033	7	N_Shore	<i>SDK1;SDK1</i>	TSS200;Body	− 0.02664	0.00557	2.60E−06
cg20090464	16	Island	<i>CCDC102A</i>	TSS200	− 0.02338	0.00489	2.62E−06
cg17163592	21	N_Shelf	<i>U2AF1;U2AF1</i>	5'UTR;Body	− 0.00457	0.00096	2.95E−06
cg08755040	1	Island	<i>QSOX1</i>	Body	− 0.03634	0.00765	3.00E−06
cg05489160	16	OpenSea	<i>KARS</i>	Body	− 0.00597	0.00126	3.16E−06
cg04111761	3	OpenSea	<i>CCR3</i>	TSS1500	− 0.00560	0.00119	3.45E−06
cg00253735	1	OpenSea	<i>CTSK</i>	5'UTR	− 0.01240	0.00267	4.76E−06
cg17575972	22	OpenSea			− 0.00981	0.00212	5.22E−06

Chr chromosome, CpG cytosine-phosphate-guanine

<sup>a</sup> CpG islands are regions of approximately one thousand base pairs long, enriched in CpG dinucleotides. In humans, 72% of gene promoters contain CpG islands. CpG can then be localized toward these islands: Island = inside the island; Shelf = 2–4 kb from the island; Shore = 0–2 kb from the island; N = upstream (5') from the island; S = downstream (3') from the island; OpenSea ≥ 4 kb from the islands

<sup>b</sup> Multiple lists of the same gene name indicate splicing variants

<sup>c</sup> Gene is divided into several regions: TSS200 = 0–200 bases upstream of the transcription start site (TSS); TSS1500 = 200–1500 bases upstream of the TSS; 5'UTR = in the 5' untranslated region, between the TSS and the ATG start site; Body = between the ATG and the stop codon, regardless of the presence of introns, exons, TSS or promoters; 3'UTR = between the stop codon and the poly A signal

outcomes to be studied (unlike the case–control design) using the same sub-cohort, which improves the cost-effectiveness of longitudinal epidemiological investigations. This design is not frequently used in literature, but it could be of great interest, especially for omics studies [25]. However, this design is not without limitations. It may be prone to selection bias if the sub-cohort is not representative of the entire cohort, and there is also a risk of confounding bias if certain variables are not adequately controlled for. To mitigate these limitations, participants were randomly selected and stratified on key characteristics to ensure that the selection is as representative as possible of the entire cohort. Additionally, we constructed a directed acyclic graph to identify potential confounding factors. Finally, with this design data analysis requires more complex statistical methods that are now well established but may be more challenging in specific situations. The weights used in our study, designed to correct the over-representation of cases, may influence the results. As the weighting of non-cases was considerably greater than that of cases, the influence of

non-cases is more important on the study results. While this was necessary to balance the selection bias inherent in the study design, it also means that the estimated association between alcohol consumption and DNAm may be more strongly influenced by non-cases. Thus, while the case-cohort design presents some challenges, its cost-effectiveness is a significant advantage, especially in epigenetic studies, which are often expensive. This makes it a valuable tool in epidemiological research. Second, even if we did not have repeated measures for alcohol consumption or DNAm, we reduced the possibility of inverse causality bias by exploiting the prospective design of the cohort and considering the temporality of our variables of interest. The alcohol consumption was collected in 1993 and reflected the consumption in the past year, DNAm was measured in blood samples collected between 1995 and 1999, and covariates were measured at baseline (age and parity) or in 1992 (body mass index, smoking, and menopausal status). However, it is important to note that while this design helps clarify the temporal relationships between variables, the possibility that

