A DNA Methylation Biomarker of Alcohol Consumption

## Supplementary Methods

#### DNA Methylation Profiling and Processing

#### Genomic DNA was extracted from peripheral whole blood samples or isolated CD14+ monocytes. In all samples, the genomic DNA was bisulfite-converted for methylation measurement using the Illumina Infinium HumanMethylation450 BeadChip (San Diego, CA). The methylated probe intensity and total probe intensities were extracted using the Illumina Genome Studio (version 2011.1) with the methylation module (version 1.9.0). Preprocessing of the methylated signal (M) and unmethylated signal (U) was conducted in each cohort. Methylation beta-value was defined as. Normalization was performed (see Table below) in each cohort to reduce technical artefacts (see Description of Study Samples).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Study** | **Ancestry** | **Tissue** | **Unrelated  or Family** | **Normalization  method** | **Normalization**  **software** |
| ARIC | AA | Whole blood | Unrelated | SWAN |  |
| CHS | EA, AA | Whole blood | Unrelated | SWAN | minfi/R |
| EPIC-Norfolk | EA | Whole blood | Unrelated | SWAN | minfi |
| FHS | EA | Whole blood | Family | DASEN | wateRmelon/R |
| GTP | EA | Whole blood | Unrelated | Quantile  normalization | limma/R |
| Inchianti | EA | Whole blood | Unrelated | DASEN | wateRmelon/R |
| KORA F4 | EA | Whole blood | Unrelated | BMIQ |  |
| LBC1936 | EA | Whole blood | Unrelated | Regression based | R |
| MESA | EA, AA, HA | CD14+ Monocytes | Unrelated | Smooth quantile  normalization | LUMI |
| NAS | EA | Whole blood | Unrelated | BMIQ | wateRmelon/R |
| PIVUS | EA | Whole blood | Unrelated | Quantile  normalisation | R |
| RS | EA | Whole blood | Unrelated | DASEN | wateRmelon/R |
| TwinsUK | EA | Whole blood | Family | BMIQ | R |

## EA, European ancestry; AA, African ancestry; HA, Hispanic samples

SWAN, Subset-quantile within array normalization; BMIQ, beta-mixture quantile method; DASEN, a data-driven approach to preprocessing Illumina 450K; BMIQ, beta-mixture quantile method**;** LUMI**,** the lumi Bioconductor package was designed to process the Illumina microarray data.

**DNA Methylation Raw Data Availability**

|  |  |
| --- | --- |
| **Cohort Name** | **DNA Methylation Raw Data Availability** |
| ARIC | Not yet |
| CHS | Not yet |
| EPIC-Norfolk | Not yet |
| FHS | dbGaP Study Accession: phs000724.v5.p10 |
| InCHIANTI | Not yet |
| KORA F4 | Not yet |
| LBC1936 | The European Genome-phenome Archive (EGA; https://www.ebi.ac.uk/ega/home)  under accession number EGAS00001000910. |
| NAS | Not yet |
| PIVUS | Not yet |
| RS | Not yet |
| TwinsUK | Not yet |
| GTP | GSE number: GSE72680 |
| MESA | The NCBI Gene Expression Omnibus and is accessible through GEO  Series accession number (GSE56046) |

**Genotyping and Genotype Imputation in FHS, KORA F4 and PIVUS**

In FHS, the Affymetrix 500K mapping array and the Affymetrix 50K gene-focused MIP array were used for genotyping. Several quality control procedures were applied to genotypes(1). MACH (version 1.0.15) was used to impute all autosomal SNPs genotyped in the 1000 Genomes Project using the publicly available reference set (Sanger version). The single nucleotide polymorphisms (SNPs) that were within 100 kb of the significant CpGs were selected (minor allele frequency >5% and imputation quality score (r2 > 0.5) from 1000G imputation data.

KORA F4 samples were genotyped using the Affymetrix Axiom chip array. Genotypes were called with the Affymetrix software and were annotated to NCBI build 37. Individuals with call rates <97% were excluded, as were mismatches of phenotypic and genetic gender and heterozygosity rates +/-5 SD from the mean. Before imputation SNPs with call rates <98%, HWE p-values<5E-6, and minor allele frequency <1% were excluded. Pre-phasing was performed with SHAPEIT v2. Imputation was performed with IMPUTE v2.3.0 using the 1000G phase1 (v3) reference panel.

In PIVUS, individuals were genotyped using the Illumina OmniExpress and Illumina Metabochip microarrays. Prior to imputation, quality control was performed. Exclusion of samples were performed based on the following criteria: genotype call rate <95%; heterozygosity >3 SD; gender discordance; duplicated samples; identity-by-descent match; and ethnic outliers. Monomorphic SNPs; or SNPs with Hardy-Weinberg equilibrium *P*<1x10-6; genotype call rate<0.99 (SNPs with MAF<5%) or <0.95 (SNPs with MAF≥5%); MAF<1% were excluded from analysis. Genotype data were imputed to 1000G (version: March 2012) using Impute v.2.2.2(2). To perform analysis in R, genotype probabilities (from Impute) were transformed to posterior mean genotypes (MACH format). Only SNPs with a minor allele frequency (MAF) >5%, INFO (from imputation) >0.8 and within 100kb of the CpG probe were included in analysis.

**Gene Expression Profiling in FHS and KORA F4**

In FHS, the gene expression profiling measured in peripheral blood samples from a total of 5,626 participants from the Offspring cohort at examination eight (2005-2008) and the Third Generation at examination two (2008-2011)(3). Fasting peripheral whole blood samples (2.5ml) were collected in PAXgene™ tubes (PreAnalytiX, Hombrechtikon, Switzerland). RNA expression profiling was conducted using the Affymetrix Human Exon Array ST 1.0 (Affymetrix, Inc., Santa Clara, CA). The expression values for ~ 18,000 transcripts were obtained from the total 1.2 million core probe sets. Quality control procedures for transcripts have previously been described(4, 5). All gene expression data from the FHS used herein are available online in dbGaP (<http://www.ncbi.nlm.nih.gov/>gap; accession number phs000007). The blood cell proportions were calculated from DNA methylation data using the method of Houseman et al.(6).

In KORA, gene expression profiling was performed using the Illumina Human HT-12 v3 Expression BeadChip as described elsewhere(7). Briefly, total RNA was extracted from whole blood under fasting conditions using the PAXgene Blood miRNA Kit (Quiagen). Purity and integrity of the RNA was assessed on the Agilent Bioanalyzer with the 6000 Nano LabChip reagent set (Agilent Technologies, Germany). 500 ng of RNA was reverse transcribed and biotin-UTP-labeled into cRNA using the Illumina TotalPrep-96 RNA Amp Kit (Ambion, Darmstadt, Germany). A total of 3000 ng of cRNA was hybridized to the Illumina Human HT-12 v3 Expression BeadChip, followed by washing steps as described in the Illumina protocol. The raw intensity data were exported from Illumina software Genome Studio to R and processed (log transformation and quantile normalization(8)) using the lumi:1.12.4 package(9) from the Bioconductor open source software (http://www.bioconductor.org). The blood cell proportions were calculated from DNA methylation data using the method of Houseman et al.(6).

## 

## Supplementary Results

All epigenome association analyses were performed with the alcohol phenotype as the independent variable and the methylation proportion of individual CpG sites as the response variable, adjusting for covariates (see Methods). All meta-analysis used the inverse variance weighted random effects model.

**Epigenome-wide Association of Alcohol Intake in European Whole Blood Samples**

The meta-analysis that employed the inverse variance weighted random effects model yielded 363 CpGs at *P*<0.05/475000=1x10-7 (**Figure 2A**, **Table 2**, and **Supplementary Table 2**) for association between the continuous alcohol phenotype (g per day) and methylation levels using all individuals of European ancestry (EA) (n=9,643, **Table 1**). Additional loci at *P*<1x10-4 are displayed in **Supplementary Table 5**. The majority of the 363 CpGs (~87%) associated with alcohol intake at *P*<1x10-7 displayed an inverse relationship between DNA methylation levels and alcohol consumption (**Figure 3B**). In meta-analysis of EA drinkers, 258 significant CpGs were identified when smoking was not a covariate (n=7,245) (**Supplementary Table 5**, **Supplementary Figures 10 and 11**). In meta-analysis of EA drinkers, 157 significant CpGs were identified when smoking was included as an additional covariate (n=7,241) (**Supplementary Table 5**, **Supplementary Figures 10 and 11**). Notably a few of the CpGs were no longer significant after smoking adjustment (e.g. CpGs in *AHRR*(10, 11) on Chr 5) have been previously identified as markers of DNA methylation in relation to smoking (**Supplementary Table 5**).

Fewer significant CpGs were identified in the analysis of the categorical alcohol phenotype which used the non-drinker category as the control and compared light (0<to ≤28 g per day in men and 0 to ≤14 g per day in women), at-risk (28 to <42 g per day in men and 14 to <28 g per day in women), and heavy drinkers (g per day ≥42 in men and ≥28 in women) versus non-drinkers (g per day=0) (**Supplementary Tables 9**). The majority of the significant CpGs identified in the analysis of the categorical alcohol phenotype were significant or nominally significant in the analysis of the continuous alcohol phenotype (Supplementary Tables 5). Analysis of light-drinkers (>0 to ≤28 grams per day in men and > 0 to ≤14 grams per day in women; n=5,539) versus non-drinkers (n=3,334) did not yield any significantly associated CpGs (*P*<1x10-7). Analysis of at-risk (28 to <42 grams per day in men and 14 to <28 grams per day in women, n=1,005) and heavy drinkers (≥42 grams per day men and ≥28 grams per day in women, n=694) versus non-drinkers yielded one (at-risk drinkers) and 11 (heavy drinkers) significant CpGs (*P*<1x10-7) (**Supplementary Table 9 and Supplementary Figure 4**). Additional CpG sites (*P*<1x10-4) were displayed in **Supplementary Tables 5** for the continuous alcohol phenotype and **Supplementary Tables 12** for the categorical alcohol phenotypes.

**Epigenome-wide Association of Alcohol Intake in African Whole Blood Samples**

For the continuous alcohol phenotype (g per day), meta-analysis (inverse variance random effects model) of all individuals of the whole blood AA cohorts (n=2,423, **Table 1**) yielded 165 CpGs at *P*<1x10-7 (**Supplementary Table 3 and Supplementary Figure 2A**). About 62% of these significant CpGs were inversely correlated with alcohol intake (**Supplementary Figure 2B**). Meta-analysis of drinkers-only yielded 30 significant CpGs (*P*<1x10-7) (n=703) and 35 significant CpGs when smoking was included as an additional covariate (n=551) (**Supplementary Table 6**, **Supplementary Figures 12 and 13**). For categorical alcohol phenotypes, we found one CpG with *P*<1x10-7 for light drinkers (n=241) versus non-drinkers (n=1,687), 18 CpGs with *P*<1x10-7 for at-risk drinkers (n=69) versus non-drinkers, and 96 CpGs with *P*<1x10-7 for heavy drinkers (n=421) versus non-drinkers (**Supplementary Table 6**). Additional CpG sites (*P*<1x10-4) are displayed in **Supplementary Tables 6 and 13** for continuous alcohol and categorical alcohol phenotypes, respectively.

