Author/s:
Dugué, PA;Wilson, R;Lehne, B;Jayasekara, H;Wang, X;Jung, CH;Joo, JHE;Makalic, E;Schmidt, DF;Baglietto, L;Severi, G;Gieger, C;Ladwig, KH;Peters, A;Kooner, JS;Southey, MC;English, DR;Waldenberger, M;Chambers, JC;Giles, GG;Milne, RL

Title:
Alcohol consumption is associated with widespread changes in blood DNA methylation: Analysis of cross-sectional and longitudinal data

Date:
2021-01-01

Citation:
Dugué, P. A., Wilson, R., Lehne, B., Jayasekara, H., Wang, X., Jung, C. H., Joo, J. H. E., Makalic, E., Schmidt, D. F., Baglietto, L., Severi, G., Gieger, C., Ladwig, K. H., Peters, A., Kooner, J. S., Southey, M. C., English, D. R., Waldenberger, M., Chambers, J. C. ,... Milne, R. L. (2021). Alcohol consumption is associated with widespread changes in blood DNA methylation: Analysis of cross-sectional and longitudinal data. Addiction Biology, 26 (1), https://doi.org/10.1111/adb.12855.

Persistent Link:
https://hdl.handle.net/11343/286704

Dugué Pierre-Antoine (Orcid ID: 0000-0003-2736-3023)
Milne Roger (Orcid ID: 0000-0001-5764-7268)

Alcohol consumption is associated with widespread changes in blood DNA methylation: analysis of cross-sectional and longitudinal data

Pierre-Antoine Dugué[1,2,3*], Rory Wilson[4,5], Benjamin Lehne[6], Harindra Jayasekara[1,7], Xiaochuan Wang[1], Chol-Hee Jung[8], JiHoon E Joo[7], Enes Makalic[2], Daniel F Schmidt[2], Laura Baglietto[9], Gianluca Severi[10], Christian Gieger[4,5], Karl-Heinz Ladwig[5,11], Annette Peters[4,5,12], Jaspal S Kooner[13,14,15,16], Melissa C Southey[3,17], Dallas R English[1,2], Melanie Waldenberger[4,5,12], John C Chambers[6,13,14,15,18], Graham G Giles[1,2,3], Roger L Milne[1,2,3]

[1] Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia
[2] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC, Australia
[3] Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia
[4] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, D-85764 Neuherberg, Germany
[5] Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, D-85764 Neuherberg, Germany
[6] Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK
[7] Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Melbourne, VIC, Australia
[8] Melbourne Bioinformatics, The University of Melbourne, Parkville, VIC, Australia
[9] Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy
[10] CESP, INSERM U1018, Univ. Paris-Sud, UVSQ, Université Paris-Saclay, Villejuif, France; Gustave Roussy, Villejuif, France
[11] Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie des Klinikums Rechts der Isar der TUM, Munich, Germany
[12] German Center for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany
[13] Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK
[14] Imperial College Healthcare NHS Trust, London W12 0HS, UK
[15] MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, UK
[16] National Heart and Lung Institute, Imperial College London, London W12 0NN, UK
[17] Genetic Epidemiology Laboratory, Department of Clinical Pathology, University of Melbourne, VIC, Australia
[18] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore

* Corresponding author: Pierre-Antoine Dugué, Cancer Epidemiology Division, Cancer Council Victoria, pdugue@cancervic.org.au

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/adb.12855

ABSTRACT

DNA methylation may be one of the mechanisms by which alcohol consumption is associated with the risk of disease. We conducted a large-scale, cross-sectional, genome-wide DNA methylation association study of alcohol consumption and a longitudinal analysis of repeated measurements taken several years apart. Using the Illumina HumanMethylation450 BeadChip, DNA methylation was measured in blood samples from 5,606 Melbourne Collaborative Cohort Study (MCCS) participants. For 1,088 of them, these measures were repeated using blood samples collected a median of 11 years later. Associations between alcohol intake and blood DNA methylation were assessed using linear mixed-effects regression models. Independent data from the LOLIPOP (N=4,042) and KORA (N=1,662) cohorts were used to replicate associations discovered in the MCCS. Cross-sectional analyses identified 1,414 CpGs associated with alcohol intake at $P<10^{-7}$, 1,243 of which had not been reported previously. Of these novel associations, 1,078 were replicated ($P<0.05$) using LOLIPOP and KORA data. Using the MCCS data, we also replicated 403 of 518 previously reported associations. Interaction analyses suggested that associations were stronger for women, non-smokers, and participants genetically predisposed to consume less alcohol. Of the 1,414 CpGs, 530 were differentially methylated ($P<0.05$) in former compared with current drinkers. Longitudinal associations between the change in alcohol intake and the change in methylation were observed for 513 of the 1,414 cross-sectional associations. Our study indicates that alcohol intake is associated with widespread changes in DNA methylation across the genome. Longitudinal analyses showed that the methylation status of alcohol-associated CpGs may change with alcohol consumption changes in adulthood.

KEYWORDS

Epigenome-wide association study, EWAS, alcohol consumption, DNA methylation, HM450 assay, cross-sectional data, longitudinal data

INTRODUCTION

Deoxyribonucleic acid (DNA) methylation is the addition of methyl groups to the 5'carbon of cytosine in cytosine-guanine dinucleotides (CpGs) and is thought to play a role in the development of disease through its influence on gene expression and cellular function [1, 2] DNA methylation is strongly affected by the underlying genetic DNA sequence,[3] sex, age and ethnicity [4, 5] and is modified by lifestyle factors and environmental exposures such as smoking and adiposity.[6-10]

Alcohol consumption is a major lifestyle risk factor contributing to the worldwide burden of disease, responsible for an estimated 2.7 million deaths and 4% of the global burden of disease annually.[11] Even modest use of alcohol may increase disease risk, but greatest risks are observed with heavy and long-term drinking. Alcohol consumption is a potentially modifiable risk factor that can be targeted with preventive interventions at both the policy and the individual levels.[12] Molecular mechanisms such as DNA methylation are thought to underlie or enhance a predisposition to addictions and substance abuse, including alcohol drinking,[13] but although there is a plausible relationship between alcohol intake and altered one-carbon metabolism and DNA methylation,[14] most evidence comes from studies that either had small sample size, were not specific to humans, or were carried out using tissues other than blood.[15-18] Identifying associations between alcohol consumption and DNA methylation may be useful to derive biomarkers of alcohol-related conditions. To our knowledge, only one large genome-wide study of blood DNA methylation and alcohol consumption has been conducted.[19]

In the present study, we sought to (i) identify novel associations between alcohol consumption and blood DNA methylation and assess whether these were modified by factors such as age, sex, or genetic prediction to drink alcohol, (ii) replicate previously reported associations, (iii) assess the

reversibility of associations, (iv) assess associations with changes in alcohol consumption using longitudinally collected data, and (v) assess overrepresentation of the discovered genes in biological pathways to investigate the potential functional implications of our findings. We used samples from the Melbourne Collaborative Cohort Study (MCCS) to discover potential associations and sought to replicate the findings using samples from the Cooperative Health Research in the Augsburg Region (KORA) and London Life Sciences Prospective Population (LOLIPOP) studies.

MATERIALS AND METHODS

Study participants

Between 1990 and 1994 (baseline), 41,513 participants were recruited to the Melbourne Collaborative Cohort study (MCCS). The majority (99%) were aged 40 to 69 years and 41% were men. Southern European migrants were oversampled to extend the range of lifestyle-related exposures.[20] Participants were contacted again between 2003 and 2007. Blood samples were taken at baseline and follow-up from 99% and 64% of participants, respectively. Baseline samples were stored as dried blood spots on Guthrie cards for the majority (73%), as mononuclear cell samples for 25% and as buffy coat samples for 2% of the participants. Follow-up samples were stored as buffy coat aliquots and dried blood spots on Guthrie cards.

All participants provided written informed consent and the study protocols were approved by the Cancer Council Victoria Human Research Ethics Committee.