**Table 7** List of DMPs associated with alcohol consumption in our EWAS and at least four other EWAS

CpG	Chr	CpG island region <sup>a</sup>	Gene <sup>b</sup>	Gene region <sup>c</sup>	Replicated <sup>d</sup>	n	Replicated in longitudinal analysis by Dugué et al. [5]
cg06690548	4	OpenSea	<i>SLC7A11</i>	Body	Yes	8	Yes
cg02583484	12	S_Shelf	<i>HNRNPA1; HNRNPA1L-2</i>	Body; Body	Yes	6	Yes
cg05512483	11	N_Shore	<i>SNORD30; SNORD22; SNORD29; SNORD31; SNHG1</i>	TSS1500; TSS1500; TSS1500; TSS1500; Body	No	6	Yes
cg00422488	9	Island	<i>ANP32B</i>	Body	No	6	Yes
cg12825509	3	OpenSea	<i>TRA2B</i>	Body	No*	6	No
cg18917643	5	OpenSea			No*	6	No
cg02578470	3	N_Shelf	<i>MCM2</i>	Body	No*	6	Yes
cg17058475	11	N_Shore	<i>CPT1A</i>	5'UTR	No*	5	Yes
cg09737197	11	N_Shore	<i>CPT1A</i>	5'UTR	Yes	5	No
cg16246545	1	S_Shore	<i>PHGDH</i>	Body	No	5	Yes
cg19731612	5	Island	<i>NSD1</i>	TSS1500	No	5	No
cg03339609	19	N_Shore	<i>LSM14A</i>	TSS1500	No*	5	Yes
cg03260624	1	S_Shelf	<i>CDC7</i>	Body	No*	5	Yes
cg18763536	12	OpenSea	<i>ETV6</i>	Body	Yes	5	Yes
cg04460609	4	OpenSea	<i>LDB2</i>	Body	No*	5	Yes
cg05603985	1	Island	<i>SKI</i>	1stExon	No	5	No
cg06983052	1	S_Shore	<i>LRRC8D</i>	5'UTR	No*	5	Yes
cg14482712	15	OpenSea			No*	5	Yes
cg24135793	19	Island	<i>NFIX</i>	Body	No*	5	No

Chr chromosome, CpG cytosine-phosphate-guanine; n: number of studies that identified the CpG site (including our EWAS)

<sup>a</sup> CpG islands are regions of approximately one thousand base pairs long, enriched in CpG dinucleotides. In humans, 72% of gene promoters contain CpG islands. CpG can then be localized toward these islands: Island = inside the island; Shelf = 2–4 kb from the island; Shore = 0–2 kb from the island; N = upstream (5') from the island; S = downstream (3') from the island; OpenSea ≥ 4 kb from the islands

<sup>b</sup> Multiple lists of the same gene name indicate splicing variants

<sup>c</sup> Gene is divided into several regions: TSS200 = 0–200 bases upstream of the transcription start site (TSS); TSS1500 = 200–1500 bases upstream of the TSS; 5'UTR = in the 5' untranslated region, between the TSS and the ATG start site; Body = between the ATG and the stop codon, regardless of the presence of introns, exons, TSS or promoters; 3'UTR = between the stop codon and the poly A signal

<sup>d</sup> Replicated in the case-control study

\* CpG sites with the same direction of association as the EWAS but a p value > 0.05

**Table 8** Overlap between DMRs identified in the EWAS on alcohol consumption and DMRs reported in the literature

Overlap id	Study	Chr	Start	End	n	Width in bp	Direction	Gene
1	Lohoff 2021 [16]	3	23,851,358	23,851,961	8	604	Hypomethylated	Not reported
1	Dragic 2025 (present study)	3	23,851,358	23,851,386	2	29	Hypomethylated	<i>UBE2E1</i>
2	Dugué 2019 [5]	6	30,457,711	30,460,798	44	3088	Hypomethylated	<i>HLA-E</i>
2	Dragic 2025 (present study)	20	30,458,515	30,459,158	7	644	Hypomethylated	<i>DUSP15; TLL9</i>
3	Dragic 2025 (present study)	12	49,318,784	49,319,673	13		Hypomethylated	<i>FKBP11</i>
3	Perrier 2019 [11]	12	49,319,500	49,319,673	3		Hypermethylated	<i>FKBP11</i>
4	Dugué 2019 [5]	19	56,158,209	56,160,065	12	1857	Hypomethylated	<i>CCDC106</i>
4	Dragic 2025 (present study)	19	56,159,710	56,159,817	2	108	Hypomethylated	<i>CCDC106</i>
5	Dugué 2019 [5]	11	68,606,654	68,608,813	9	2160	Hypomethylated	<i>CPT1A</i>
5	Dragic 2025 (present study)	11	68,607,257	68,608,225	6	969	Hypomethylated	<i>CPT1A</i>
5	Lohoff 2021 [16]	11	68,607,622	68,607,737	4	116	Hypomethylated	Not reported
6	Lohoff 2021 [16]	5	138,634,374	138,634,507	2	134	Hypomethylated	Not reported
6	Dragic 2025 (present study)	5	138,634,374	138,634,507	2	134	Hypomethylated	<i>MATR3</i>

Bp base pair, Chr chromosome, n number of probes in the differentially methylated region

DNAm could influence alcohol consumption cannot be ruled out, as DNAm can be hereditary. Third, we used a nested case–control study to validate our findings. Fourth, we have compared our results with the literature by using our previous systematic review.