**Association Analysis of Alcohol on Methylation in Monocytes**

In DNA samples derived from CD14+ monocytes (**Table 1**, n=1,251), we found 62 CpGs significantly (*P*<1x10-7) differentially methylated in response to continuous alcohol intake (n=1,251) (**Supplementary Tables 4 and Supplementary Figure 3A**). 76% of these 62 CpGs were inversely correlated with alcohol intake (**Supplementary Figure 3B**). Among drinkers (n=560), the association analysis yielded 24 CpGs with *P*<1x10-7 when smoking was not adjusted, and also 24 CpGs *P*<1x10-7 when smoking was adjusted (**Supplementary Table 7**, **Supplementary Figures 14 and 15**). In categorical analysis, four CpGs were found at *P*<1x10-7 from heavy drinkers versus non-drinkers. There was no CpGs at *P*<1x10-7 in analysis of light drinkers versus non-drinkers (**Supplementary Table 11**). Additional CpG sites (*P*<1x10-4) are displayed in **Supplementary Tables 7 and 14** for the continuous alcohol and categorical alcohol phenotypes, respectively.

**Epigenome-wide Association of Alcohol Intake in Pooled Samples**

For association analysis between the continuous alcohol phenotype (g per day) and methylation measurement, we performed meta-analysis with the inverse variance weighted random effects model for about 438,215 CpGs that passed filters (see Methods). In meta-analysis of all pooled samples (n total=13,317) including whole blood derived DNA samples of EA (n=9,643) and AA (n=2,423) ancestries, and monocyte derived DNA samples of mixed ancestries (n=1,251), we identified 88 CpGs with *P*<1x10-7; in meta-analysis of drinkers (n total=8,508) including 7,245 samples of EA whole blood, 703 samples of AA whole blood and 560 samples of monocyte samples, we identified only 11 CpGs with *P*<1x10-7 when smoking was not adjusted; in meta-analysis of drinkers (n total=8,353) including 7,241 samples of EA whole blood, 551 samples of AA whole blood and 560 samples of monocyte samples, we identified only 8 CpGs with *P*<1x10-7 when smoking was adjusted as an additional variable (**Supplementary Table 28** and **Supplementary Figure** **16**).

For association analysis between the categorical alcohol phenotypes and methylation levels, we performed meta-analysis with the inverse variance weighted random effects model for pooled samples that included DNA derived from the EA and AA whole blood samples and from the monocyte samples of mixed ancestries. The meta-analysis of association of light drinkers (n=6,224) versus non-drinkers (n=5,712) did not identify any CpGs at *P*<1x10-7; the meta-analysis of at-risk drinkers (1,071) versus non-drinkers identified one CpGs at *P*<1x10-7; and the meta-analysis of heavy drinkers (n=1,166) versus non-drinkers identified 52 CpGs at *P*<1x10-7 (**Supplementary Table 29 and Supplementary Figure 17**).

**Sensitivity tests to investigate if cardiovascular diseases (CVD) and cancers confound the relationship between DNA methylation and alcohol consumption**

To investigate if cardiovascular diseases confound the relationship between DNA methylation and alcohol consumption, we performed a sensitivity test in the Framingham Heart Study (FHS). In total, 174 prevalent CVD cases were found in 2,427 individuals with DNA methylation data. The sensitivity test compared the regression coefficients and p-values between methylation (outcome) and alcohol intake (independent variable) in an EWAS model that did not adjust for CVD status to a second model that adjusted for CVD (other covariates remained the same in both models). For genome-wide methylation probes, we found no difference in the regression coefficients (Pearson correlation= 0.9996) and –log10 (p-values) (Pearson correlation=0.9992) between the two models. The model that did not adjust for CVD yielded 82 CpGs at *P*<1e-7 while the model adjusting for CVD yielded 83 CpGs at *P*<1e-7; 82 CpGs overlapped between the models. For these top CpGs, the Pearson correlation was 0.99997 for the regression coefficients and 0.99993 for the –log10 (*P*-values). Therefore, cardiovascular disease status was unlikely to confound the relationship between alcohol consumption and DNA methylation in whole blood.

To investigate if cancers confound the relationship between DNA methylation and alcohol consumption, we also performed a sensitivity test adjusting for all types of prevalent cancer in FHS. In total, 654 prevalent cancer cases were found in 2,427 individuals with DNA methylation data. The sensitivity test compared the regression coefficients and p-values between methylation (outcome) and alcohol intake (independent variable) in an EWAS model that did not adjust for cancer status to a second model that adjusted for cancer (other covariates remained the same in both models). For genome-wide methylation, we found no difference in the regression coefficients (Pearson correlation=0.998) and –log10 (p-values) (Pearson correlation=0.967) between the two models. The two models yielded 82 CpGs at *P*<1e-7, and 81 CpGs overlapped between the two models. For the top CpGs from both models the Pearson correlation was 0.993 for the regression coefficients, and 0.992 for the –log10 (p-values). Therefore, cancer status was unlikely to confound the relationship between alcohol consumption and DNA methylation in whole blood

**Description of Study Samples**

The CHARGE methylation-alcohol working group consists of 13 cohorts: ARIC, CHS, EPIC-Norfolk, FHS, GTP, InCHIANTI, KORA F4, LBC, MESA, NAS, PIVUS, RS, and TwinsUK (**Table 1**). All cohorts used Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA)(12) to measure DNA methylation levels. Cohort descriptions (**Table 1**), DNA methylation and quality control procedures, methods for statistical analyses for all participating studies are described below. Informed consent for genetic studies was obtained from all subjects. The protocol for the study was approved by the institutional review boards of all individual cohorts.

*Acknowledgements, general*: Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. This work was supported in part by the Intramural Research Program of the National Institutes of Health (NIH): National Heart Lung and Blood Institute, National Institute on Aging and the National Institute of Environmental Health Sciences.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health

***1.1 ARIC***

**The Atherosclerosis Risk in Communities study**(13) is a population-based, prospective cohort study designed to investigate cardiovascular disease and its risk factors (<http://www.cscc.unc.edu/aric/>). Sponsored by the National Heart, Lung, and Blood Institute (NHLBI), ARIC comprises 15,792 men and women of Caucasian, African-American, and other race/ethnicity, aged 45-64 at baseline (1987-89), chosen by probability sampling in four U.S. communities. Participants were reexamined every three years between 1987 and 1998 (four examinations in total). Annual follow-up has been conducted to monitor clinical events. The follow-up consists of two components 1) the Cohort Component, which follows study participants; and 2) Community Surveillance Component, which assesses cardiovascular disease in the four recruitment communities. The protocol for the study was approved by the institutional review boards of all centers, and all participants provided written informed consent that included consent for genetic studies.

*DNA methylation and quality control*: The DNA methylation was measured for 2,950 African-American samples. All covariates were recorded or measured at Visit 2. BMI was calculated using the formula weight (kilograms) divided by height (meters) squared. White blood cell count (WBC) was assessed for a subset of participants at Visit 2 (n=187) by automated particle counters within 24 hours after venipuncture in the local hospital hematology laboratory. The reliability coefficient for the WBC count measurement was greater than 0.96. The measured WBC differentials were then used in the imputation of differential WBCs for the remaining subjects (see below for description). Among 2,950 samples, 2,003 had both DNA methylation and alcohol consumption phenotypes.

Genomic DNA was extracted from peripheral blood leukocyte samples using the Gentra Puregene Blood Kit (Qiagen; Valencia, CA, USA) according to the manufacturer’s instructions (www.qiagen.com). Bisulfite conversion of 1 µg genomic DNA was performed using the EZ-96 DNA Methylation Kit (Deep Well Format) (Zymo Research; Irvine, CA, USA) according to the manufacturer's instructions (www.zymoresearch.com). Bisulfite conversion efficiency was determined by PCR amplification of the converted DNA before proceeding with methylation analyses on the Illumina platform using Zymo Research’s Universal Methylated Human DNA Standard and Control Primers.

DNA was isolated from white blood cells as per the standard DNA extraction procedure (Autopure LS, Qiagen). 500 ng of extracted DNA was bisulfite-modified using the EZ DNA Methylation kit (Zymo Research, D5004) following the manufacturer’s instructions.The Illumina Infinium HumanMethylation450 Beadchip array (HM450) was used to measure DNA methylation (Illumina Inc.; San Diego, CA, USA). Raw methylation data were extracted using Illumina GenomeStudio software (version 2011.1, Methylation module 1.9.0). The methylation score for each CpG was represented as a beta value according to the fluorescent intensity ratio of methylated to overall signals. Background subtraction was conducted with the GenomeStudio software using built-in negative control bead types on the array.

Subset quantile within array normalization (SWAN)(14) was used to normalize DNA methylation beta-values. Further, in this study, we conducted all analyses at the single probe level, and therefore, any differences in probe type should not strongly influence the results.

A total of 2,950 study participants had HM450 data available for further quality control analyses. We removed poor-quality samples with pass rate <99%, that is, if the sample had at least 1% of CpG sites with detection *P* value >0.01 or missing (n=32), indicative of lower DNA quality or incomplete bisulfite conversion. At the target level, we flagged poor-quality CpG sites with average detection *P* value >0.01, and calculated the percentage of samples having detection *P* value >0.01 for each autosomal and X chromosome CpG site. There were 9,399 autosomal and X chromosomal markers where >1% of samples showed detection *P* value >0.01, and these sites were excluded. In addition, we filtered 370 CpG sites on the Y chromosome with average detection *P* value >0.01, and after SWAN normalization, a total of 473,788 CpG sites remained for analysis.

*Phenotype definition*: Alcohol consumption was accessed by self-administered questionnaire. The total number of consumption of beers, wine and spirit was collected. A continuous alcohol consumption phenotype, “grams per day”, was defined as alcohol consumed per day (grams/day) from consumption of beers, wine and liquor by using the following conversion factor: one bottle of beer was equivalent to 14 grams of ethanol; one glass of wine was equivalent to 15 grams of ethanol; and one drink of liquor was equivalent to 14 grams of ethanol. A categorical variable including four categories [non-drinkers (grams per day=0); light drinkers (0<grams per day ≤28 in men and ≤14 in women); at-risk drinkers (28 <grams per day<42 in men and 14 <grams per day <28 in women); and Heavy drinkers (grams per day ≥42 men and ≥28 in women)] was defined using “grams per day”.

*Statistical analysis*: The association analysis used the alcohol phenotype as the independent variable and DNA methylation proportion as outcome, adjusting for age, gender, BMI, total white blood cell count, and estimated white blood cell proportions.

*Acknowledgements*: The Atherosclerosis Risk in Communities study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, HHSN268201100012C, R01HG006292 and R01HL129132). Funding support for ‘Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium’ was provided by the National Institutes of Health (NIH) through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419).

The authors thank the staff and participants of the ARIC study.

***1.2 CHS***

**The Cardiovascular Health Study** is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥65 years conducted across four field centers(15). The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility list. Subsequently, an additional predominantly African-American cohort of 687 persons was enrolled. CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease.

*DNA methylation profiling and Quality Control*: DNA methylation was measured on a randomly selected subset of 200 African-Americans participants from study year 5 with available DNA. In addition, two hundred participants of European ancestry were selected from participants without presence of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack at study baseline or lack of available DNA at study year 5.