The present study sample comprised MCCS participants selected for inclusion in one of seven previously conducted nested case-control studies of DNA methylation.[21-25] Controls were matched to

incident cases of prostate, colorectal, gastric, lung or kidney cancer, urothelial cell carcinoma or mature B-cell neoplasms on: sex, year of birth, country of birth, baseline sample type and smoking status (the latter for the lung cancer study only). Participants included in each nested case-control study were free of cancer at baseline. After quality control, methylation data for baseline blood samples (baseline study) were available for 5,606 MCCS participants. Methylation measures were repeated in DNA extracted from blood samples collected on Guthrie cards at follow-up (longitudinal study) and, after quality control, were available for a subset of 1,088 of the controls who also had their baseline sample collected on a Guthrie card (***Table 1***).

Alcohol and other variables

At both baseline and follow-up, participants completed questionnaires that included detailed questions on demographic characteristics, medical history, cigarette smoking, alcohol consumption, physical activity and diet, the latter using food frequency questionnaires. On both occasions, anthropometric measurements were obtained by trained personnel using standard procedures. Height was only measured at baseline.

Alcohol intake at baseline was recorded as frequency and quantity of intake per drinking occasion by type (beer, wine, spirits) and by decade of age starting from 20 years. Participants were also asked about their alcohol intake on each day during the previous week, in terms of the number, measure and type of drink (e.g. two glasses of wine). At follow-up, the frequency and quantity of intake, by type of drink, during the previous calendar year were assessed as described above. Grams per day (g/day) were calculated as reported previously.[26] Participants who reported a weekly or current intake >200g/day were excluded. Four alcohol consumption variables were considered: g/day in the last week (continuous), g/day in the current decade (continuous), g/day over the lifetime

(continuous), and drinking status (never, former, current; based on current decade and lifetime variables).

DNA methylation and genetic data

Methods relating to DNA extraction and bisulfite conversion, DNA methylation data processing, normalisation and quality control, and genotyping are described in the ***Supplementary Methods***.

Methylome-wide association study (EWAS)

We assessed cross-sectional associations at each individual CpG by regressing DNA methylation M-values on alcohol consumption using linear mixed-effects regression models, using the function *lmer* from the R package *lme4*. Alcohol intake was represented using three continuous variables (consumption in the previous year, consumption in the previous week, and lifetime consumption, in g/day) that were modelled separately. Models were adjusted by fitting fixed effects for age (continuous), sex, smoking status (never, former ≥15 years ago, former <15 years ago, current <20 cigarettes per day, current ≥20 cigarettes per day), body mass index (BMI) (≤25 kg/m^2, >25 to ≤30, >30), country of birth (Australia/New-Zealand, Italy, Greece, United Kingdom), sample type (peripheral blood mononuclear cells, dried blood spots, buffy coats) and white blood cell composition (percentage of CD4+ T cells, CD8+ T cells, B cells, NK cells, monocytes and granulocytes, estimated using the Houseman algorithm[27]), and random effects for study, plate, and chip. A significance threshold of P-value<10^{-7} was used to account for multiple testing.[7] The False Discovery Rate (FDR) was used to identify suggestive associations.

Interaction terms were tested between alcohol consumption (previous week) and each of age (continuous), sex, smoking status (pseudo-continuous variable 1: Never, to 5: Current smoker >20

cigarettes/day), BMI (continuous), country of birth (as categorised above), future cancer case status, and a polygenic score for alcohol consumption, the latter derived as described below. These analyses were restricted to CpGs associated at $P<10^{-7}$ in the cross-sectional analysis with any of the three continuous alcohol variables ('last week', 'current decade' and 'lifetime' intake).

Associations for each type of alcohol (beer, wine, spirits) were assessed by including the three variables in a same model, so that intake of each type was adjusted for the two others. This was done for the 'current decade', and 'lifetime' alcohol intake variables because type of alcoholic beverage was not determined for the 'alcohol last week' variable.

Sensitivity analyses were conducted to assess potential confounding by: i) fitting the same models without adjustment for smoking or BMI; ii) fitting the same models with additional adjustment for socioeconomic status, educational attainment, a physical activity score based on metabolic equivalents,[28] and a score of healthy dietary habits;[29] iii) examining whether methylation was associated with these covariates; and iv) using less stringent adjustment for white blood cell composition by omitting each one of the six cell type variables.

Differentially methylated regions

We also tested for regions of contiguous CpGs associated with alcohol using the DMRcate package,[30] with default options (bandwidth of λ=1000 base pairs and a scaling factor C=2 for the Gaussian kernel smoother). The models were adjusted using as fixed effects age (continuous), sex, smoking status, country of birth, sample type, white cell composition (all categorized as described above), study and plate. Associations with region-wise Stouffer P-value$<10^{-7}$ were considered statistically significant.

Replication of novel associations

CpGs associated with alcohol intake at $P<10^{-7}$ in the MCCS were selected for replication using data from the KORA cohort (N=1,662, assessed in 1999-2001) including German participants and the LOLIPOP cohort (N=4,042, assessed in 2003-2008), including predominantly Asian participants, respectively.[31, 32] Alcohol intake in KORA was defined as the average of alcohol intake 'in previous day' and 'in previous week'. Alcohol intake in LOLIPOP was defined over a week (***Supplementary Methods***). Each cohort applied a normalisation method based on control probes [33] and adjusted models for the same covariates defined in a similar way to those used in the MCCS analyses, ***Supplementary Table 1, Supplementary Methods.*** Results from the two cohorts were pooled using fixed-effects meta-analysis with inverse-variance weights.[34] An association was considered replicated if P<0.05 and the direction of association was the same as in the MCCS.[7]

Polygenic score for alcohol consumption

A polygenic score for alcohol consumption was constructed using MCCS data based on the genome-wide association study by Clarke and colleagues, which identified 14 single nucleotide polymorphisms (SNPs) associated with alcohol consumption using UK Biobank data.[35] Data for 13 out of the 14 SNPs were available and were combined using the formula: *Polygenic score[i]* = $b_1 * d_{1,i} + \ldots b_{13} * d_{13,i,}$ where $d_{k,i}$ is the imputed allele dosage of variant *k* for person *i*, and b_k the per-allele regression coefficient reported for SNP *k.*

Replication of previously reported associations with alcohol consumption

Using MCCS data, we assessed replication (P<0.05) of associations with alcohol consumption (g/day) reported in a recent pooled, large-scale analysis of Europeans and African Americans.[19] A

total of 518 CpGs were considered, comprising 363 identified for participants of European ancestry and a further 155 CpGs identified for those of African ancestry.

Reversibility of associations

We calculated regression coefficients for comparisons of '*former'* to '*never'*, '*current'* to '*never'*, and '*current'* to '*former'* drinkers using MCCS data. As there were too few never-drinkers in the KORA data (N=22), we only considered the comparison '*former'* to '*current'*; we pooled the latter using fixed-effects meta-analysis.

In the MCCS, we calculated a 'reversibility coefficient', expressed as a percentage defined as: coefficient ('*former'* compared with '*current'*) / coefficient ('*never'* compared with '*current'*). These analyses were undertaken for CpGs with $P<10^{-7}$ in the MCCS EWAS.

Longitudinal associations

We further examined longitudinal associations with alcohol consumption for CpGs with $P<10^{-7}$ in the MCCS EWAS, incorporating data from follow-up. Linear mixed-effects regression models were used to assess the association between changes in DNA methylation (outcome) and changes in alcohol consumption (exposure) at each CpG. The change in alcohol consumption was computed in g/day as the difference between follow-up (alcohol intake in previous year) and baseline (previous week intake); study was included as a random effect and the following variables were included as fixed effects: baseline alcohol intake, baseline BMI and change in BMI (continuous), baseline age and change in age (continuous), sex, smoking status at baseline (as defined previously), smoking status at follow-up (yes/no), country of birth (as defined previously), baseline cell composition (as defined previously), change in each cell type composition (continuous) and baseline methylation M-

value. The change in methylation was calculated as the difference between follow-up and baseline ComBat-normalised methylation M-values. The same analyses were conducted for the KORA cohort, in which methylation measures taken approximately seven years later (2006-2008) were available for 1,332 participants (***Supplementary Methods***). As adjustment for baseline methylation in analyses of change in methylation may lead to bias in some circumstances,[36] we conducted a sensitivity analysis using models without adjustment for baseline methylation in the MCCS.