## Conclusions

In conclusion, this study used a case-cohort study to identify blood DNAm changes associated with alcohol consumption at individual sites across the genome. The widespread DNAm alterations identified seem to differ according to the type of alcohol consumed and the quantity consumed. Differences between countries in the type and quantity of alcohol consumed may therefore explain the heterogeneity of results in the literature. Despite such heterogeneity, 19 CpGs have been identified in at least five studies. In particular, higher alcohol consumption was consistently associated with the hypomethylation of sites located in *SLC7A11*, *HNRNPA1*, *SNORD30*, *ANP32B*, and *MCM2* genes. The link between these sites and chronic diseases should be investigated as they may help to understand the molecular mechanisms underlying the harmful effects of alcohol consumption on health. Future research should focus on elucidating the biological pathways through which these DNAm changes contribute to disease development.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-025-01893-1>.

**Additional file 1: Fig. S1.** Workflow of the case-cohort selection. **Fig. S2.** Plot of the associations between alcohol consumption and DNA methylation at 715,986 CpG sites with and without education adjustment. **Fig. S3.** Plot of the associations between alcohol consumption and DNA methylation at 715,986 CpG sites in non-cases only ( $n = 612$ ) and in all participants ( $n = 1,538$ ). **Fig. S4.** Plot of the associations between alcohol consumption and DNA methylation at 715,986 CpG sites in all participants after excluding incident cancers occurring within the first five years after blood collection ( $n = 1,186$ ) and in all participants ( $n = 1,538$ ). **Fig. S5.** Plot of the associations between alcohol consumption and DNA methylation at 591 CpG sites with and without cellular composition adjustment. **Table S2.** Overlap of CpG sites ( $n = 30$ ) associated with alcohol consumption in EWAS on total alcohol consumption and by alcohol subtype. **Table S5.** Characteristics of the case-control study by alcohol consumption ( $n = 354$ ).

**Additional file 2: Table S1.** EWAS results for the 19,255 CpG sites associated with alcohol consumption in g/day ( $FDR < 0.05$ ). **Table S3.** Information on the CpG sites included in the 90 differentially methylated regions identified. **Table S4.** Gene over-representation analyses using KEGG pathways collection using DMPs associated with **a** alcohol consumption in continuous ( $n = 19,255$ ); **b** heavy drinking category ( $n = 13,101$ ); **c** moderate drinking category ( $n = 13,172$ ); **d** wine and champagne consumption ( $n = 19,632$ ); **e** cider and beer consumption ( $n = 13,131$ ); **f** spirits consumption ( $n = 13,387$ ); **g** alcohol consumption in continuous and replicated in the case-control study ( $n = 1,528$ ); **h** alcohol consumption in continuous, replicated in the case-control study and identified by other studies in the literature ( $n = 85$ ). **Table S6.** Replicated CpG sites in the case-control study. **Table S7.** List of the 8,411 DMPs identified in the literature. The studies identifying these sites are reported.

## Acknowledgements

We are grateful to the E3N-Generations study participants for their continued participation. We also thank all members of the E3N-Generations study group, the data-management team and, in particular Amandine Gelot, for data management and technical assistance. We acknowledge the contribution of the French National Centre for Research on Human Genomics (CNRGH) in Evry, France, where the methylation measurements were conducted and the Fondation Jean-Dausset-CEPH where DNA have been extracted from buffy coats.

## Author contributions

DD, FA, CD, and GS contributed to the study conception and design, and the elaboration of the statistical analysis plan. Material preparation and data analysis were performed by DD. DD drafted the first version of the manuscript. All authors read, revised, and approved the manuscript.