Methylation measurements were performed at the Institute for Translational Genomics and Population Sciences at the Harbor-UCLA Medical Center Institute for Translational Genomics and Population Sciences using the Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA). Quality control was performed in the minfi R package(16, 17) (version 1.12.0, <http://www.bioconductor.org/packages/release/bioc/html/minfi.html>). Methylation values were normalized using the SWAN quantile normalization method(14). Samples with low median intensities of below 10.5 (log2) across the methylated and unmethylated channels, samples with a proportion of probes falling detection of greater than 0.5%, samples with QC probes falling greater than 3 standard deviation from the mean, sex-check mismatches, or failed concordance with prior genotyping were removed. In total, 11 samples were removed for sample QC resulting in a sample of 191 European-ancestry and 198 African-American samples. The white blood cell proportions were estimated from the methylation data using the Houseman method(18).

*Phenotype definition:* The Alcohol consumption in CHS was assessed by self-report at each yearly visit.  Questions regarding drink frequency and number of drinks on one occasion were used to calculate the number of cans of beer, shots of liquor, and glasses of wine consumed per week.  These quantities, in turn, were converted to grams per day using the following conversions: one beer (12 oz) ~14 gm ethanol, or one glass of wine (5 oz) ~ 15 gm ethanol, or one drink of spirit (1.5 oz 80 proof alcohol) ~ 14 gm ethanol.

*Statistical analysis*: Analyses were stratified by race and all analyses were adjusted for age, gender, BMI, total white blood cell count, study clinic and estimated white blood cell proportions. Surrogate variable analysis was performed to identify batch effects(19). No surrogate variables were found to be associated (p-value <0.01) with the outcome among European-ancestry participant and one surrogate variable was used in the analyses of African-Americans. African-American analyses were additionally adjusted for two genetic principal components. All association analyses were performed in R using linear models with DNA methylation beta values as the outcome.

*Acknowledgements:* The CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL092111, R01HL105756, R01HL103612, R01HL111089, R01HL116747 and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA) as well as Laughlin Family, Alpha Phi Foundation, and Locke Charitable Foundation. A full list of principal CHS investigators and institutions can be found at [CHS-NHLBI.org](http://chs-nhlbi.org/). The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

***1.3 EPIC-Norfolk***

## The European Prospective Investigation into Cancer-Norfolk study(20) is a large multi-center follow-up cohort primarily looking at the connection between diet, lifesyle factors and cancer. Between 1993 and 1997, this study enrolled more than 25,000 community-based men and women (40-79 years old) in and around the city of Norwich (Norfolk, UK). The study design and follow up of participants has been reported previously(20). All participants provided written informed consent for genetic study. The study complies with the principles of the Declaration of Helsinki. The ethical approval was given by the Norfolk Local Research Ethics Committee and the East Norfolk and Waveney NHS Research Governance Committee.

## *DNA methylation profiling and quality control*: DNA was isolated from white blood cells as per the standard DNA extraction procedure (Autopure LS, Qiagen). 500 ng of extracted DNA was bisulfite-modified using the EZ DNA Methylation kit (Zymo Research, D5004) following the manufacturer’s instructions. DNA methylome profiling was carried out using the Illumina Infinium HumanMethylation450 (HM450) BeadChip assay. The minfi R package was used to preprocess, normalize (SWAN), and calculate beta values.

## *Alcohol phenotype*: Personal medical history was assessed using the question in the Health and Lifestyle Questionnaire. Participants were asked about their weekly frequency and amount of alcohol consumption of different alcoholic beverages over the past 12 months. The alcohol consumption data were summarized as units per week, and then converted to grams/day (one unit of alcohol in the UK is defined as 7.9 grams(21)). The categorical phenotype is the same as the FHS definition.

## *Statistical analysis*: All association analyses were performed in R using linear models with DNA methylation beta values as the outcome. Analyses were analyses were adjusted for age, gender, and estimated white blood cell proportions, as well as plate number and position.

### *Acknowledgements*: the EPIC-Norfolk study is supported by program grants from the Medical Research Council (MRC) [G9502233; G0401527] and Cancer Research UK [C864/A8257]. The generation and management of the Illumina 450K methylation array data in this cohort is supported through the MRC Cambridge initiative in metabolomic science [MR/L00002/1]. CEE, KKO and NJW are supported by MRC program grants [MC\_UU\_12015/1 and MC\_UU\_12015/2].

***1.4 FHS***

**The Framingham Heart Study is** a population-based, prospective study that began in 1948 with the recruitment of an original cohort(22) of 5,209 men and women from Framingham, MA. The offspring cohort was recruited in 1971, including 5,124 offspring and spouses of offspring of the FHS Original cohort(23). Participants underwent examinations every four years (except eight years between the first and the second examinations) to collect demographic and clinical measures and medical history(23). DNA methylation was measured in 2,846 FHS Offspring cohort participants who attended the eighth examination cycle from 2005-2008. Of the 2,846 individuals, 2,427 had both measurements on alcohol consumption and methylation. All participants included in this study provided written consent for genetic research.

*DNA Methylation profiling and quality control*: DNA methylation was measured on 2,846 participants who gave consent for genetic studies. Peripheral whole blood samples were collected from these participants at the eighth examination. Buffy coat fractions were obtained and genomic DNA was extracted using the Gentra Puregene DNA extraction kit (Qiagen, Venlo, Netherlands). Bisulfite conversion of genomic DNA was performed with the EZ DNA Methylation Kit (Zymo Research, Irvine, CA). After whole genome amplification, fragmentation, array hybridization, and single-base pair extension, DNA methylation was quantified in two laboratories. The first laboratory analyzed 576 samples (denoted as S1) that were previously selected for a cardiovascular disease (CVD) case-control study(24). The second laboratory analyzed 2, 270 samples (denoted as S2) from the remainder of the Offspring cohort.

In FHS, alcohol consumption was measured by asking participants questions on drinking of beer, wine and spirits (80 proof) at each examination cycle. The questions included, “On average, how many drinks did you have per week over course of last year?” A drink was defined as having 12 oz of beer (in bottle, can, or glass), 5 oz of wine, or 1.5 oz of liquor. One beer (12 oz) ~14 gm ethanol, one glass of wine (5 oz) ~ 15 gm ethanol, and one drink of spirit (1.5 oz 80 proof alcohol) ~ 14 gm ethanol. The continuous variable was “grams per day”. The categorical phenotype is “drink\_cat” was defined as having the following categories: non-drinkers (grams per day=0); light drinkers (0<grams per day ≤28 in men and ≤14 in women); at-risk drinkers (28 <grams per day<42 in men and 14 <grams per day <28 in women); and Heavy drinkers (grams per day ≥42 men and ≥28 in women).

Due to possible laboratory effects on methylation arrays, data was processed within each laboratory using the DASEN methodology implemented in the *wateRmelon* R package(25) (version 3.0.2, http://www.bioconductor.org/packages/release/bioc/html/wateRmelon.html). The first step of data processing was to adjust for methylated and unmethylated fluorescent intensities (M and U) and technical variations. The second step was to perform quantile normalization of the M and U values with consideration of two types of probe technologies. Beta values were then derived as the ratio of methylated probe intensity to the overall intensity. For quality control purposes, we retrieved overlapped SNPs (Number of SNPs=65) from previous genotyping effort. We excluded samples with a probe missing rate >1% (n=45), poor SNP matching to the 65 SNP control probe locations (n=79), and outliers by multi-dimensional scaling techniques (n=73). At the probe level, we excluded those with missing rate >20% at p<0.01 (n=466 from S1 1 and n=366 from S2), as well as probes previously identified to map to multiple locations3 or have an underlying SNP (minor allele frequency >5% in European ancestry (EUR) 1000 genomes project data) at the CpG site or <=10 bp of the single base extension (n=42,251). After quality control, 2,651 (522 in S1 and 2,129 in S2) samples with ~ 440,000 CpG probes were used for subsequent analyses.

*Statistical analysis***:** The association analysis was performed using non-transformed DNA methylation beta values (DNAm) as the outcome variables. Alcohol consumption was analyzed either as continuous (g per day) or categorical variables (light drinker, at risk drinkers, and heavy drinkers) in the regression model. Covariates included age, sex, BMI, white blood cell differential count, and surrogate variables that were calculated to account for hidden confounders. All association analysis was conducted using linear mixed regression models with the R function lme () to account for familial correlation (<http://cran.r-project.org/>).

*Acknowledgements*: The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD. The analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the Center for Information Technology, National Institutes of Health, Bethesda, MD. Dr. Liang’s was supported by P30 DK46200 and partially supported by P30 DK46200.

***1.5 GTP***

**The Grady Trauma Project** (http://gradytraumaproject.com/) is a population-based prospective study that began in 2005 with the goal of investigating the demographic and trauma exposure characteristics of a population of urban, low-income, and predominantly African-American men and women(26). Participants were recruited from the general medical clinics of Grady Memorial Hospital, a publicly funded hospital that serves economically disadvantaged individuals in Atlanta, Georgia. All participants in this study have given informed consent for genetic studies. The study was approved by the institutional review boards of Emory University School of Medicine and Grady Memorial Hospital.

DNA was extracted from whole blood at the Max Planck Institute in Munich using the Gentra Puregene Kit (Qiagen). Genomic DNA was then bisulfite converted using the Zymo EZ-96 DNA Methylation Kit (Zymo Research). DNA methylation was measured according to the instructions of the manufacturer (Illumina Inc.). For each CpG site and individual, we collected two data points: M (the total methylated signal), and U (the total unmethylated signal). We set points with a detection p-value greater than 0.001 to missing, and removed individual samples from analysis if they were outliers in a hierarchical clustering analysis or had 1) a mean total signal less than half of the median of the overall mean signal or 2,000 arbitrary units, or 2) a missingness rate above 5%. Similarly, we removed from analysis CpG sites with a missingness rate above 10%. We used limma to perform quantile normalization of the combined signal data, and subsequently computed β-values for each individual at each CpG site as the total methylated signal divided by the total signal: .

*Phenotype definition*: Alcohol consumption was measured from the Alcohol Use Disorders Identification Test (AUDIT). From the larger dataset, 231 had both measurements on alcohol consumption and methylation. Table 1 summarizes the characteristics of these 231 individuals. In order to convert the data from the AUDIT survey (“*How often do you have a drink with alcohol*?*”*) to number of grams of alcohol consumed per day, we used the following conversion: for those endorsing 0-2 drinks per day, we assigned 14.33 grams/day (though we classified individuals endorsing “never” in a separate category using the question, “*In the year you drank the most, how often did you have a drink with alcohol?”*)*;* for those endorsing 3-4 drinks per day, we assigned 50.155 grams/day; for those endorsing 5-6 drinks per day, we assigned 78.815 grams/day; for those endorsing 7-9 drinks per day, we assigned 114.64 grams/day; and for 10+ drinks per day, we assigned 143.3 grams/day.

*Statistical analysis*: The statistical analysis using the methylation proportion as the outcome and alcohol phenotype as the independent variable, adjusting for age, sex, BMI, Chip ID, Chip row, and 6 cell type proportions (estimated via the Houseman method).