Pathway analyses

We used the *gometh* function from the *missMethyl* package [37] for pathway analyses assessing over-representation relative to all Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) pathways.[38, 39] To investigate potentially different biological pathways underlying i) acute compared with chronic alcohol consumption and ii) most dynamic compared with least dynamic methylation sites, *gometh* was applied, respectively, to i) CpGs associated with 'last week' and 'lifetime' alcohol intake, and ii) CpGs with a reversibility coefficient greater and lower than 50%. A nominal *P*-value lower than 0.05 was considered to indicate a potentially relevant pathway.

All statistical analyses were performed using the software R (version 3.4.0).

RESULTS

Altogether, 5,606 MCCS participants were included in the cross-sectional analysis; their median age was 61 years (IQR: 54-65), 68% were males, and alcohol intakes were wide-ranging (***Table 1)***. There were moderate-to-high correlations between the baseline alcohol variables, as well as between the baseline and follow-up variables, with Spearman correlations ranging from 0.68 to 0.76

(***Supplementary Table 2***). Participants in the longitudinal analysis were younger and generally had healthier lifestyle.

Methylome-wide association study

Across the three cross-sectional continuous alcohol variables considered, we observed 1,414 associations with $P<10^{-7}$. The most statistically significant associations are presented in ***Table 2***. There were 1,318, 358, and 392 CpGs associated with alcohol intake over the last week, decade and lifetime, respectively ***(Supplementary Table 3, Supplementary Figure 1)***. Associations were consistently stronger for alcohol intake in the previous week (***Supplementary Figure 2***).

Of these 1,414 associations, 1,243 were novel (i.e. not reported in Liu et al. [19]), and the CpGs were located in 842 genes; 241 CpGs were intergenic. Compared to the rest of the HM450 assay, alcohol-associated CpGs were over-represented in gene bodies (binomial proportion test, P=0.001), promoter regions TSS1500 ($P=5\times10^{-6}$) and 5'UTR ($P=3\times10^{-8}$); unannotated regions were underrepresented ($P=2\times10^{-5}$). These 1,414 associations corresponded to 1,084 unique clusters when considering that CpGs separated by less than 50kb of each other formed a single genomic region (***Supplementary Table 4***). For the overwhelming majority (99%) of associations, greater alcohol intake was associated with lower methylation. The number of CpGs associated with alcohol intake over the last week, decade and lifetime was 16,732, 5,751 and 6,585, respectively when considering an FDR-adjusted $P<0.05$ (***Supplementary Table 3***).

Genomic regions associated with alcohol intake: Applying DMRcate to the MCCS data, association with alcohol intake in last week was observed for methylation at 221 regions of contiguous CpGs ($P<1.0\times10^{-7}$) (***Supplementary Table 19***). Several of the most statistically significant regions

contained a large number of CpGs annotated to genes that were not ranked among the highest in the EWAS; for example, a cluster was identified in each of *MAB21L1* (23 CpGs, $P=2x10^{-38}$); *HOXA4* (25 CpGs, $P=3x10^{-31}$); GAS5 (9 CpGs, $P=4x10^{-30}$); SNHG1(9 CpGs, $P=6x10^{-30}$).

Replication of novel associations using external data: Of the 1,243 novel associations, 1,078 (87%) were replicated using data from KORA and LOLIPOP. Replication rates were 87%, 89% and 93% for CpGs associated with alcohol intake over the last week, decade and lifetime, respectively ($P<0.05$; ***Table 2*** and ***Supplementary Table 5***).

Interaction analyses: Using the Bonferroni correction for multiple testing ($P=0.05/1,414=3.5x10^{-5}$) and the 'last week' alcohol intake variable, we observed stronger associations for women at cg13446906 (*MIR548F5*), and cg22363327 (*SFRS13B*), and weaker associations for smokers at cg05104080 (*ILKAP*), cg17058475 (*CPT1A*), and cg01395047 (*TLR9*). At $P<0.05$, stronger associations were observed for women at 200 CpGs (test for binomial proportions, $P=6x10^{-29}$); weaker associations were observed for participants with a higher smoking score (N= 159, $P=6x10^{-20}$) and a higher BMI (N=165, P=0.003) (***Table 3; Supplementary Table 15***).

For the MCCS sample with genetic and DNA methylation data (N=3,859), the polygenic score was positively associated with alcohol intake in previous week ($P=9x10^{-5}$) (***Supplementary Table 13***). Weaker associations ($P<0.05$) for participants with higher polygenic score were observed for 156 CpGs ($P=2x10^{-19}$), although none passed the Bonferroni correction (***Table 3 and Supplementary Tables 12-13***). There was only weak evidence of association with methylation at CpGs located in regions including SNPs in the polygenic score ***(Supplementary Table 14).***

Alcohol types: For alcohol intake in the previous decade, the regression coefficients were the greatest for beer intake for 917 (65%) CpGs, for spirit intake for 309 (22%) CpGs and for wine intake for 188 (13%) CpGs, whereas these percentages were 27%, 32% and 41%, respectively, for lifetime alcohol intake (***Supplementary Table 16***).

Sensitivity analyses: Age, sex, smoking, BMI, country of birth, and cell composition were associated with methylation at many alcohol-associated CpG sites. Adjustment for smoking, but not BMI, made a substantial difference to the estimated coefficients; for the 'previous week' alcohol intake variable, 1,985 associations were observed when no adjustment for smoking was made. Adjusting the models for additional health-related variables or not adjusting for all white blood cell types made virtually no difference to the results (***Supplementary Material***).

Replication of previously reported associations using MCCS data

We examined the replication of the associations between alcohol intake and whole-blood DNA methylation previously reported in Liu et al. with $P<10^{-7}$. [19] Of the 518 associations, we replicated 403 (78%) at $P<0.05$, using the MCCS 'previous week' alcohol intake variable; 169 reached genome-wide significance $P<10^{-7}$ (***Table 4, Supplementary Table 5***). Replication was substantially higher for associations identified in European-ancestry individuals (335/363, 92%), compared with those of African ancestry (68/155, 44%).

Reversibility of associations

The 'current decade' alcohol consumption variable of the MCCS was classified into current, former, and never. Of the 1,414 CpGs considered, 280 were differentially methylated ($P<0.05$) between former and never drinkers, and 282 were differentially methylated ($P<0.05$) between former and

current drinkers, with only 2 overlapping CpGs. The reversibility coefficients, assessing, for former drinkers, the degree of return to the average methylation levels of never drinkers, were wide-ranging (median: 47%, IQR=19% to 79%) (***Supplementary Table 7***). In KORA, there was also substantial reversibility, as 332 CpGs were differentially methylated when comparing former to current drinkers. A total of 86 CpGs were differentially methylated in both datasets for the comparison of former to current drinkers, and 530 CpGs when pooling results from both cohorts (P<0.05); the most significant associations are presented in ***Table 5***.

Longitudinal associations

We further tested the 1,414 CpGs with cross-sectional associations for longitudinal associations using the MCCS and KORA. Repeated methylation measures and alcohol information were available a median of 11 and 7 years apart in the MCCS and KORA, respectively. Change in alcohol intake was associated with change in methylation (P<0.05) for 267 CpGs in the MCCS, and for 331 CpGs in KORA, with 92 overlapping associations. After pooling the results, we observed evidence of change over time for 513 CpG sites (***Supplementary Table 8***). The most statistically significant longitudinal associations are shown in ***Table 6***. Fewer associations were observed in the MCCS when no adjustment for baseline methylation levels was made (N=125, ***Supplementary Material***). The analyses of alcohol cessation (N=88 participants, 8%) and uptake (N=107, 10%) revealed 147 and 40 associations, respectively, overlapping little with the 513 identified CpGs (N=65 and N=19, respectively). CpG sites that showed stronger evidence of association in the longitudinal analysis appeared somewhat more reversible in the cross-sectional analysis ***(Supplementary Figure*** 5) and 245 CpGs appeared differentially methylated in both analyses (***Supplementary Table 10***).