## Funding

This work was conducted with data from the E3N-Generations cohort (Inserm) and supported by the Mutuelle Générale de l'Éducation Nationale (MGEN), Gustave Roussy Institute, and French League Against Cancer for the constitution and maintenance of the cohort. The cohort has benefited from state funding managed by the French National Research Agency (ANR) under the program "Investment in the future" bearing the reference ANR-10-COHO-0006, as well as from a subsidy from the Ministry of Higher Education, Research and Innovation for public service charges bearing the reference number 2102918823, 2103236497, and 2103586016. The management of biological collections and activities such as extraction, quality control and quantification of DNA have benefited from funding from the project French Cohorts Biobank (BioCF, Equipex+). Dzevka Dragic holds a Scholarship of Excellence from the University Laval Cancer Research Center and a doctoral funding from the University Paris-Saclay Doctoral School of Public Health EDSP. The methylation analyses were funded from grants from the Ligue Nationale Contre le Cancer (PRE2016), the ARC Foundation for Cancer Research (CANCAIR201601245), and the French National Institute of Cancer (SHSESP 2017-30).

## Availability of data and materials

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

The study received ethical approval from the French National Commission for Data Protection and Privacy (CNIL) and all participants gave written informed consent. The study identifier at ClinicalTrials.gov is NCT03285230.

### Consent for publication

Not applicable.

### Competing interest

The authors declare no competing interests.

Received: 19 August 2024 Accepted: 6 May 2025

Published online: 07 July 2025

## References

1. Barbería-Latasa M, Gea A, Martínez-González MA. Alcohol, drinking pattern, and chronic disease. *Nutrients*. 2022;14(9):693.
2. Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. *Science*. 2001;293(5532):1068–70.
3. Alegría-Torres JA, Baccarelli A, Bollati V. Epigenetics and lifestyle. *Epigenomics*. 2011;3(3):267–77.
4. Dragic D, Chang SL, Ennour-Idrissi K, Durocher F, Severi G, Diorio C. Association between alcohol consumption and DNA methylation in blood: a systematic review of observational studies. *Epigenomics*. 2022;14(12):793–810.