*Acknowledgements*: This work was primarily supported by the National Institute of Mental Health (MH096764 and MH071537 to K.J.R.). Support was also received from Emory and Grady Memorial Hospital General Clinical Research Center, the Max Planck Society, the Behrens-Weise Foundation, NIH National Centers for Research Resources (M01RR00039), and the National Center for Advancing Translational Sciences of the NIH (UL1TR000454).

***1.6 InCHIANTI***

The **InCHIANTI** study is a population-based epidemiological study aimed at evaluating the factors that influence mobility in the older population living in the Chianti region in Tuscany, Italy(27). Briefly, 1,616 residents were selected from the population registry of Greve in Chianti (a rural area; 11,709 residents with 19.3% of the population greater than 65 years of age), and Bagno a Ripoli (Antella village near Florence; 4,704 inhabitants, with 20.3% greater than 65 years of age). The participation rate was 90% (n=1453), and the subjects ranged between 21-102 years of age. All participants signed an informed participation consent.

*DNA methylation profiling and quality control*: Genomic DNA of 514 individuals was collected from peripheral whole blood samples. The extracted DNA was bisulfite-modified using the EZ DNA Methylation kit (Zymo Research, D5004) following the manufacturer’s instructions. The DNA methylation profiling used the Illumina Infinium HumanMethylation450 Beadchip array (HM450). In this study, 496 individuals had both alcohol consumption and DNA methylation. The study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review.

Quality control filtering and normalization was accomplished using the wateRmelon package(25). Markers were excluded if the bead count was less than 3 in ≥5% of samples. Samples and markers were also excluded if ≥5% of detection p-values were greater than 0.01. A background adjustment and quantile normalization were applied to the filtered data set; the selected method normalizes both methylated and unmethylated probes as well as type I and II assays separately. Locations were annotated using the FDb.InfiniumMethylation.hg19 database. Methylation markers on the X and Y chromosome, as well as markers with potentially cross-reactive probes and probes that may be polymorphic in European populations (allele frequency ≥.01)(28) were excluded from analyses.

*Phenotype definition*: Alcohol consumption was assessed using a 236-item food frequency questionnaire designed for the European Prospective Investigation into Cancer and nutrition (EPIC) study(29). The questionnaire was administered by a trained interviewer. For alcohol, the question included the intake frequency of wine, liquor and beer for the past year. Specific software created for the EPIC study was used to transform the data to grams of alcohol per day.

*Statistical analysis*: the association analysis was performed using non-transformed DNA methylation beta values (DNAm) as the outcome variables. Alcohol consumption was analyzed either as continuous (g per day) or categorical variables (light drinker, at risk drinkers, and heavy drinkers) in the regression model. Covariates included age, sex, BMI, and technical covariates including experimental batch and white blood cell differential count. Association analysis was conducted using linear regression models. All statistical analyses were performed using R software (<http://cran.r-project.org/>).

*Acknowledgements*: The InCHIANTI study baseline (1998-2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336). Our research utilized the high-performance computational capabilities of the Biowulf system at the NIH, Bethesda, MD.

***1.7 KORA F4***

**The Cooperative Health Research in the Region of Augsburg** **study** is an independent population-based cohort from the region of Augsburg, Southern Germany(30). Whole blood samples were obtained from the KORA F4 survey (examination 2006-2008), a seven-year follow-up study of the KORA S4 cohort. Participants gave written informed consent for genetic study and the study was approved by the local ethics committee (Bayerische Landesärztekammer). In total, 1,799 subjects had both methylation and alcohol consumption data.

*DNA methylation profiling and quality control*: Whole blood genomic DNA was bisulfite converted using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA) according to the manufacturer’s procedure, with the alternative incubation conditions recommended when using the Illumina Infinium Methylation Assay. Raw methylation data were extracted using the Illumina GenomeStudio software (version 2011.1, Methylation module 1.9.0). Preprocessing was performed with R (version 3.0.1). Probes with signals from less than three functional beads, and probes with a detection p‑value >0.01 were defined as low-confidence probes. Probes that covered SNPs (MAF in Europeans >5%) were excluded from the data set. A color bias adjustment was performed with the R package lumi (version 2.12.0) by smooth quantile normalization and background correction based on negative control probes present on the Infinium HumanMethylation BeadChip. This was performed separately for the two color channels and chips. β-values corresponding to low-confidence probes were set to missing. A 95% call rate threshold was applied on samples and CpG sites. Beta-mixture quantile normalization (BMIQ) was applied by using the R package wateRmelon, version 1.0.3. Because KORA F4 samples were processed on 20 96-well plates in 9 batches, plate and batch effects were investigated by principle component analysis and eigenR2 analysis. Association between methylation and alcohol intake was determined using linear models with methylation beta values as the outcome variables, and alcohol as independent variable, additionally including age, sex, BMI, smoking category, estimated white blood cell proportion (Houseman et al. 2012) and plate information as covariates.

*Phenotype definition*: Alcohol consumption was measured by questionnaires on drinking of beer, light beer, wine, spirits (80 proof), and mixed drinks during the previous weekend (Saturday and Sunday) and the day before the last working day. The continuous phenotype, “grams of alcohol consumed per day,” is calculated by using the following converters between drinks and grams: one beer was equivalent to 40.0 g alcohol**;** one light beer to 22.0 g alcohol**;** one alcohol-free beer to 3.0 g alcohol**; one bottle of** wine to 100.0 g alcohol; **and one glass of** spirits (0.02L) to 6.2 g alcohol. Grams per day were calculated as: (alcoholweekend + 5 \* alcoholworking day )/7.

*Statistical analysis*: The association between DNA methylation level as the outcome and alcohol intake as the independent variable was performed using the linear regression models, adjusting for age, BMI, smoking, blood cells counts, and technical covariates (analytic plate, chip position on plate, and columns and rows).

*Acknowledgements*: 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 was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreements n° 261433 (BioSHaRE) and n° 603288 (sysVASC). The Geman Diabetes Center was supported by the Ministry of Science and Research of the State of North Rhine-Westphalia (MIWF NRW) and the German Federal Ministry of Health (BMG). This study was supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V.).

***1.8 LBC***

**The Lothian Birth Cohort** of 1936 is a longitudinal study of ageing(31-33). It derives from the Scottish Mental Survey of 1947, when nearly all 11 year-old children in Scotland completed a test of general cognitive ability(33). Survivors living in the Lothian area of Scotland were recruited in later life at mean age 70 (n=1,091). Follow-up has taken place at ages 70, 73, and 76. Collected data include genetic information, longitudinal epigenetic information, longitudinal brain imaging, and numerous blood biomarkers, anthropomorphic, and lifestyle measures. Post QC, DNA methylation data were available for 920 participants at age 70. Data on alcohol consumption and methylation were available on 920 participants. All subjects gave written, informed consent for genetic studies.

*DNA methylation profiling and quality control*: Detailed information about the collection and QC steps undertaken on the LBC methylation data have been reported previously(34). Briefly, DNA from whole blood samples was bisulfite converted using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA) according to the manufacturer’s procedure and then used for measuring DNA methylation. Background correction was performed and QC was used to remove probes with a low detection rate, low quality (manual inspection), low call rate, and samples with a poor match between genotypes and SNP control probes, and incorrect predicted sex. To account for technical variability in the measurement of the methylation CpGs, we adjusted the probes for plate, array, position on the chip, and hybridisation date (all treated as fixed effects).

*Phenotype definition*: In LBC, alcohol consumption was measured by asking questions on drinking of beer, wine, and spirits at the same wave as blood collection for methylation. Four specific questions were asked: (1) Do you ever drink alcohol (yes/no) (2) How many times a week do you drink? (3) What do you normally drink? (4) How many glasses/pints on average? From these data, we estimated weekly alcohol consumption in units. This was then converted into g per day using the formula 1 unit of alcohol = 7.9 grams i.e. units per week \*7.9/7. A glass of wine (175ml), beer (568ml), or measure of spirit (25ml) corresponded to 2, 2, and 1 units respectively. Conversion of grams per day into the drink\_cat variable was identical to the FHS analysis.

*Statistical analysis*: DNA methylation proportion was used as the outcome and alcohol intake was used as the independent variable in linear regression models. Covariates included age, sex, BMI, smoking, blood cells counts, and technical covariates (analytic plate, chip position on plate, and columns and rows).

*Acknowledgements*: Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Methylation typing was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. REM, PMV, and IJD are members of the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE). CCACE is supported by funding from the BBSRC, the Economic and Social Research Council (ESRC), the Medical Research Council (MRC), and the University of Edinburgh as part of the cross-council Lifelong Health and Wellbeing initiative (MR/K026992/1).

***1.9 MESA***

**The Multi-Ethnic Study of Atherosclerosis (MESA)** was designed to investigate the prevalence, correlates, and progression of subclinical cardiovascular disease in a population cohort of 6,814 participants. Since its inception in 2000, five clinic visits collected extensive clinical, socio-demographic, lifestyle, behavior, laboratory, nutrition, and medication data(35). The study protocol was approved by the Institutional Review Board at each site. All participants signed informed consent for genetic research.

*DNA methylation profiling*: DNA was measured in purified (CD14+) monocyte samples from the April 2010 to February 2012 examination (exam 5) of 1,264 randomly selected MESA participants [55–94 years old, Caucasian (47%), African American (21%), Hispanic (32%) and female (51%)] from four MESA field centers (Baltimore, MD; Forsyth County, NC; New York, NY; and St. Paul, MN) as previously described(36). DNA was bisulfite converted using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA) according to the manufacturer’s procedure. Illumina HumanMethylation450 BeadChips and HiScan reader were used to perform the epigenome-wide methylation analysis. The EZ-96 DNA Methylation™ Kit (Zymo Research, Orange, CA, USA) was used for bisulfate conversation with 1 µg of input DNA (at 45 µl). An amount of 4 µl of bisulfite-converted DNA was used for DNA methylation assays, following the Illumina Infinium HD Methylation Protocol. This consisted of a whole-genome amplification step followed by enzymatic end-point fragmentation, precipitation and resuspension. The resuspended samples were hybridized on HumanMethylation 450 BeadChips at 48°C for 16 h. The individual samples were assigned to the BeadChips and to chip position, using the same sampling scheme as that for the expression BeadChips.