Gene description and pathway analyses

The role of each of the 842 identified genes is summarized in ***Supplementary Table 20.*** The *gometh* function was applied separately to CpG sites associated with 'lifetime' (N=392) and 'last week' alcohol intake (N=1,318), and for associations with reversibility coefficients lower and greater than 50% (N=753 and N=661, respectively) (***Supplementary Table 11)***. The most significant KEGG pathways were for the 'last week' variable: '*Chronic myeloid leukemia*', '*Ribosome*', '*Glycosaminoglycan biosynthesis*'; lifetime variable: '*Regulation of actin cytoskeleton*', '*Biosynthesis of amino acids*', '*Platelet activation*'; persistent associations: '*Ribosome'*, '*Human cytomegalovirus infection*'; reversible associations: "*Cellular senescence*", "*MAPK signaling pathway*". ~~Nominally significant associations were also observed for pathways directly relevant to alcohol drinking such as '*GABAergic synapse*' or '*Glutamatergic synapse*'~~. There was strong evidence for enrichment of many GO pathways, and the strongest associations were observed for the 'last week' variable: '*negative regulation of gene expression*'; persistent associations: '*homophilic cell adhesion via plasma membrane adhesion molecules'*; reversible associations: "*translation*" **(*Supplementary Table 12***).

DISCUSSION

Our study identified 1,481 methylation sites associated with alcohol consumption, including 1,078 discovered in the MCCS and replicated in independent cohorts and 403 replicated from a previous large EWAS. An additional 513 CpGs discovered in the cross-sectional analysis were indirectly replicated in the longitudinal analysis. The findings using a less conservative significance threshold than $P<10^{-7}$ (i.e. 16,732 associations observed using FDR-adjusted P<0.05 for previous week variable) indicate that many more alcohol-associated CpGs are likely to exist across the genome. The

majority of CpG sites we identified were hypomethylated with increased alcohol intake. The overrepresentation of gene body probes in our analysis could be explained by an overall hypomethylation resulting from the potential disruption of one-carbon metabolism induced by increased alcohol consumption, and might not have specific functional implications. In contrast, aberrant hypomethylation of gene promoter CpGs, which were also overrepresented, might affect specific gene functions, but the functional relevance of our findings could not be tested in the present study.

Alcohol-related hypomethylation appears to be largely reversible upon alcohol cessation; this was inferred from three analyses. First, we observed substantially more and stronger associations with alcohol consumed in the last week than in the last decade or lifetime, indicating that the alcohol intake most relevant to DNA methylation was that closest to blood draw. Similar data were not available from other cohorts to replicate this finding. Second, the cross-sectional comparison of current and former to never drinkers revealed that the difference in terms of DNA methylation between former and current drinkers was on average half that between never and current drinkers, with wide-ranging estimates; we also identified CpG sites that were consistently differentially methylated in former compared with current drinkers in MCCS and KORA. Third, using longitudinal data taken several years apart (11 years in the MCCS and 7 years in KORA), we identified a set of 513 CpG sites that varied with change in alcohol consumption, and 245 of these corresponded to differentially methylated sites in the comparison of former to current drinkers (cross-sectional). The longitudinal analysis had less power due to a lower number of included participants and relatively small variation in drinking status over the periods considered, and because the variable reflecting changes in alcohol drinking may have been measured with error. In the MCCS, the assessment of

alcohol consumption was not identical at baseline (alcohol in the previous week), and follow-up (alcohol in the previous year). These findings taken together indicate a substantial degree of reversibility in the associations, which was not assessed by previous studies.

Another potential limitation of our study is residual confounding, most notably by smoking or white blood cell type composition, which are both strongly associated with alcohol drinking and DNA methylation.[8] We observed that many CpG sites associated with alcohol drinking were also associated with other factors such as smoking, white blood cell composition, BMI and other factors, which may indicate that these loci are very sensitive to the environment. Cell composition was estimated with the widely used Houseman algorithm modified by Jaffe and Irizarry [27, 40] and we did not assess sensitivity to the method used for deriving cell composition.[41] Although our adjustment for smoking was relatively comprehensive, our sensitivity analyses demonstrate that alcohol and smoking may exert joint influences on many CpGs across the genome.

We observed less substantial replication for CpGs discovered in individuals of African ancestry, which may indicate that alcohol-associated methylation changes are not generalizable to all human populations. Associations in individuals of African ancestry in the study by Liu et al. were discovered by pooling data from a smaller number of studies so might have been less replicable by nature.

The associations between alcohol consumption and methylation were stronger for people with genetic predisposition to consume less alcohol, as defined by a 13-SNP polygenic score. Several genes included in the polygenic score are involved in alcohol metabolism, for example those of the alcohol dehydrogenase family (*ADH1B, ADH1C,* and *ADH5)* and *GCKR* (glucokinase regulatory protein, involved in glucose metabolism), which could provide a biological explanation for the

interaction between the polygenic score and alcohol intake, given links between alcohol metabolism pathway and epigenetic mechanisms.[42]

Associations appeared weaker for smokers and men, consistent with the observation that these population subgroups tend to drink more alcohol in most cultures, perhaps due to being less susceptible to the harmful effects of alcohol.[43] Women have previously been reported to have slower alcohol metabolism than men[44] These findings should be confirmed by further studies. We included in the analysis participants who later developed cancer, which could give rise to collider bias when both DNA methylation and alcohol are associated with cancer risk.[45] We found no evidence of differences in associations by case/control status in our study. Further, that most discovered associations were replicated in independent cohorts of healthy participants with distinct ethnic origin is a strong testament that our findings were not driven by the inclusion of future cancer cases.

Our study is difficult to compare with studies other than that by Liu et al., due to differences in terms of study design, measurement of DNA methylation, and tissue analysed. Auta et al. [46] measured global methylation levels in liver and cerebellum tissue of rats and concluded that each tissue was characterised by a distinct methylation signature of alcohol consumption, but no results at specific genetic loci were reported. In a US study including over 600 samples, the association between alcohol dependence and epigenetic aging in various sets of blood, liver, and prefrontal cortex tissue was investigated.[47] The authors found some evidence of accelerated epigenetic aging although results were only partly consistent across tissues; significant associations were found in one set of blood samples and in liver. In a previous study using MCCS data, we also found accelerated epigenetic aging in the blood of heavy drinkers.[48] Lohoff and colleagues analysed data from multiple cohorts and tissues, and found that alcohol use disorder was associated with differential methylation at the

promoter region of *PCSK9*.[49] Interestingly, the association was observed in blood, brain, and liver, and *PCSK9* is a gene primarily expressed in the liver. However, *PCSK9* was not differentially methylated with respect to alcohol consumption in our study or that of Liu et al.

Another example of general lack of replication is a cross-sectional analysis of DNA methylation measures (Illumina 27K assay) in the blood of 81 drinkers and 81 age- and sex-matched controls, which reported 906 differentially-methylated CpG sites.[50] However, these findings were not independently replicated, and as reported by the authors, only two of these were identified at $P<10^{-7}$ in the study by Liu et al.;[19] these were also identified in our study. Tay and colleagues studied 1,287 adolescents for whom blood DNA methylation was measured using the HM450K assay [51] to identify associations with psychosocial stress and substance use, including alcohol consumption. There was weak evidence of an association between alcohol use and methylation of the *SPDEF* gene, the latter being associated with number of stressful life events. The *SPDEF* gene was not associated with alcohol consumption in our study.