5. Dugué PA, Wilson R, Lehne B, Jayasekara H, Wang X, Jung CH, et al. Alcohol consumption is associated with widespread changes in blood DNA methylation: analysis of cross-sectional and longitudinal data. *Addict Biol*. 2021;26(1):e12855.
6. Stephenson M, Bollepalli S, Cazaly E, Salvatore JE, Barr P, Rose RJ, et al. Associations of alcohol consumption with epigenome-wide DNA methylation and epigenetic age acceleration: individual-level and co-twin comparison analyses. *Alcohol Clin Exp Res*. 2021;45(2):318–28.
7. Liang X, Justice AC, So-Armah K, Krystal JH, Sinha R, Xu K. DNA methylation signature on phosphatidylethanol, not on self-reported alcohol consumption, predicts hazardous alcohol consumption in two distinct populations. *Mol Psychiatry*. 2021;26(6):2238–53.
8. Cronjé HT, Elliott HR, Nienaber-Rousseau C, Pieters M. Replication and expansion of epigenome-wide association literature in a black South African population. *Clin Epigenetics*. 2020;12(1):6.
9. Langdon R, Richmond R, Elliott HR, Dudding T, Kazmi N, Penfold C, et al. Identifying epigenetic biomarkers of established prognostic factors and survival in a clinical cohort of individuals with oropharyngeal cancer. *Clin Epigenetics*. 2020;12(1):95.
10. Wilson LE, Xu Z, Harlid S, White AJ, Troester MA, Sandler DP, et al. Alcohol and DNA methylation: an epigenome-wide association study in blood and normal breast tissue. *Am J Epidemiol*. 2019;188(6):1055–65.
11. Perrier F, Viallon V, Ambatipudi S, Ghantous A, Cuenin C, Hernandez-Vargas H, et al. Association of leukocyte DNA methylation changes with dietary folate and alcohol intake in the EPIC study. *Clin Epigenetics*. 2019;11(1):57.
12. Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai PC, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. *Mol Psychiatry*. 2018;23(2):422–33.
13. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, et al. Epigenetic prediction of complex traits and death. *Genome Biol*. 2018;19(1):136.
14. Clark SL, Aberg KA, Nerella S, Kumar G, McClay JL, Chen W, et al. Combined Whole Methylome and Genomewide Association Study Implicates CNTN4 in Alcohol Use. *Alcohol Clin Exp Res*. 2015;39(8):1396–405.
15. Flanagan JM, Brook MN, Orr N, Tomczyk K, Coulson P, Fletcher O, et al. Temporal stability and determinants of white blood cell DNA methylation in the breakthrough generations study. *Cancer Epidemiol Biomarkers Prev*. 2015;24(1):221–9.
16. Lohoff FW, Clarke T-K, Kaminsky ZA, Walker RM, Birmingham ML, Jung J, et al. Epigenome-wide association study of alcohol consumption in N = 8161 individuals and relevance to alcohol use disorder pathophysiology: identification of the cystine/glutamate transporter SLC7A11 as a top target. *Mol Psychiatry*. 2022;27(3):1754–64.
17. Clavel-Chapelon F. Cohort Profile: The French E3N Cohort Study. *Int J Epidemiol*. 2015;44(3):801–9.
18. van Liere MJ, Lucas F, Clavel F, Slimani N, Villeminot S. Relative validity and reproducibility of a French dietary history questionnaire. *Int J Epidemiol*. 1997;26(Suppl 1):S128–36.
19. Fagherazzi G, Vilier A, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. Alcohol consumption and breast cancer risk subtypes in the E3N-EPIC cohort. *Eur J Cancer Prev*. 2015;24(3):209–14.
20. Groen K, Lea RA, Maltby VE, Scott RJ, Lechner-Scott J. Letter to the editor: blood processing and sample storage have negligible effects on methylation. *Clin Epigenetics*. 2018;10(1):22.
21. Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. *Bioinformatics*. 2016;32(17):2659–63.
22. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data*. 2016;9:22–4.
23. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11(1):587.
24. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC Bead-Array. *Genome Biol*. 2018;19(1):64.
25. O'Brien KM, Lawrence KG, Keil AP. The case for case-cohort: an applied epidemiologist's guide to reframing case-cohort studies to improve usability and flexibility. *Epidemiology*. 2022;33(3):354–61.
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B (Methodol)*. 1995;57(1):289–300.
27. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ*. 1995;310(6973):170.
28. Matthew S, James RS, Robert F, Ryan A, Andrew S, Kate T. dmrff: identifying differentially methylated regions efficiently with power and control. *bioRxiv*. 2018:508556.
29. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 2015;32(2):286–8.
30. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. *Genome Biol*. 2021;22(1):173.
31. Mancini FR, Cano-Sancho G, Mohamed O, Cervenka I, Omichessan H, Marchand P, et al. Plasma concentration of brominated flame retardants and postmenopausal breast cancer risk: a nested case-control study in the French E3N cohort. *Environ Health*. 2020;19(1):54.
32. Omichessan H, Perduca V, Polidoro S, Kvaskoff M, Truong T, Cano-Sancho G, et al. Associations between plasma levels of brominated flame retardants and methylation of DNA from peripheral blood: a cross-sectional study in a cohort of French women. *Environ Res*. 2022;210:112788.
33. Tu H, Tang LJ, Luo XJ, Ai KL, Peng J. Insights into the novel function of system Xc<sup>-</sup> in regulated cell death. *Eur Rev Med Pharmacol Sci*. 2021;25(3):1650–62.
34. Todorov IT, Attaran A, Kearsey SE. BM28, a human member of the MCM2-3-5 family, is displaced from chromatin during DNA replication. *J Cell Biol*. 1995;129(6):1433–45.
35. Shen SM, Yu Y, Wu YL, Cheng JK, Wang LS, Chen GQ. Downregulation of ANP32B, a novel substrate of caspase-3, enhances caspase-3 activation and apoptosis induction in myeloid leukemic cells. *Carcinogenesis*. 2010;31(3):419–26.
36. Shin CS, Mishra P, Watrous JD, Carelli V, D'Aurelio M, Jain M, et al. The glutamate/cystine xCT antiporter antagonizes glutamine metabolism and reduces nutrient flexibility. *Nat Commun*. 2017;8:15074.
37. Zhang Y, Yu H, Zhang J, Gao H, Wang S, Li S, et al. Cui4A-DDB1-mediated monoubiquitination of phosphoglycerate dehydrogenase promotes colorectal cancer metastasis via increased S-adenosylmethionine. *J Clin Invest*. 2021;131(21):63.
38. Yang H, Zhao H, Ren Z, Yi X, Zhang Q, Yang Z, et al. Overexpression CPT1A reduces lipid accumulation via PPARα/CD36 axis to suppress the cell proliferation in ccRCC. *Acta Biochim Biophys Sin (Shanghai)*. 2022;54(2):220–31.
39. Munemasa Y, Suzuki T, Aizawa K, Miyamoto S, Imai Y, Matsumura T, et al. Promoter region-specific histone incorporation by the novel histone chaperone ANP32B and DNA-binding factor KLF5. *Mol Cell Biol*. 2008;28(3):1171–81.
40. Dreyfuss G, Matunis MJ, Piñol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem*. 1993;62:289–321.
41. Bochman ML, Schwacha A. The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev*. 2009;73(4):652–83.
42. Lopez RG, Carron C, Oury C, Gardellin P, Bernard O, Ghysdael J. TEL is a sequence-specific transcriptional repressor. *J Biol Chem*. 1999;274(42):30132–8.
43. Vallerga CL, Zhang F, Fowdar J, McRae AF, Qi T, Nabais MF, et al. Analysis of DNA methylation associates the cystine-glutamate antiporter SLC7A11 with risk of Parkinson's disease. *Nat Commun*. 2020;11(1):1238.
44. Wang C, Liu H, Xu S, Deng Y, Xu B, Yang T, et al. Ferroptosis and Neurodegenerative Diseases: Insights into the Regulatory Roles of SLC7A11. *Cell Mol Neurobiol*. 2023;43(6):2627–42.
45. Chen X, Calandrelli R, Girardini J, Yan Z, Tan Z, Xu X, et al. PHGDH expression increases with progression of Alzheimer's disease pathology and symptoms. *Cell Metab*. 2022;34(5):651–3.
46. Salapa HE, Thibault PA, Libner CD, Ding Y, Clarke JWE, Denomy C, et al. hnRNP A1 dysfunction alters RNA splicing and drives neurodegeneration in multiple sclerosis (MS). *Nat Commun*. 2024;15(1):356.
47. Naruse H, Ishiura H, Mitsui J, Date H, Takahashi Y, Matsukawa T, et al. Molecular epidemiological study of familial amyotrophic lateral sclerosis in Japanese population by whole-exome sequencing and identification of novel HNRNPA1 mutation. *Neurobiol Aging*. 2018;61:255.e9–e16.
48. Villa C, Fenoglio C, De Riz M, Clerici F, Marcone A, Benussi L, et al. Role of hnRNP-A1 and miR-590-3p in neuronal death: genetics and expression