Data pre-processing and QC analyses were performed in R (http://www.r-project.org/) using Bioconductor (http://www.bioconductor.org/) packages. Bead-level methylation data were summarized in GenomeStudio. Because a two-channel system and both Infinium I and II assays were used, normalization was performed in several steps using the lumi package. Smooth quantile normalization was used to adjust for color bias. Next, the data were background adjusted by subtracting the median intensity value of the negative control probes. Lastly, data were normalized across all samples by standard quantile normalization applied to the bead-type intensities and combined across Infinium I and II assays and both colors. QC measures included checks for sex and race/ethnicity mismatches, and outlier identification by multidimensional scaling plots. To estimate residual sample contamination for data analysis, we generated separate enrichment scores for neutrophils, B cells, T cells, monocytes, and natural killer cells. We implemented a Gene Set Enrichment Analysis(37) as previously described(36) to calculate the enrichment scores using the gene signature of each blood cell type from previously defined lists(38). To remove technical error in methylation levels associated with batch effects across the multiple chips, positional effects of the sample on the chip, and residual sample contamination with non-monocyte cell types, we adjusted methylation values for chip, sample position on the chip, and estimated residual sample contamination with neutrophils, B cells, T cells, monocytes, and natural killer cells. The final methylation value for each methylation probe was computed as the beta-value, essentially the proportion of the methylated to the total intensity

*Phenotype definition*: Alcohol consumption was measured at exam 5 by asking questions on drinking of beer, wine, and liquor or mixed drinks. The questions included, “Do you presently drink alcoholic beverages?”, “How many glasses of red wine do you usually have per week?”, “How many glasses of white wine do you usually have per week?”, “How many cans, bottles, or glasses of beer do you usually have per week?”, and “How many drinks of liquor or mixed drinks do you usually have per week?”. The continuous phenotype, “grams of alcohol consumed per day,” is calculated by using the following converters between drinks and grams: one beer ~14 gm ethanol, one glass of wine ~ 15 gm ethanol, and one drink of liquor or mixed drink ~ 14 gm ethanol.

*Statistical analysis*: the association analysis was performed using non-transformed DNA methylation beta values (DNAm) as the outcome variables. Alcohol consumption was analyzed either as continuous (g per day) or categorical variables (light drinker, at risk drinkers, and heavy drinkers) in the regression model. Covariates included age, sex, BMI, current smoking status.

*Acknowledgements*: MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-001079, UL1-TR-000040, and DK063491. The MESA Epigenomics & Transcriptomics Study was funded by NIA grant 1R01HL101250-01 to Wake Forest University Health Sciences.

***1.10 NAS***

**The Normative Aging Study (NAS)** is a longitudinal study which studies the effects of aging on various health issues of healthy male volunteers. It was initiated by the Department of Veterans Affairs in Boston in the 1960s. Participants in the study have undergone medical examinations every three to five years, also answering questions about behaviors affecting health, in particular stress, smoking, and cardiac health. Participants provided written informed consent for genetic research, and the VA Boston Healthcare System Institutional Review Board approved the study.

*DNA methylation profiling and quality control*:Buffy coat DNA was bisulfite converted (EZ-96 DNA Methylation Kit,Zymo Research, Orange, CA, USA), and DNA methylation was measured using the Infinium HumanMethylation450 BeadChip. Samples were organized using a stratified randomization in blocks for Illumina BeadChips (sets of 12) and analytic plates (sets of 8 BeadChips) so that multiple samples from the same individual were always on the same BeadChip, stratified by quartile of age, with statistical checks to ensure balance across chips and plates of age, MMSE, and CHD. Quality control samples included replicate pairs as well as a single sample that was run within and between chips and plates to characterize batch effects. All analytic plates were run consecutively by the same technician and were processed and read on the same scanner at the genomics core facility at Northwestern University.

Quality control procedures included the empirical removal of failing samples/probes using the pfilter command from package wateRmelon (excluded 15 samples with >1% of probes with detection p-values > 0.05), and subsequently excluded 949 probes with >1% of samples (after excluding samples dropped above) with detection p-value >0.05. Probes were further flagged for removal based on design/annotation, resulting in: 1) exclusion of 65 genotyping probes; 2) exclusion of 3,091 CpG sites; and 3) exclusion of 3,688 probes with a SNP in the last 10 bases with minor allele frequency >0.01 in CEU dataset. As some of these exclusions overlapped with the set excluded by pfilter, the resulting dataset included 477,927 probes on the first available 450k sample from 623 participants.

Data preprocessing included three additional steps: 1) Background correction using the out-of-band (noob) method of Triche et al 2013(39); 2) Dye-bias adjustment from methylumi; and 3) Probe-type adjustment using BMIQ (40) as implemented in the wateRmelon package.

*Phenotype definitions***:** In NAS, the NHS2 style FFQs was used to assess average alcohol consumption. It was a self-administered validated semi-quantitative food frequency questionnaire adapted from the questionnaire used in the Nurses’ Health Study. A glass was defined as a 12-oz (355-ml) bottle of beer or a 4-oz (118-ml) glass of wine, and a shot of liquor. For each beverage, participants reported their usual average consumption in the preceding year, with nine response categories. The alcohol intake was determined by multiplying the consumption of each beverage by its ethanol content (12.8 g for beer, 11.3g for light beer, 11.0 g for wine, and 14.0 g for liquor) and summing all beverages.

*Association analysis***:** To examine the association between DNA methylation levels and alcohol intake we used linear regression models. As outcome we used non-transformed DNA methylation beta values and we adjusted for age, BMI, smoking, blood cells counts, and technical covariates (analytic plate, chip position on plate, and columns and rows).

*Acknowledgements*: The present work on the US Department of Veterans Affairs (VA) Normative Aging Study has been supported by funding from the U.S. National Institute of Environmental Health Sciences (NIEHS) (R01ES015172, R01ES021733). The VA Normative Aging Study is supported by the Cooperative Studies Program/ERIC, US Department of Veterans Affairs, and is a research component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). Additional support to the VA Normative Aging Study was provided by the US Department of Agriculture, Agricultural Research Service (contract 53-K06-510). The views expressed in this paper are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs.

***1.11 PIVUS***

**The Prospective Investigation of the Vasculature in Uppsala Seniors** **study** is a community-based prospective cohort of participants from Uppsala, Sweden(41). Baseline medical examination (including anthropometric measurements) was performed at 70 years of age (between years 2001 and 2004). All participants gave written informed consent for genetic study and the Ethics Committee of Uppsala University approved the study protocols.

*DNA methylation profiling and quality control*: Blood for DNA methylation assay were collected at the 70 years visit. Genomic DNA was extracted from peripheral blood leukocyte samples and bisulfite conversion of 500 ng genomic DNA was performed using the EZ-96 DNA Methylation Gold Kit from Zymo Research Product. The equivalent of approximately 200ng of bisulfite converted DNA, was removed, evaporated to a volume of <4μl, and used for methylation profiling according to the protocol from the supplier. The results were analyzed with GenomeStudio 2011.1 from Illumina Inc. After exclusion of replicates a total of 1,002 study participants had methylation data available for quality control procedures. Three samples were excluded based on poor bisulfite conversion efficiency, 12 samples due to low pass rate of CpG sites (<98.5% with a detection P-value> 0.01) and a further six samples based on low SNP genotype match (>1 SNP mismatches) between genotypes from the methylation array and Omni/Metabochip genotyping chips, leaving n=981 samples. Following removal of participants with abnormal leukocyte cell counts (>10x109 cells/L; n=14) and those with missing data on alcohol usage, a final set of 818 individuals were used in alcohol analysis. The signal intensities for the methylated and unmethylated state were quantile normalized for each probe type separately, and beta values were calculated.

*Phenotype definition*: Alcohol data was collected at the medical examination visit. The conversions were based on a database from the Swedish Food Agency listing "all" drinks and food, with very precise measures per type of drinks, but we have converted these measures into similar broad categories as in the other studies: Beer (4.5 volume %) ~ 4.5 grams/100 ml, Wine~ 10 grams/100 ml.

*Statistical analysis*: Association between methylation beta values at each CpG site and alcohol usage (grams/day) or alcohol categories were analyzed using linear models adjusted for sex, age, BMI, smoking, predicted white cell counts (Houseman algorithm), Bisulfite conversion plate (96-well plate) and Bisulfite conversion efficiency mean (calculated from control probes). In total, 818 subjects were used for association analysis.

*Acknowledgements*: This study is supported by the Swedish Research Council (grant no. 2012-1397), Knut och Alice Wallenberg Foundation (grant no. 2013.0126), Swedish Heart-Lung Foundation (grant no. 20140422), and Swedish Diabetes Foundation (grant no. 2013-024). Genome-wide DNA methylation profiling in PIVUS was funded by the Uppsala University Hospital (ALF-medel) and was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

***1.12. RS***

**The Rotterdam Study** is a large prospective, population-based cohort study aimed at assessing the occurrence of and risk factors for chronic (cardiovascular, endocrine, hepatic, neurological, ophthalmic, psychiatric, dermatological, oncological, and respiratory) diseases in the elderly(42). The study comprises 14,926 subjects in total, living in the well-deﬁned Ommoord district in the city of Rotterdam in the Netherlands. In 1989, the first cohort, Rotterdam Study-I (RS-I) comprised of 7,983 subjects with age 55 years or above. In 2000, the second cohort, Rotterdam Study-II (RS-II) was included with 3,011 subjects who had reached an age of 45 years since 1989. In 2006, the third cohort, Rotterdam Study-III (RS-III) was further included with 3,932 subjects with age 45 years and above. Each participant gave an informed consent for a genetic study and the study was approved by the medical ethics committee of the Erasmus University Medical Center, Rotterdam, the Netherlands.

*DNA Methylation profiling and quality control:* At the Genetic Laboratory (Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands), the DNA methylation dataset was generated for a subset of 747 individuals of RS-III at baseline. The second DNA methylation dataset was generated in another subset of 767 864 individuals comprising of individuals at their fifth, third and second visit of RS-I, RS-II and RS3 respectively, between 2009 and 2013. Genomic DNA was extracted from whole peripheral blood by standardized salting out methods. This was followed by a bisulfide conversion using the Zymo EZ-96 DNA-methylation kit (Zymo Research, Irvine, CA, USA). The genome for each sample was then amplified, fragmented and hybridized to the Infinium Illumina Human Methylation 450k arrays according to the manufacturer’s protocol.

The quality control was performed to first filter out samples using the Methylation Module of the GenomeStudio software (http://www.illumina.com/applications/microarrays/microarray-software/genomestudio.html). Data was extracted into beta values from raw IDAT files. We excluded samples based on the detection p-value criteria >99% (n=7), poor bisulfite conversion based on control dashboard check (n=5) and failed chromosome X & Y clustering (n=4). Further data preprocessing was performed using an R programming pipeline which is based on the pipeline developed by Tost & Toulemat(43), which includes additional parameters and options to preprocess and normalize methylation data directly from idat files. The beta values were extracted using methylumi(44). We excluded 10408 probes which had a detection p-value >0.01 in >95% of samples. 11648 probes at X and Y chromosomes were excluded to avoid gender bias. The filtering criteria left us with 731 samples and 463456 probes, respectively. Finally, the raw beta values were background corrected and normalized using the DASEN option of the WateRmelon R-package(25).

*Phenotype definition:* Alcohol consumption in the RS was measured by asking questions on alcohol consumption during a home interview with the participants before each examination cycle. Questions involved specific questions on the type of alcoholic beverages consumed, i.e. beer, white wine, red wine and strong alcoholic beverages, and how much of these types of beverages were consumed during a week. This information was used to calculate a continuous alcohol phenotype “grams per day”.

*Statistical analysis*: The association between DNA methylation level as the outcome and alcohol intake as the independent variable was performed using the linear regression models, adjusting for age, BMI, smoking, blood cells counts, and technical covariates (analytic plate, chip position on plate, and columns and rows).