The review by Zhang and Gelernter identified a number of small studies (N<200 participants) that examined associations of DNA methylation at candidate genes with various alcohol use or disorder outcomes.[18] Even for the largest of these studies, associations were generally weak and not replicated in our study, for example reported associations for the *MAOA* gene,[52] for *OPRM1*,[53] for *POMC* [54] at which an association in the opposite direction was observed in our study (cg09916783 promoter of POMC: $P=1.9x10^{-6}$ for last week intake). Several studies investigated the *SLC6A3/4* genes for which there was very weak evidence of association in our study ($P>4x10^{-4}$). A study of 383 heavy drinkers whose DNA methylation was measured in saliva samples using a custom assay targeting 6 CpG sites in the promoter region of the dopamine D2 receptor (*DRD2*) gene reported associations of DNA

methylation with activation of neural reward regions, as well as severity alcohol use disorder symptoms.[55] In our setting, we did not observe associations between *DRD2* DNA methylation and alcohol consumption, and only marginally significant associations with other *DRD* genes. Finally, Bruckmann et al. observed hypomethylation of *GDAP1* in alcohol-dependent patients, with a dose-response relationship by clinical severity;[56] interestingly, we also observed hypomethylation of *GDAP1* promoter at cg23779890 (last week intake, $P=3.1x10^{-9}$).

The newly discovered CpG sites with strongest evidence of association with alcohol consumption were all located in genes, including in regulatory regions. The 5 most strongly differentially methylated newly identified CpGs located in regulatory regions were in *SFRS13A* (Serine and Arginine Rich Splicing Factor 10, a splicing factor that in its dephosphorylated form acts as a general repressor of pre-mRNA splicing), *CPNE1* (encoding Copine 1, a calcium-dependent phospholipid-binding protein that plays a role in calcium-mediated intracellular processes), *SLC1A5* (Solute Carrier Family 1 Member 5, encoding a sodium-dependent neutral amino acid transporter), *FAM49A* (Family with sequence similarity 49, member A), and *PRELP* (Proline And Arginine Rich End Leucine Rich Repeat Protein, which encodes a leucine-rich repeat protein present in connective tissue extracellular matrix). We searched the literature for these genes and found little evidence of obvious links to alcohol metabolism, consumption, or alcohol-related diseases. We did not examine causality in our study; we hypothesise that if DNA methylation were the cause of alcohol drinking, it would likely be at a restricted number of loci involved in addiction mechanisms and alcohol metabolism. Our investigation of biological pathways using two enrichment databases did not clearly identify pathways or mechanisms known to play a key role in alcohol metabolism or alcohol-related diseases such as cancer, cardiovascular disease, and mental health and addiction pathologies. Further

analyses at the gene expression level, in particular for the most significant pathways we identified, are required to investigate to what extent the methylation changes associated with alcohol drinking identified in our study are involved in epigenetic regulation of genes or networks of genes. Our study identified methylation patterns in blood whose usefulness as biomarkers to predict or monitor alcohol-related disease should be validated in further studies.

Our study shows that alcohol consumption is associated with widespread changes in blood DNA methylation. These changes appear more pronounced in women, non-smokers, and individuals with lower genetic predisposition to drink alcohol, and are at least partially reversible.

ACKNOWLEDGEMENTS

MCCS: This work was supported by the Australian National Health and Medical Research Council (NHMRC) [grant 1088405]. MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases were ascertained through the Victorian Cancer Registry (VCR) and the Australian Cancer Database (Australian Institute of Health and Welfare). The nested case-control methylation studies were supported by the NHMRC grants 1011618, 1026892, 1027505, 1050198, 1043616 and 1074383. M.C.S. is an NHMRC Senior Research Fellow (1061177).

KORA: The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research has been supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. This work has received funding from the European Foundation for Alcohol Research (ERAB 2018 – EA1817). We thank all members of field staffs who were involved in the planning and conduct of the MONICA/KORA Augsburg studies.

LOLIPOP: The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z, 090532 & 098381) the NIHR (RP-PG-0407-10371), the NIHR Official Development Assistance (ODA, award 16/136/68), the European Union FP7 (EpiMigrant, 279143) and H2020 programs (iHealth-T2D, 643774). We acknowledge support of the MRC-PHE

Centre for Environment and Health, and the NIHR Health Protection Research Unit on Health Impact of Environmental Hazards. The work was carried out in part at the NIHR/Wellcome Trust Imperial Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the Imperial College Healthcare NHS Trust, the NHS, the NIHR or the Department of Health. We thank the participants and research staff who made the study possible. JC is supported by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research Investigator (STaR) Award (NMRC/STaR/0028/2017).

AUTHORS CONTRIBUTION

PAD, JEJ, EM, DRE, GGG, and RLM were responsible for the study concept and design. CHJ, JEJ, EM, DFS, LB, GS, CG, KHL, AP, JSK, MCS, DRE, MW, JCC, GGG, and RLM contributed to the acquisition of data. HJ, XW, CHJ, and JEJ contributed to the preparation of the data. PAD, RW, and BL performed the statistical analyses. PAD and RLM drafted the manuscript. All authors provided critical revision of the manuscript for important intellectual content. All authors critically reviewed content and approved final version for publication.

REFERENCES

1. Meissner A, Mikkelsen TS, Gu H, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008;454: 766-770.
2. Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005;6: 597-610.
3. van Dongen J, Nivard MG, Willemsen G, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. Nat Commun. 2016;7: 11115.
4. Boks MP, Derks EM, Weisenberger DJ, et al. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. PLoS One. 2009;4: e6767.
5. Fraga MF, Esteller M. Epigenetics and aging: the targets and the marks. Trends Genet. 2007;23: 413-418.
6. Geurts YM, Dugue PA, Joo JE, et al. Novel associations between blood DNA methylation and body mass index in middle-aged and older adults. Int J Obes (Lond). 2017.
7. Wahl S, Drong A, Lehne B, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541: 81-86.
8. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016;9: 436-447.
9. Shenker NS, Polidoro S, van Veldhoven K, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22: 843-851.
10. Dugue PA, Jung CH, Joo JE, et al. Smoking and blood DNA methylation: an epigenome-wide association study and assessment of reversibility. Epigenetics. 2019: 1-11.
11. Ezzati M, Riboli E. Behavioral and dietary risk factors for noncommunicable diseases. N Engl J Med. 2013;369: 954-964.
12. LoConte NK, Brewster AM, Kaur JS, Merrill JK, Alberg AJ. Alcohol and Cancer: A Statement of the American Society of Clinical Oncology. J Clin Oncol. 2018;36: 83-93.
13. Ron D, Barak S. Molecular mechanisms underlying alcohol-drinking behaviours. Nat Rev Neurosci. 2016;17: 576-591.
14. Giovannucci E. Alcohol, one-carbon metabolism, and colorectal cancer: recent insights from molecular studies. J Nutr. 2004;134: 2475S-2481S.
15. Berkel TD, Pandey SC. Emerging role of epigenetic mechanisms in alcohol addiction. Alcohol Clin Exp Res. 2017.
16. Cecil CA, Walton E, Smith RG, et al. DNA methylation and substance-use risk: a prospective, genome-wide study spanning gestation to adolescence. Transl Psychiatry. 2016;6: e976.
17. Cervera-Juanes R, Wilhelm LJ, Park B, Grant KA, Ferguson B. Alcohol-dose-dependent DNA methylation and expression in the nucleus accumbens identifies coordinated regulation of synaptic genes. Transl Psychiatry. 2017;7: e994.
18. Zhang H, Gelernter J. DNA methylation and alcohol use disorders: Progress and challenges. Am J Addict. 2016.
19. Liu C, Marioni RE, Hedman AK, et al. A DNA methylation biomarker of alcohol consumption. Mol Psychiatry. 2016.
20. Milne RL, Fletcher AS, MacInnis RJ, et al. Cohort Profile: The Melbourne Collaborative Cohort Study (Health 2020). Int J Epidemiol. 2017.
21. Baglietto L, Ponzi E, Haycock P, et al. DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. Int J Cancer. 2016.