- analysis in patients with Alzheimer disease and frontotemporal lobar degeneration. *Rejuvenation Res.* 2011;14(3):275–81.
49. Ren Q, Guo M, Yang F, Han T, Du W, Zhao F, et al. Association of CPT1A gene polymorphism with the risk of gestational diabetes mellitus: a case-control study. *J Assist Reprod Genet.* 2021;38(7):1861–9.
  50. Lai CQ, Parnell LD, Smith CE, Guo T, Sayols-Baixeras S, Aslibekyan S, et al. Carbohydrate and fat intake associated with risk of metabolic diseases through epigenetics of CPT1A. *Am J Clin Nutr.* 2020;112(5):1200–11.
  51. Wang W, Cai Y, Deng G, Yang Q, Tang P, Wu M, et al. Allelic-specific regulation of xCT expression increases susceptibility to tuberculosis by modulating microRNA-mRNA interactions. *mSphere.* 2020;5(2):69.
  52. Kattimani Y, Veerappa AM. Complex interaction between mutant HNRNPA1 and gE of varicella zoster virus in pathogenesis of multiple sclerosis. *Autoimmunity.* 2018;51(4):147–51.
  53. Staller E, Sheppard CM, Baillon L, Frise R, Peacock TP, Sancho-Shimizu V, et al. A natural variant in ANP32B impairs influenza virus replication in human cells. *J Gen Virol.* 2021;102(9):45.
  54. Zheng J. Diagnostic value of MCM2 immunocytochemical staining in cervical lesions and its relationship with HPV infection. *Int J Clin Exp Pathol.* 2015;8(1):875–80.
  55. Gessner BD, Gillingham MB, Wood T, Koeller DM. Association of a genetic variant of carnitine palmitoyltransferase 1A with infections in Alaska Native children. *J Pediatr.* 2013;163(6):1716–21.
  56. Koppula P, Zhuang L, Gan B. Cystine transporter SLC7A11/xCT in cancer: ferroptosis, nutrient dependency, and cancer therapy. *Protein Cell.* 2021;12(8):599–620.
  57. Shan Q, Zhang C, Li Y, Li Q, Zhang Y, Li X, et al. SLC7A11, a potential immunotherapeutic target in lung adenocarcinoma. *Sci Rep.* 2023;13(1):18302.
  58. Nath P, Alfarsi LH, El-Ansari R, Masisi BK, Erkan B, Fakroun A, et al. The amino acid transporter SLC7A11 expression in breast cancer. *Cancer Biol Ther.* 2024;25(1):2291855.
  59. Siculella L, Giannotti L, Di Chiara SB, Spedicato F, Calcagnile M, Quarta S, et al. A comprehensive understanding of hnRNP A1 role in cancer: new perspectives on binding with noncoding RNA. *Cancer Gene Ther.* 2023;30(3):394–403.
  60. Erdem M, Ozgul İ, Dioken DN, Gurcuoglu I, Guntekin Ergun S, Cetin-Atalay R, et al. Identification of an mRNA isoform switch for HNRNPA1 in breast cancers. *Sci Rep.* 2021;11(1):24444.
  61. Li T, Wang N, Li S, Yan H, Gao S, Gao W, et al. ANP32B promotes lung cancer progression by regulating VDAC1. *Gene.* 2023;859:147200.
  62. Wang K, Luo L, Fu S, Wang M, Wang Z, Dong L, et al. PHGDH arginine methylation by PRMT1 promotes serine synthesis and represents a therapeutic vulnerability in hepatocellular carcinoma. *Nat Commun.* 2023;14(1):1011.
  63. Mullarky E, Mattaini KR, Vander Heiden MG, Cantley LC, Locasale JW. PHGDH amplification and altered glucose metabolism in human melanoma. *Pigment Cell Melanoma Res.* 2011;24(6):1112–5.
  64. Sun Y, Cheng Z, Liu S. MCM2 in human cancer: functions, mechanisms, and clinical significance. *Mol Med.* 2022;28(1):128.
  65. Aiderus A, Black MA, Dunbier AK. Fatty acid oxidation is associated with proliferation and prognosis in breast and other cancers. *BMC Cancer.* 2018;18(1):805.
  66. Liang JZ, Li YH, Zhang Y, Wu QN, Wu QL. Expression of ETV6/TEL is associated with prognosis in non-small cell lung cancer. *Int J Clin Exp Pathol.* 2015;8(3):2937–45.
  67. Zhou F, Chen B. Acute myeloid leukemia carrying ETV6 mutations: biologic and clinical features. *Hematology.* 2018;23(9):608–12.
  68. Lee S, Hu Y, Loo SK, Tan Y, Bhargava R, Lewis MT, et al. Landscape analysis of adjacent gene rearrangements reveals BCL2L14-ETV6 gene fusions in more aggressive triple-negative breast cancer. *Proc Natl Acad Sci U S A.* 2020;117(18):9912–21.
  69. Curtis BJ, Zahs A, Kovacs EJ. Epigenetic targets for reversing immune defects caused by alcohol exposure. *Alcohol Res.* 2013;35(1):97–113.
  70. Mason JB, Choi SW. Effects of alcohol on folate metabolism: implications for carcinogenesis. *Alcohol.* 2005;35(3):235–41.
  71. Albano E. Alcohol, oxidative stress and free radical damage. *Proc Nutr Soc.* 2006;65(3):278–90.
  72. Shen L, Zhang J, Zheng Z, Yang F, Liu S, Wu Y, et al. PHGDH Inhibits Ferroptosis and Promotes Malignant Progression by Upregulating SLC7A11 in Bladder Cancer. *Int J Biol Sci.* 2022;18(14):5459–74.
  73. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* 2013;8(2):203–9.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.