*Acknowledgment/Support:* The generation and management of the Illumina 450K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk, and Lisette Stolk PhD for their help in creating the methylation database. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

***1.13 TwinsUK***

The TwinsUK cohort was established in 1992 to recruit MZ and DZ same-sex twins(45). The majority of participants are healthy female Caucasians (age range from 16 to 98 years old). There are more than 13,000 twin participants from all regions across the United Kingdom and many have multiple visits over the years. All study participants provided written informed consent for genetic studies. DNA methylation was measured on 877 individuals, and 597 of them had both DNA methylation measurements and alcohol consumption information. The average time between the DNA extraction and collection of information regarding alcohol consumption was about two years.

*DNA methylation profiling and quality control***:** Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA) was used to measure DNA methylation. Details of experimental approaches have been previously described(46). To correct the technical issues caused by the two Illumina probe types, the beta- mixture quantile dilation (BMIQ) method was performed(40). DNA methylation probes that mapped incorrectly or to multiple locations in the reference sequence were removed. Probes with detection P-value >0.05 were also removed. All the probes were with non-missing values. After quality control, 597 subjects with 458,479 probes were used for subsequent analyses.

*Phenotype definitions***:** In TwinsUK, the alcohol consumption data were collected using self-reported questionnaires. Participants were asked for their average weekly amount of alcohol intakes (e.g. wine, beer, spirits, and liquors). The alcohol consumption data were summarized as units per week, and then converted to grams/day (one unit of alcohol in the UK is defined as 7.9 grams(21). The categorical phenotype is the same as the FHS definition.

*Statistical**analysis:*Association analysis was performed using non-transformed DNA methylation beta values as the outcome variables. Covariates including batch effects (plate, position on the plate), age, BMI, smoking status, 4 cell counts (eosinophils, lymphocytes, monocytes, and neutrophils), family structure and zygosity structure. A linear mixed effect regression model was applied as the data contained MZ and DZ twins. Family structure and zygosity were taken as random-effect terms, while all the other covariates were included as fixed-effect terms. All statistical analyses were performed using R software.

*Acknowledgements*: Funding support for this project was obtained from the European Research Council (project number 250157) and in part by TwinsUK, which is funded by the Wellcome Trust; European Community’s Seventh Framework Programme (FP7/2007-2013); and also receives support from the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR. TDS is the holder of an ERC Advanced Principal Investigator award (ERC 250157).

**Supplementary Tables**

Supplementary Tables 8, 22-27 are listed in Supplementary Information. Supplementary Tables 1-7, 9-21, 28, 29 are uploaded as separate excel files

**Supplementary Table 8.** Genomic Inflation Factors for Epigenome-wide Association Study and Meta-analysis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Study** | **All** | **Drink** | **Smoking** | **Light** | **At-risk** | **Heavy** |
| *Whole blood, European ancestry (EA)* | |  |  |  |  |  |
| CHS | 0.88 | 0.87 | 0.83 | n.a. | n.a. | n.a. |
| EPIC Norfolk | 1.00 | 1.01 | 1.07 | n.a. | n.a. | n.a. |
| FHS | 1.02 | 0.91 | 0.88 | 1.10 | 1.02 | 0.82 |
| InCHIANTI | 1.04 | 1.25 | 1.25 | 0.92 | 0.81 | 0.78 |
| KORA | 0.98 | 0.90 | 0.89 | 0.85 | 0.87 | 1.02 |
| LBC 1936 | 1.34 | 1.29 | 1.30 | 0.89 | n.a. | 0.57 |
| NAS | 1.08 | 1.17 | 1.17 | 0.96 | 1.21 | 1.00 |
| PIVUS | 1.13 | 1.14 | 1.12 | 1.11 | 1.09 | n.a. |
| RS | 0.86 | 0.89 | 0.91 | n.a. | n.a. | n.a. |
| TwinsUK | 1.20 | 1.12 | 1.15 | 1.02 | 1.08 | 1.17 |
| **Meta in EA whole blood** | 1.13 | 1.11 | 1.08 | 0.92 | 0.97 | 0.96 |
| *Whole blood, African ancestry (AA)* | |  |  |  |  |  |
| ARIC | 0.98 | 0.80 | 0.80 | 0.81 | 1.19 | 1.30 |
| CHS | 0.92 | 0.87 | 0.87 | n.a. | n.a. | n.a. |
| GTP | 1.31 | 1.06 | n.a. | 1.15 | n.a. | 1.19 |
| **Meta in AA whole blood** | 0.90 | 0.70 | 0.74 | 0.80 | 1.19 | 1.09 |
| *Monocyte, mixed ancestries* |  |  |  |  |  |  |
| Monocyte | 0.90 | 0.83 | 0.83 | 0.98 | n.a. | 1.01 |
| **Meta in pooled samples** | 1.07 | 0.94 | 0.92 | 0.88 | 0.80 | 1.05 |

All, association and meta-analyses used all individuals. The continuous alcohol phenotype (g per day) was used as the main predictor and the methylation proportion was used as the outcome variable, adjusting for age, sex, BMI, technical covariate, and white blood cell counts.

Drink, association and meta-analyses used drinkers. The same variables as in “All” were used.

Smoking, association and meta-analyses used drinkers. In addition to the same variables stated in All, current smoking was adjusted as an additional covariate.

Light/at-risk/heavy, light/at-risk/heavy drinkers versus non-drinkers. The categorical phenotype was used as the main predictor and the methylation proportion was used as the outcome variable, adjusting for age, sex, BMI, technical covariate, and white blood cell counts.

**Supplementary Table 22.** Genomic Features of the Significant CpGs with *P*<1x10-7 in Meta-analysis of Whole Blood Samples of European Ancestry for Continuous Alcohol Phenotype

|  |  |  |  |
| --- | --- | --- | --- |
| **Features** | **Significant  CpGs1 (%) (n=328)** | **All CpGs in the Genome2 (%) (n=438,782)** | ***P* value** |
|  | | | |
| CGI | 53(16) | 139753 (32) | 1.1x10-6 |
| CGI Shore | 156 (48) | 103434 (24) | 7.3x10-12 |
| CGI Shelf | 24 (7) | 41691 (10) | 0.25 |
| Others | 95 (29) | 153904 (35) | − |
|  |  |  |  |
| Enhancer | 96 (29) | 97967 (22) | 0.003 |
| Non Enhancer | 232 (31) | 340815 (78) | − |
|  |  |  |  |
| Promoter | 10 (3) | 30473 (7) | 0.009 |
| Non promoter | 318 (97) | 408309 (93) | − |
|  |  |  |  |
| DHS | 44 (13) | 56096 (12) | 0.62 |
| Non DHS | 284 (87) | 382686 (88) | − |

The significant CpGs were identified from whole blood samples of European ancestry after removing 35 CpGs whose methylation levels might also be affected by smoking (Supplementary Table 1); Several criteria were applied to exclude CpGs (see Methods) for the remaining genomewide CpGs (n=438,782).

**Supplementary Table 23.** The Enriched Biological Processes for Genes Mapped to the 328 CpGs with *P* < 1x10-7 in Meta-analysis of Whole Blood Samples of European Ancestry for Continuous Alcohol Phenotype

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **GO biological process experimental only** | **REF LIST  (20814)** | **Observed** | **Expected** | **Over/under** | **Fold  Enrichment** | **P-value** |
| negative regulation of transcription from RNA polymerase II promoter | 722 | 32 | 7.8 | + | 4.1 | 1.46E-07 |
| negative regulation of nucleobase-containing compound metabolic process | 1310 | 42 | 14.16 | + | 2.97 | 1.96E-06 |
| negative regulation of transcription, DNA-templated | 1104 | 38 | 11.93 | + | 3.18 | 2.29E-06 |
| positive regulation of biological process | 5302 | 100 | 57.31 | + | 1.74 | 4.25E-06 |
| negative regulation of biosynthetic process | 1401 | 43 | 15.14 | + | 2.84 | 4.37E-06 |
| negative regulation of cellular macromolecule biosynthetic process | 1240 | 40 | 13.4 | + | 2.98 | 4.74E-06 |
| negative regulation of nucleic acid-templated transcription | 1138 | 38 | 12.3 | + | 3.09 | 5.30E-06 |
| negative regulation of RNA biosynthetic process | 1153 | 38 | 12.46 | + | 3.05 | 7.61E-06 |
| negative regulation of macromolecule biosynthetic process | 1318 | 41 | 14.25 | + | 2.88 | 8.04E-06 |
| positive regulation of cellular process | 4586 | 90 | 49.57 | + | 1.82 | 8.06E-06 |
| single-organism process | 12755 | 180 | 137.88 | + | 1.31 | 9.95E-06 |
| negative regulation of nitrogen compound metabolic process | 1406 | 42 | 15.2 | + | 2.76 | 1.60E-05 |
| negative regulation of RNA metabolic process | 1193 | 38 | 12.9 | + | 2.95 | 1.93E-05 |
| negative regulation of cellular biosynthetic process | 1381 | 41 | 14.93 | + | 2.75 | 3.07E-05 |
| negative regulation of gene expression | 1404 | 41 | 15.18 | + | 2.7 | 4.90E-05 |
| single-organism cellular process | 11415 | 165 | 123.4 | + | 1.34 | 6.61E-05 |
| regulation of transcription from RNA polymerase II promoter | 1716 | 46 | 18.55 | + | 2.48 | 6.86E-05 |
| positive regulation of metabolic process | 3529 | 73 | 38.15 | + | 1.91 | 8.49E-05 |
| cellular process | 14147 | 190 | 152.93 | + | 1.24 | 1.08E-04 |
| biological regulation | 11133 | 161 | 120.35 | + | 1.34 | 1.73E-04 |
| biological\_process | 16542 | 209 | 178.82 | + | 1.17 | 1.82E-04 |
| negative regulation of metabolic process | 2481 | 57 | 26.82 | + | 2.13 | 1.91E-04 |
| localization | 4683 | 87 | 50.62 | + | 1.72 | 2.81E-04 |
| negative regulation of macromolecule metabolic process | 2207 | 52 | 23.86 | + | 2.18 | 4.24E-04 |
| regulation of metabolic process | 6588 | 110 | 71.22 | + | 1.54 | 4.36E-04 |
| cellular response to organic substance | 1978 | 48 | 21.38 | + | 2.24 | 6.57E-04 |
| negative regulation of cellular metabolic process | 2213 | 51 | 23.92 | + | 2.13 | 1.18E-03 |
| cellular response to stimulus | 6227 | 104 | 67.31 | + | 1.54 | 1.49E-03 |
| negative regulation of osteoblast differentiation | 40 | 7 | 0.43 | + | > 5 | 2.75E-03 |
| positive regulation of cellular metabolic process | 2904 | 60 | 31.39 | + | 1.91 | 3.11E-03 |
| anatomical structure development | 4699 | 84 | 50.8 | + | 1.65 | 3.24E-03 |
| system development | 4011 | 75 | 43.36 | + | 1.73 | 3.40E-03 |
| response to stress | 3648 | 70 | 39.43 | + | 1.78 | 3.84E-03 |
| organic acid transmembrane transport | 88 | 9 | 0.95 | + | > 5 | 5.12E-03 |
| regulation of biological process | 10614 | 151 | 114.74 | + | 1.32 | 5.60E-03 |
| multicellular organismal development | 4571 | 81 | 49.41 | + | 1.64 | 8.69E-03 |
| single-organism developmental process | 5209 | 89 | 56.31 | + | 1.58 | 8.96E-03 |
| developmental process | 5291 | 90 | 57.2 | + | 1.57 | 9.07E-03 |
| negative regulation of biological process | 4372 | 78 | 47.26 | + | 1.65 | 1.19E-02 |
| anion transmembrane transport | 239 | 13 | 2.58 | + | > 5 | 2.10E-02 |
| response to stimulus | 7621 | 116 | 82.38 | + | 1.41 | 2.68E-02 |
| positive regulation of macromolecule metabolic process | 2741 | 55 | 29.63 | + | 1.86 | 2.74E-02 |
| regulation of osteoblast differentiation | 109 | 9 | 1.18 | + | > 5 | 2.89E-02 |
| cellular response to chemical stimulus | 2404 | 50 | 25.99 | + | 1.92 | 3.23E-02 |
| regulation of cellular process | 10120 | 143 | 109.4 | + | 1.31 | 3.58E-02 |
| organ development | 2784 | 55 | 30.1 | + | 1.83 | 4.37E-02 |
| regulation of gene expression, epigenetic | 219 | 12 | 2.37 | + | > 5 | 4.74E-02 |

The enrichment analysis was performed by using Gene Ontology (http://geneontology.org/).