22. Dugué PA, Brinkman MT, Milne RL, et al. Genome-wide measures of DNA methylation in peripheral blood and the risk of urothelial cell carcinoma: a prospective nested case-control study. Br J Cancer. 2016.
23. Dugué PA, English DR, MacInnis RJ, et al. Reliability of DNA methylation measures from dried blood spots and mononuclear cells using the HumanMethylation450k BeadArray. Sci Rep. 2016;6: 30317.
24. FitzGerald LM, Naeem H, Makalic E, et al. Genome-Wide Measures of Peripheral Blood Dna Methylation and Prostate Cancer Risk in a Prospective Nested Case-Control Study. Prostate. 2017.
25. Wong Doo N, Makalic E, Joo JE, et al. Global measures of peripheral blood-derived DNA methylation as a risk factor in the development of mature B-cell neoplasms. Epigenomics. 2015.
26. Jayasekara H, Macinnis RJ, Hodge AM, et al. Alcohol consumption for different periods in life, intake pattern over time and all-cause mortality. Journal of Public Health. 2015;37: 625-633.
27. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13: 86.
28. MacInnis RJ, English DR, Hopper JL, Haydon AM, Gertig DM, Giles GG. Body size and composition and colon cancer risk in men. Cancer Epidemiol Biomarkers Prev. 2004;13: 553-559.
29. Chiuve SE, Fung TT, Rimm EB, et al. Alternative dietary indices both strongly predict risk of chronic disease. J Nutr. 2012;142: 1009-1018.
30. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin. 2015;8: 6.
31. Wilson R, Wahl S, Pfeiffer L, et al. The dynamics of smoking-related disturbed methylation: a two time-point study of methylation change in smokers, non-smokers and former smokers. BMC Genomics. 2017;18: 805.
32. Chambers JC, Loh M, Lehne B, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol. 2015;3: 526-534.
33. Lehne B, Drong AW, Loh M, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol. 2015;16: 37.
34. Becker BJ, Wu M-J. The synthesis of regression slopes in meta-analysis. Statistical Science. 2007: 414-429.
35. Clarke TK, Adams MJ, Davies G, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). Mol Psychiatry. 2017;22: 1376-1384.
36. Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. Am J Epidemiol. 2005;162: 267-278.
37. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32: 286-288.
38. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28: 27-30.
39. Gene Ontology C. The Gene Ontology: enhancements for 2011. Nucleic Acids Res. 2012;40: D559-564.
40. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15: R31.
41. McGregor K, Bernatsky S, Colmegna I, et al. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. Genome Biol. 2016;17: 84.
42. Zakhari S. Alcohol metabolism and epigenetics changes. Alcohol Res. 2013;35: 6-16.

43. Marmot M. Inequality, deprivation and alcohol use. Addiction. 1997;92 Suppl 1: S13-20.
44. Frezza M, di Padova C, Pozzato G, Terpin M, Baraona E, Lieber CS. High blood alcohol levels in women. The role of decreased gastric alcohol dehydrogenase activity and first-pass metabolism. N Engl J Med. 1990;322: 95-99.
45. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology. 2003;14: 300-306.
46. Auta J, Zhang H, Pandey SC, Guidotti A. Chronic Alcohol Exposure Differentially Alters One-Carbon Metabolism in Rat Liver and Brain. Alcohol Clin Exp Res. 2017;41: 1105-1111.
47. Rosen AD, Robertson KD, Hlady RA, et al. DNA methylation age is accelerated in alcohol dependence. Transl Psychiatry. 2018;8: 182.
48. Fiorito G, Polidoro S, Dugue PA, et al. Social adversity and epigenetic aging: a multi-cohort study on socioeconomic differences in peripheral blood DNA methylation. Sci Rep. 2017;7: 16266.
49. Lohoff FW, Sorcher JL, Rosen AD, et al. Methylomic profiling and replication implicates deregulation of PCSK9 in alcohol use disorder. Mol Psychiatry. 2018;23: 1-11.
50. Chen J, Hutchison KE, Bryan AD, et al. Opposite Epigenetic Associations With Alcohol Use and Exercise Intervention. Front Psychiatry. 2018;9: 594.
51. Tay N, Macare C, Liu Y, et al. Allele-Specific Methylation of SPDEF: A Novel Moderator of Psychosocial Stress and Substance Abuse. Am J Psychiatry. 2019;176: 146-155.
52. Philibert RA, Gunter TD, Beach SR, Brody GH, Madan A. MAOA methylation is associated with nicotine and alcohol dependence in women. Am J Med Genet B Neuropsychiatr Genet. 2008;147B: 565-570.
53. Zhang H, Herman AI, Kranzler HR, Anton RF, Simen AA, Gelernter J. Hypermethylation of OPRM1 promoter region in European Americans with alcohol dependence. J Hum Genet. 2012;57: 670-675.
54. Muschler MA, Hillemacher T, Kraus C, Kornhuber J, Bleich S, Frieling H. DNA methylation of the POMC gene promoter is associated with craving in alcohol dependence. J Neural Transm (Vienna). 2010;117: 513-519.
55. Bidwell LC, Karoly HC, Thayer RE, et al. DRD2 promoter methylation and measures of alcohol reward: functional activation of reward circuits and clinical severity. Addict Biol. 2018.
56. Bruckmann C, Di Santo A, Karle KN, Batra A, Nieratschker V. Validation of differential GDAP1 DNA methylation in alcohol dependence and its potential function as a biomarker for disease severity and therapy outcome. Epigenetics. 2016;11: 456-463.

Table 1. Characteristics of participants in the Melbourne Collaborative Cohort Study (MCCS) at baseline and follow-up visits

	Cross-sectional analysis (N=5,606)		Longitudinal analysis Baseline data (N=1,088)		Longitudinal analysis Follow-up data (N=1,088)	
Age in years, median [IQR]	61 [54-65]		59 [51-64]		70 [63-76]	
Sex, male	3,793	68%	740	68%	740	68%
Country of birth						
AU/NZ/Other	3,744	67%	831	76%	831	76%
Greece	434	8%	44	4%	44	4%
Italy	819	15%	89	8%	89	8%
UK	609	11%	124	11%	124	11%
BMI (kg/m²), median [range]	26.9 [24.5-29.5]		26.3 [24.1-29.0]		26.8 [24.2-29.4]	
Smoking status						
Never	2,519	45%	549	50%	545	50%
Former ≥15 years ago	1,152	21%	230	21%	485	45%
Former <15 years ago	1,100	20%	196	18%		
Current <20 cig/day	322	6%	62	6%	58	5%
Current ≥20 cig/day	513	9%	51	5%		
Drinking status						
Lifetime abstainers	1,314	24%	209	20%	168	16%
Former drinkers	569	10%	94	9%	111	11%
Current drinkers	3,626	66%	759	71%	744	73%
Missing	97	2%	26	2%	65	6%
Median [IQR] for intake in last decade (g/day)	4.6 [0.0-21.5] missing: N=97		7.5 [0.0-21.3] missing: N=26			
Median [IQR] for intake in last week (g/day)	4.3 [0.0-18.7] missing: N=41		7.4 [0.0-19.0] missing: N=23			
Median [IQR] for lifetime intake (g/day)	8.4 [0.3-23.5] missing: N=115		8.8 [1.3-22.3] missing: N=42			
Median [IQR] for intake in last year (g/day)					8.9 [0.3-22.7] missing: N=70	
Median [IQR] for the difference between follow-up and baseline (last week) (g/day)					0.0 [-2.5 - 6.4]	
Median [IQR] for the difference between follow-up and baseline (lifetime) (g/day)					0.0 [-5.5 - 5.0]	
Alcohol uptake (non-drinkers at baseline who were drinkers at follow-up) (yes)					107	10%
Alcohol cessation (drinkers at baseline who were non-drinkers at follow-up) (yes)					88	8%

Table 2. The 15 most statistically significant novel associations (N=1,243) between alcohol intake and blood DNA methylation discovered in the MCCS and replication in external cohorts KORA and LOLIPOP

CpG	Chr	Position	Gene	Location	Mean	Discovery dataset (MCCS)						Replication datasets (KORA and LOLIPOP)						
						Lifetime		Current decade		Previous week		KORA		LOLIPOP		Pooled		
						Est.	P	Est.	P	Est.	P	Est.	P	Est.	P	Est.	P	Replicated
cg26856289	1	24307516	*SFRS13A*	TSS1500	0.34	-1.8	$4x10^{-17}$	-1.6	$5x10^{-18}$	-2.1	$3x10^{-25}$	-1.3	$4x10^{-5}$	-2.1	$3x10^{-19}$	-1.8	$2x10^{-22}$	yes
cg00422488	9	100747767	*ANP32B*	Body	0.11	-2.4	$2x10^{-12}$	-2.3	$3x10^{-13}$	-3.3	$7x10^{-23}$	-1.9	$4x10^{-4}$	-2.9	$8x10^{-17}$	-2.6	$3x10^{-19}$	yes
cg11826008	20	34249284	*CPNE1*	5'UTR	0.30	-1.7	$3x10^{-13}$	-1.3	$8x10^{-11}$	-2.2	$1x10^{-22}$	-0.9	$3x10^{-2}$	-1.4	$5x10^{-10}$	-1.3	$6x10^{-11}$	yes
cg03607573	11	93471889	*TAF1D*	Body	0.39	-1.5	$1x10^{-11}$	-1.4	$8x10^{-12}$	-2.1	$2x10^{-21}$	-1.7	$2x10^{-5}$	-2.1	$1x10^{-21}$	-2.0	$8x10^{-26}$	yes
cg20625334	2	152991620	*STAM2*	Body	0.58	-1.9	$9x10^{-13}$	-1.6	$2x10^{-12}$	-2.4	$6x10^{-21}$	-1.7	$3x10^{-4}$	-1.9	$3x10^{-14}$	-1.9	$3x10^{-17}$	yes
cg23684449	16	46919194	*GPT2*	Body	0.53	-1.0	$5x10^{-9}$	-0.9	$3x10^{-9}$	-1.5	$8x10^{-21}$	-0.8	$2x10^{-4}$	-0.9	$5x10^{-10}$	-0.9	$4x10^{-13}$	yes
cg15114651	19	47289410	*SLC1A5*	TSS1500	0.56	-1.0	$3x10^{-15}$	-0.8	$6x10^{-12}$	-1.1	$2x10^{-20}$	-1.0	$8x10^{-8}$	-1.0	$2x10^{-11}$	-1.0	$7x10^{-18}$	yes
cg06829760	2	16845412	*FAM49A*	5'UTR	0.48	-1.3	$5x10^{-13}$	-1.2	$2x10^{-14}$	-1.5	$6x10^{-20}$	-0.7	$4x10^{-3}$	-1.1	$3x10^{-10}$	-0.9	$1x10^{-11}$	yes
cg26841068	1	203456691	*PRELP*	3'UTR	0.41	-1.2	$1x10^{-12}$	-1.0	$3x10^{-12}$	-1.4	$6x10^{-20}$	-1.1	$6x10^{-6}$	-1.4	$7x10^{-15}$	-1.3	$3x10^{-19}$	yes
cg05288253	16	89552259	*ANKRD11*	5'UTR	0.59	-1.1	$4x10^{-12}$	-1.1	$3x10^{-15}$	-1.4	$1x10^{-19}$	-1.1	$5x10^{-8}$	-1.1	$1x10^{-12}$	-1.1	$3x10^{-19}$	yes
cg13408712	11	129817591	*PRDM10*	TSS200	0.61	-0.8	$3x10^{-10}$	-0.9	$3x10^{-13}$	-1.2	$4x10^{-19}$	-0.9	$4x10^{-7}$	-0.4	$2x10^{-3}$	-0.6	$5x10^{-8}$	yes
cg04367503	19	13121571	*NFIX*	Body	0.52	-3.1	$4x10^{-13}$	-2.8	$2x10^{-13}$	-3.7	$6x10^{-19}$	-2.6	$4x10^{-5}$	-3.7	$2x10^{-14}$	-3.3	$9x10^{-18}$	yes
cg08899105	15	93450671	*CHD2*	Body	0.47	-0.9	$4x10^{-8}$	-0.9	$3x10^{-9}$	-1.4	$1x10^{-18}$	-0.2	$4x10^{-1}$	-0.8	$3x10^{-9}$	-0.7	$1x10^{-8}$	yes
cg04261072	13	79977499	*RBM26*	Body	0.49	-1.4	$2x10^{-9}$	-1.3	$7x10^{-11}$	-1.9	$2x10^{-18}$	-0.4	$4x10^{-1}$	-1.1	$6x10^{-8}$	-1.0	$2x10^{-7}$	yes
cg07577824	20	17943403	*SNX5*	Body	0.38	-1.1	$2x10^{-8}$	-1.1	$1x10^{-9}$	-1.7	$3x10^{-18}$	-0.7	$2x10^{-2}$	-1.3	$1x10^{-7}$	-1.0	$2x10^{-8}$	yes

Abbreviations: Chr.: Chromosome; Est.: regression coefficients from linear mixed regression models (MCCS) or linear regression models (KORA and LOLIPOP).
All models were adjusted by fitting fixed effects for age, sex, smoking status, BMI, country of birth, sample type and white blood cell composition (percentage of CD4+ T cells, CD8+ T cells, B cells, NK cells, monocytes and granulocytes, estimated using the Houseman algorithm), and batch effects.
Est. pooled and P pooled are results from the fixed-effects meta-analysis of results from KORA and LOLIPOP.
Regression coefficients are given for intakes in grams per day and multiplied by 1000.
The full results are presented in the Supplementary file.

Table 3: Interaction between alcohol intake and other factors in association with DNA methylation changes for 1,414 alcohol-related CpG sites

Variable	**Positive interaction (P<0.05)**	**Negative interaction (P<0.05)**	**P****	**Genes annotated to CpGs for which strongest evidence of interaction observed (P<0.001)** Full results in Supplementary Table 13
Sex (F vs M)	200	5	$6x10^{-29}$	*MIR548F5, SFRS13B, MIR548F5, MGAT5B, NSD1, ABI3, SARM1, NFIX, SC65, NTF3, ADRA2A, ESRP2, ZNF532*
Age	53	11	0.07	*SLC7A11*
Smoking score*	7	159	$6x10^{-20}$	*TLR9, CPT1A, ILKAP, MYB, SLC7A11, HDAC1, GPR39, AKR1A1, RPP21, C6orf227*
Body mass index	7	65	0.003	*BCAN, JAK1, TOP1MT*
Case status (Case vs. control)	35	16	1	-
PRS for alcohol consumption	8	156	$2x10^{-19}$	*RPL6, DOPEY2, NPM1, MSI2, VPS54*

[a] Smoking was coded as a pseudo-continuous score: 0: Never, 1: Former, ≥15 years ago, 2: Former <15 years ago, 3: Current <20 cig/day, 4: Current ≥20 cig/day.

** *P*-value from test of binomial proportions that the greatest number (e.g. 200 for sex) is greater than 35 (number of associations expected by chance). P-values were calculated assuming independence between methylation values at the 1,414 CpGs.

Table 4. Replication of the strongest associations ($P<10^{-15}$) reported in Liu et al., Mol. Psychiatry, 2016 using MCCS data.

CpG	Chr.	Position	Gene	Location	Discovery set ancestry	Est. Liu	P Liu	P for alcohol intake in previous week (MCCS)	Replicated
cg11610002	3	126080368			African	-0.10	$1x10^{-28}$	0.06	
cg15061231	3	186353575			African	-0.09	$9x10^{-28}$	0.6	
cg22524735	8	144779094	*BREA2*	TSS200	African	-0.03	$1x10^{-24}$	0.8	
cg25729907	3	146263345	*PLSCR1*	TSS1500	African	-0.05	$4x10^{-24}$	0.7	
cg25120484	3	197186156			African	-0.05	$2x10^{-22}$	1	
cg07710247	18	55268669	*NARS*	3'UTR	African	-0.04	$1x10^{-20}$	0.8	
cg05538701	6	116866658	*FAM26D*	5'UTR	African	-0.04	$1x10^{-19}$	0.5	
cg02583484	12	54677008	*HNRNPA1*	Body	white	-0.39	$2x10^{-19}$	$2x10^{-29}$	yes
cg13729116	4	1859262	*LETM1*	TSS1500	white	-0.18	$7x10^{-18}$	$1x10^{-6}$	yes
cg13057576	14	64692154	*SYNE2*	Body	African	-0.04	$9x10^{-18}$	0.1	
cg07470207	8	132828746			African	0.04	$1x10^{-17}$	0.5	
cg09935388	1	92947588	*GFI1*	Body	African	-0.17	$2x10^{-17}$	$2x10^{-6}$	yes
cg05593667	6	35490744			white	-0.25	$4x10^{-16}$	$3x10^{-7}$	yes
cg02003183	14	103415882	*CDC42BPB*	Body	African	0.10	$7x10^{-16}$	0.04	yes
cg13620705	17	73559492	*LLGL2*	Body	African	-0.09	$9x10^{-16}$	0.2	

Abbreviations: Chr.: chromosome; Est. Liu: regression coefficient reported in the study by Liu et al., Mol. Psychiatry, 2016

The full results are presented in the Supplementary file.