Of the 328 CpGs that were identified in EA whole blood samples, 300 CpGs reside in 257 genes that are enriched for 95 biological processes

(Bonferroni corrected P<0.05). REF LIST, all genes in the database in Gene Ontology analysis. Observed, observed number of genes in the testing set. Expected, the expected number of genes for the testing set. Over/Under, + denotes over-/under-representation when compared the observe versus expected number of genes.

**Supplementary Table 24.** Genes in the Most Significant Biological Process in Supplementary Table 23

|  |  |  |
| --- | --- | --- |
| **Gene** | **Full Name** | **Function** |
| *DNMT3B* | DNA (CYTOSINE-5)-METHYLTRANSFERASE 3B (PTHR23068:SF9) |  |
| *HEXIM1* | PROTEIN HEXIM1 (PTHR13469:SF7) |  |
| *ARID1A* | AT-RICH INTERACTIVE DOMAIN-CONTAINING PROTEIN 1A (PTHR12656:SF12) |  |
| *GATAD2B* | TRANSCRIPTIONAL REPRESSOR P66-BETA (PTHR13455:SF4) | chromatin/chromatin-binding protein |
| *NR2F6* | NUCLEAR RECEPTOR SUBFAMILY 2 GROUP F MEMBER 6 (PTHR24083:SF44) | nuclear hormone receptor;receptor;nucleic acid binding |
| *TBL1XR1* | F-BOX-LIKE/WD REPEAT-CONTAINING PROTEIN TBL1XR1 (PTHR22846:SF40) |  |
| *EZR* | EZRIN (PTHR23281:SF13) | actin family cytoskeletal protein |
| *HDAC4* | HISTONE DEACETYLASE 4 (PTHR10625:SF100) | reductase;nucleic acid binding;deacetylase |
| *PRDM16* | PR DOMAIN ZINC FINGER PROTEIN 16 (PTHR24393:SF5) | KRAB box transcription factor |
| *PARP1* | POLY [ADP-RIBOSE] POLYMERASE 1 (PTHR10459:SF62) | glycosyltransferase;DNA ligase;DNA ligase |
| *BMP2* | BONE MORPHOGENETIC PROTEIN 2 (PTHR11848:SF143) | growth factor |
| *RERE* | ARGININE-GLUTAMIC ACID DIPEPTIDE REPEATS PROTEIN (PTHR13859:SF12) |  |
| *SOX18* | TRANSCRIPTION FACTOR SOX-18 (PTHR10270:SF204) | HMG box transcription factor;nucleic acid binding |
| *RREB1* | RAS-RESPONSIVE ELEMENT-BINDING PROTEIN 1 (PTHR23233:SF42) | zinc finger transcription factor;nucleic acid binding |
| *TNF* | TUMOR NECROSIS FACTOR (PTHR11471:SF23) | tumor necrosis factor family member |
| *JARID2* | PROTEIN JUMONJI (PTHR10694:SF44) | zinc finger transcription factor |
| *ZFHX3* | ZINC FINGER HOMEOBOX PROTEIN 3 (PTHR24208:SF84) | homeobox transcription factor;zinc finger transcription factor;  RNA binding protein; actin family cytoskeletal protein |
| *SMAD3* | MOTHERS AGAINST DECAPENTAPLEGIC HOMOLOG 3 (PTHR13703:SF25) | transcription factor |
| *RARA* | RETINOIC ACID RECEPTOR ALPHA (PTHR24082:SF115) | nuclear hormone receptor;receptor;nucleic acid binding |
| *FLCN* | FOLLICULIN (PTHR31441:SF2) |  |
| *NSD1* | HISTONE-LYSINE N-METHYLTRANSFERASE,  H3 LYSINE-36 AND H4 LYSINE-20 SPECIFIC (PTHR22884:SF312) | methyltransferase;DNA binding protein |
| *NFIX* | NUCLEAR FACTOR 1 X-TYPE (PTHR11492:SF3) | transcription factor;nucleic acid binding |
| *WT1* |  |  |
| *TXNIP* | THIOREDOXIN-INTERACTING PROTEIN (PTHR11188:SF14) |  |
| *ZNF217* | ZINC FINGER PROTEIN 217 (PTHR23233:SF6) | zinc finger transcription factor;nucleic acid binding |
| *JDP2* | JUN DIMERIZATION PROTEIN 2 (PTHR23351:SF10) | transcription factor;nucleic acid binding |
| *KANK2* | KN MOTIF AND ANKYRIN REPEAT  DOMAIN-CONTAINING PROTEIN 2 (PTHR24168:SF0) |  |
| *ETV6* | TRANSCRIPTION FACTOR ETV6 (PTHR11849:SF19) | transcription factor;signaling molecule;nucleic acid binding |
| *SKI* | SKI ONCOGENE (PTHR10005:SF15) | transcription factor |
| *NFIC* | NUCLEAR FACTOR 1 C-TYPE (PTHR11492:SF2) | transcription factor;nucleic acid binding |
| *ID3* | DNA-BINDING PROTEIN INHIBITOR ID-3 (PTHR11723:SF16) | transcription factor |
| *DNMT3A* | DNA (CYTOSINE-5)-METHYLTRANSFERASE 3A (PTHR23068:SF10) |  |

**Supplementary Table 25.** The Enriched Biological Processes for Genes Mapped to the Genes in Supplementary Table 19

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **GO biological process experimental only** | **REF LIST  (20814)** | **Observed** | **Expected** | **Over/under** | **Fold  Enrichment** | **P-value** |
| negative regulation of transposition | 9 | 4 | 0.04 | + | > 5 | 8.04E-04 |
| regulation of transposition | 9 | 4 | 0.04 | + | > 5 | 8.04E-04 |
| defense response to virus | 159 | 8 | 0.72 | + | > 5 | 5.60E-03 |
| negative regulation of viral genome replication | 46 | 5 | 0.21 | + | > 5 | 1.91E-02 |
| DNA cytosine deamination | 6 | 3 | 0.03 | + | > 5 | 2.45E-02 |
| negative regulation of single stranded viral  RNA replication via double stranded DNA intermediate | 6 | 3 | 0.03 | + | > 5 | 2.45E-02 |
| immune response | 1430 | 20 | 6.46 | + | 3.1 | 3.99E-02 |

Results are for the 100 cis-genes whose transcript levels were significantly associated with 83 CpGs

(Bonferroni corrected P<0.05); REF LIST, all genes in the database in Gene Ontology analysis. Observed, observed number of genes in the testing set. Expected, the expected number of genes for the testing set. Over/Under, + denotes over-/under-representation when compared the observe versus expected number of genes.