Table 5. Reversibility of associations (cross-sectional Current vs Former vs Never) for the fifteen most significant associations in the MCCS EWAS for the 'current decade' alcohol intake variable.

			Current decade intake (g/day)		Former vs. never		Current vs. never		
CpG	**Chr.**	**Position**	**Est.**	**P**	**Est.**	**P**	**Est.**	**P**	**Reversibility coefficient[a]**
cg06690548	4	139162808	-6.6	$7x10^{-70}$	-0.01	0.8	-0.12	$2x10^{-8}$	95%
cg14476101	1	120255992	-3.4	$6x10^{-34}$	-0.02	0.4	-0.10	$2x10^{-10}$	81%
cg12825509	3	185648568	-2.4	$2x10^{-26}$	-0.05	0.009	-0.07	$2x10^{-8}$	31%
cg02711608	19	47287964	-2.1	$2x10^{-24}$	-0.01	0.5	-0.04	$8x10^{-5}$	76%
cg18120259	6	43894639	-1.9	$1x10^{-20}$	-0.04	0.009	-0.05	$6x10^{-6}$	14%
cg18336453	6	43082296	-1.3	$2x10^{-19}$	-0.03	0.01	-0.04	$5x10^{-8}$	32%
cg19693031	1	145441552	-2.5	$5x10^{-19}$	-0.02	0.3	-0.06	$3x10^{-4}$	61%
cg11376147	11	57261198	-1.5	$1x10^{-18}$	-0.04	0.001	-0.05	$3x10^{-7}$	7%
cg26856289	1	24307516	-1.6	$5x10^{-18}$	-0.01	0.5	-0.05	$4x10^{-6}$	78%
cg17058475	11	68607737	-3.6	$9x10^{-18}$	-0.07	0.03	-0.16	$5x10^{-13}$	56%
cg16246545	1	120255941	-1.7	$1x10^{-17}$	0.00	0.9	-0.04	$1x10^{-4}$	104%
cg06644515	1	173834831	-1.6	$6x10^{-17}$	0.00	0.9	-0.04	$3x10^{-4}$	106%
cg02583484	12	54677008	-1.4	$1x10^{-16}$	-0.02	0.2	-0.05	$3x10^{-7}$	62%
cg15804598	17	43224418	-1.1	$1x10^{-15}$	-0.02	0.1	-0.04	$2x10^{-7}$	52%
cg00252472	6	150739173	-2.7	$1x10^{-15}$	-0.03	0.3	-0.08	$5x10^{-6}$	63%

Abbreviations: Chr.: Chromosome; Est.: regression coefficients from linear mixed regression models (MCCS)
All models were adjusted by fitting fixed effects for age, sex, smoking status, BMI, country of birth, sample type and white blood cell composition (percentage of CD4+ T cells, CD8+ T cells, B cells, NK cells, monocytes and granulocytes, estimated using the Houseman algorithm), and batch effects.
Regression coefficients are given for intakes in grams per day and multiplied by 1000.
The full results are presented in the Supplementary file.

[a] Reversibility coefficient, expressed as a percentage defined as: coefficient ('former' compared with 'current') / coefficient ('never' compared with 'current').

Table 6. Longitudinal associations ($P<10^{-5}$) assessed in the MCCS and in KORA (data collected 11 and 7 years apart, respectively).

					Cross-sectional data		Longitudinal data					
					MCCS (previous week)		MCCS		KORA		Pooled[a]	
CpG	Chr.	Position	Gene	Location	Est.	P	Est.	P	Est.	P	Est.	P
cg06690548	4	139162808	*SLC7A11*	Body	-7.9	$1x10^{-86}$	-4.7	$2x10^{-5}$	-5.4	$7x10^{-11}$	-5.2	$5x10^{-15}$
cg18120259	6	43894639	*LOC100132354*	Body	-2.4	$4x10^{-27}$	-1.5	$5x10^{-3}$	-1.5	$5x10^{-8}$	-1.5	$8x10^{-10}$
cg02711608	19	47287964	*SLC1A5*	1stExon	-2.6	$6x10^{-30}$	-2.0	$4x10^{-4}$	-1.4	$1x10^{-6}$	-1.5	$3x10^{-9}$
cg14476101	1	120255992	*PHGDH*	Body	-4.0	$2x10^{-39}$	-2.9	$1x10^{-4}$	-1.8	$3x10^{-6}$	-2.1	$3x10^{-9}$
cg16246545	1	120255941	*PHGDH*	Body	-2.2	$4x10^{-23}$	-1.7	$2x10^{-3}$	-1.4	$7x10^{-7}$	-1.4	$4x10^{-9}$
cg11376147	11	57261198	*SLC43A1*	Body	-2.1	$2x10^{-29}$	-1.0	$8x10^{-3}$	-1.4	$1x10^{-6}$	-1.3	$3x10^{-8}$
cg20732160	3	48590040	*PFKFB4*	Body	-1.7	$3x10^{-16}$	-1.6	$3x10^{-3}$	-1.2	$1x10^{-5}$	-1.3	$1x10^{-7}$
cg03068497	7	30635838	*GARS*	Body	-3.0	$3x10^{-14}$	-2.0	$2x10^{-2}$	-2.4	$4x10^{-6}$	-2.3	$2x10^{-7}$
cg07626482	19	47289503	*SLC1A5*	TSS1500	-1.4	$4x10^{-20}$	-1.2	$2x10^{-3}$	-1.0	$4x10^{-5}$	-1.1	$3x10^{-7}$
cg13526915	14	24164078			-2.1	$8x10^{-13}$	-1.2	$8x10^{-2}$	-1.7	$1x10^{-5}$	-1.6	$3x10^{-6}$
cg14756878	2	12568736			-1.2	$3x10^{-9}$	-1.4	$2x10^{-3}$	-0.9	$4x10^{-4}$	-1.0	$3x10^{-6}$
cg04460609	4	16532808	*LDB2*	Body	-2.2	$3x10^{-19}$	-1.2	$2x10^{-2}$	-1.3	$1x10^{-4}$	-1.3	$5x10^{-6}$
cg21626848	17	39969267	*SC65*	TSS1500	-1.6	$1x10^{-16}$	-1.8	$2x10^{-4}$	-0.8	$2x10^{-3}$	-1.1	$8x10^{-6}$
cg03533472	16	46919112	*GPT2*	Body	-2.2	$1x10^{-15}$	-2.5	$9x10^{-4}$	-1.7	$2x10^{-3}$	-2.0	$9x10^{-6}$

Abbreviations: Chr.: Chromosome; Est.: regression coefficients from linear mixed regression models (MCCS) or linear regression models (KORA).
All models were adjusted by fitting fixed effects for age, sex, smoking status, BMI, country of birth, sample type and white blood cell composition (percentage of CD4+ T cells, CD8+ T cells, B cells, NK cells, monocytes and granulocytes, estimated using the Houseman algorithm), and batch effects.
Regression coefficients are given for intakes in grams per day and multiplied by 1000.
The full results are presented in the Supplementary file.

[a] Est. and P refer to the pooling of results from the MCCS and KORA using fixed-effects meta-analysis.