**Supplementary Table 26.** Enriched Biological Processes for trans-Genes Significantly Associated with cg04781796 in *GABRD* (Supplementary Table 20)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **GO biological process experimental only** | **REF LIST  (20814)** | **Observed** | **Expected** | **Over/**  **under** | **Fold  Enrichment** | **P-value** |
| lymphocyte activation | 340 | 29 | 4.31 | + | > 5 | 1.09E-11 |
| immune system process | 2163 | 73 | 27.43 | + | 2.66 | 3.19E-11 |
| immune response | 1430 | 57 | 18.14 | + | 3.14 | 8.30E-11 |
| B cell activation | 133 | 19 | 1.69 | + | > 5 | 1.53E-10 |
| leukocyte activation | 413 | 30 | 5.24 | + | > 5 | 2.16E-10 |
| regulation of immune system process | 1473 | 57 | 18.68 | + | 3.05 | 2.87E-10 |
| regulation of cell activation | 454 | 30 | 5.76 | + | > 5 | 2.35E-09 |
| regulation of leukocyte activation | 423 | 29 | 5.37 | + | > 5 | 2.45E-09 |
| cell activation | 651 | 35 | 8.26 | + | 4.24 | 7.58E-09 |
| positive regulation of immune system process | 866 | 40 | 10.98 | + | 3.64 | 1.74E-08 |
| adaptive immune response | 260 | 22 | 3.3 | + | > 5 | 3.90E-08 |
| regulation of lymphocyte activation | 370 | 25 | 4.69 | + | > 5 | 1.56E-07 |
| regulation of response to stimulus | 3555 | 89 | 45.09 | + | 1.97 | 3.61E-07 |
| regulation of cell-cell adhesion | 370 | 24 | 4.69 | + | > 5 | 9.06E-07 |
| cell surface receptor signaling pathway | 2180 | 64 | 27.65 | + | 2.31 | 1.04E-06 |
| lymphocyte differentiation | 206 | 18 | 2.61 | + | > 5 | 2.15E-06 |
| regulation of homotypic cell-cell adhesion | 297 | 21 | 3.77 | + | > 5 | 3.03E-06 |
| response to stimulus | 7621 | 146 | 96.66 | + | 1.51 | 4.02E-06 |
| regulation of T cell activation | 280 | 20 | 3.55 | + | > 5 | 6.78E-06 |
| regulation of leukocyte cell-cell adhesion | 288 | 20 | 3.65 | + | > 5 | 1.09E-05 |
| positive regulation of cell-cell adhesion | 231 | 18 | 2.93 | + | > 5 | 1.28E-05 |
| positive regulation of immune response | 585 | 28 | 7.42 | + | 3.77 | 2.12E-05 |
| regulation of immune response | 930 | 36 | 11.8 | + | 3.05 | 2.73E-05 |
| positive regulation of leukocyte activation | 273 | 19 | 3.46 | + | > 5 | 2.74E-05 |
| regulation of lymphocyte proliferation | 187 | 16 | 2.37 | + | > 5 | 2.88E-05 |
| regulation of mononuclear cell proliferation | 188 | 16 | 2.38 | + | > 5 | 3.10E-05 |
| regulation of B cell proliferation | 58 | 10 | 0.74 | + | > 5 | 4.40E-05 |
| positive regulation of cell activation | 282 | 19 | 3.58 | + | > 5 | 4.59E-05 |
| regulation of leukocyte proliferation | 195 | 16 | 2.47 | + | > 5 | 5.14E-05 |
| regulation of cell adhesion | 612 | 28 | 7.76 | + | 3.61 | 5.55E-05 |
| positive regulation of leukocyte cell-cell adhesion | 198 | 16 | 2.51 | + | > 5 | 6.35E-05 |
| leukocyte differentiation | 293 | 19 | 3.72 | + | > 5 | 8.39E-05 |
| regulation of B cell activation | 103 | 12 | 1.31 | + | > 5 | 9.86E-05 |
| single organism signaling | 5192 | 107 | 65.85 | + | 1.62 | 1.53E-04 |
| signaling | 5195 | 107 | 65.89 | + | 1.62 | 1.58E-04 |
| single-organism process | 12755 | 204 | 161.78 | + | 1.26 | 1.83E-04 |
| positive regulation of lymphocyte activation | 249 | 17 | 3.16 | + | > 5 | 2.45E-04 |
| cell communication | 5325 | 108 | 67.54 | + | 1.6 | 3.05E-04 |
| positive regulation of T cell activation | 193 | 15 | 2.45 | + | > 5 | 3.09E-04 |
| regulation of biological process | 10614 | 178 | 134.63 | + | 1.32 | 3.68E-04 |
| positive regulation of homotypic cell-cell adhesion | 197 | 15 | 2.5 | + | > 5 | 4.02E-04 |
| positive regulation of cell adhesion | 362 | 20 | 4.59 | + | 4.36 | 4.60E-04 |
| immune system development | 606 | 26 | 7.69 | + | 3.38 | 6.89E-04 |
| hemopoiesis | 527 | 24 | 6.68 | + | 3.59 | 7.60E-04 |
| biological regulation | 11133 | 183 | 141.21 | + | 1.3 | 8.85E-04 |
| hematopoietic or lymphoid organ development | 574 | 25 | 7.28 | + | 3.43 | 9.35E-04 |
| positive regulation of biological process | 5302 | 105 | 67.25 | + | 1.56 | 1.95E-03 |
| regulation of signal transduction | 2538 | 62 | 32.19 | + | 1.93 | 2.14E-03 |
| single-organism cellular process | 11415 | 185 | 144.79 | + | 1.28 | 2.25E-03 |
| regulation of cellular process | 10120 | 169 | 128.36 | + | 1.32 | 2.65E-03 |
| signal transduction | 4801 | 97 | 60.89 | + | 1.59 | 3.05E-03 |
| leukocyte cell-cell adhesion | 265 | 16 | 3.36 | + | 4.76 | 3.18E-03 |
| positive regulation of lymphocyte proliferation | 118 | 11 | 1.5 | + | > 5 | 3.59E-03 |
| positive regulation of mononuclear cell proliferation | 119 | 11 | 1.51 | + | > 5 | 3.89E-03 |
| B cell proliferation | 37 | 7 | 0.47 | + | > 5 | 4.79E-03 |
| positive regulation of leukocyte proliferation | 123 | 11 | 1.56 | + | > 5 | 5.36E-03 |
| antigen receptor-mediated signaling pathway | 128 | 11 | 1.62 | + | > 5 | 7.85E-03 |
| immune response-regulating cell surface receptor signaling pathway | 480 | 21 | 6.09 | + | 3.45 | 9.30E-03 |
| regulation of lymphocyte differentiation | 135 | 11 | 1.71 | + | > 5 | 1.31E-02 |
| system development | 4011 | 83 | 50.87 | + | 1.63 | 1.31E-02 |
| immune response-regulating signaling pathway | 578 | 23 | 7.33 | + | 3.14 | 1.40E-02 |
| regulation of signaling | 2870 | 65 | 36.4 | + | 1.79 | 1.46E-02 |
| multicellular organismal development | 4571 | 91 | 57.98 | + | 1.57 | 1.70E-02 |
| negative regulation of leukocyte activation | 141 | 11 | 1.79 | + | > 5 | 1.97E-02 |
| regulation of adaptive immune response | 114 | 10 | 1.45 | + | > 5 | 2.04E-02 |
| negative regulation of immune system process | 349 | 17 | 4.43 | + | 3.84 | 2.48E-02 |
| B cell differentiation | 91 | 9 | 1.15 | + | > 5 | 2.50E-02 |
| cell adhesion | 1037 | 32 | 13.15 | + | 2.43 | 2.98E-02 |
| defense response to protozoan | 18 | 5 | 0.23 | + | > 5 | 3.21E-02 |
| biological adhesion | 1041 | 32 | 13.2 | + | 2.42 | 3.23E-02 |
| positive regulation of response to stimulus | 2033 | 50 | 25.79 | + | 1.94 | 3.36E-02 |
| antigen processing and presentation of peptide antigen via MHC class II | 95 | 9 | 1.2 | + | > 5 | 3.52E-02 |
| homotypic cell-cell adhesion | 285 | 15 | 3.61 | + | 4.15 | 3.82E-02 |
| regulation of cell proliferation | 1475 | 40 | 18.71 | + | 2.14 | 3.86E-02 |
| humoral immune response | 154 | 11 | 1.95 | + | > 5 | 4.50E-02 |
| negative regulation of B cell activation | 34 | 6 | 0.43 | + | > 5 | 4.58E-02 |
| T cell differentiation | 126 | 10 | 1.6 | + | > 5 | 4.87E-02 |
| antigen processing and presentation of peptide or polysaccharide  antigen via MHC class II | 99 | 9 | 1.26 | + | > 5 | 4.88E-02 |

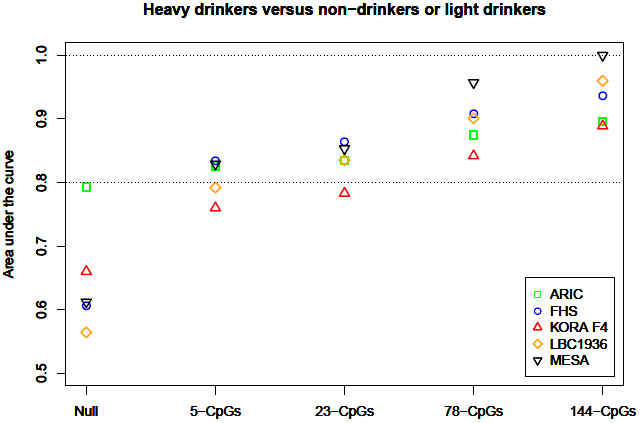
**Supplementary Table 27.** The Enriched Biological Processes for trans-Genes Significantly Associated with cg09577455 in Supplementary Table 21

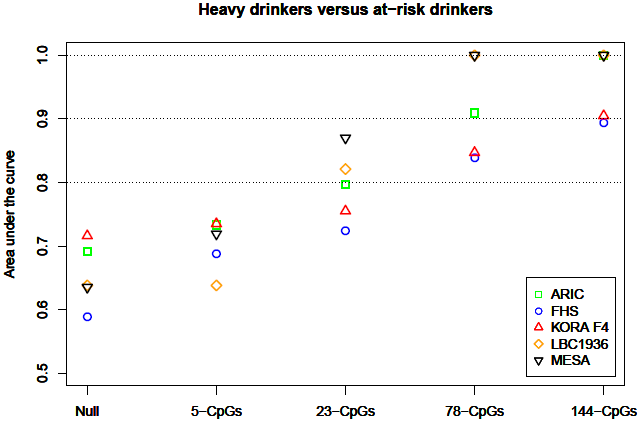
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **GO biological process experimental only** | **REF LIST  (20814)** | **Observed** | **Expected** | **Over/under** | **Fold  Enrichment** | **P-value** |
| immune response | 1430 | 10 | 1.65 | + | > 5 | 1.45E-02 |

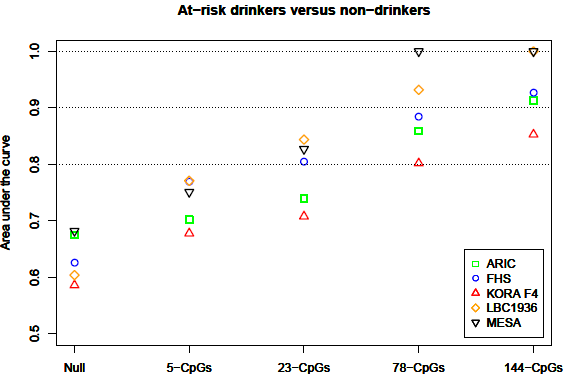
The trans-genes that were significantly associated with cg09577455 (*GABBR1*) was tested for enrichment in

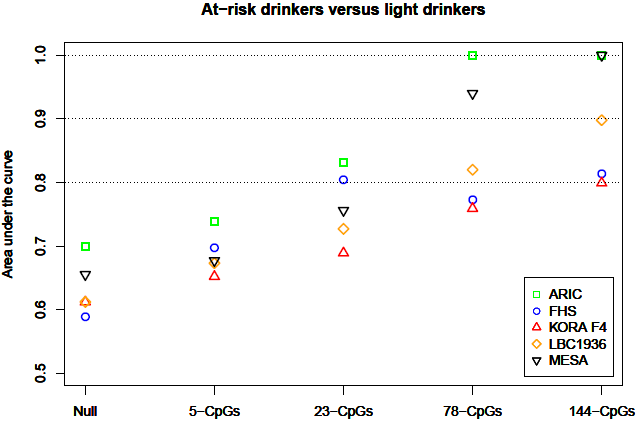
biological processes; REF LIST, all genes in the database in Gene Ontology analysis. Observed, observed number of genes in the testing set. Expected, the expected number of genes for the testing set; Over/Under, + denotes over-/under-representation when compared the observe versus expected number of genes.

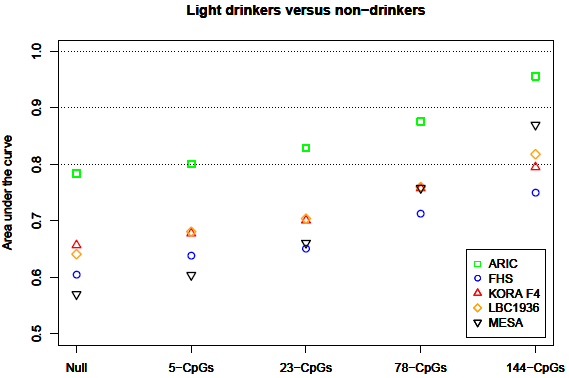
**Supplementary Figures**



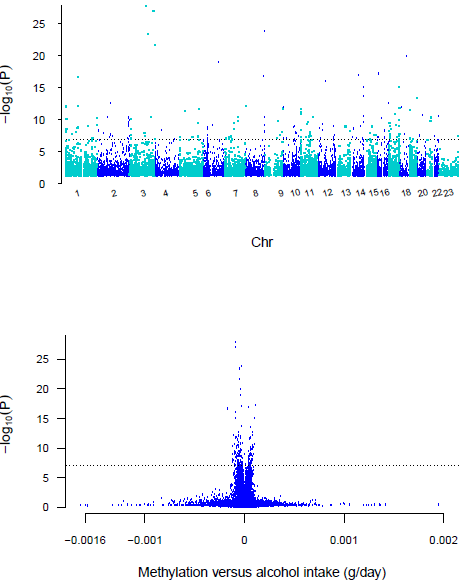




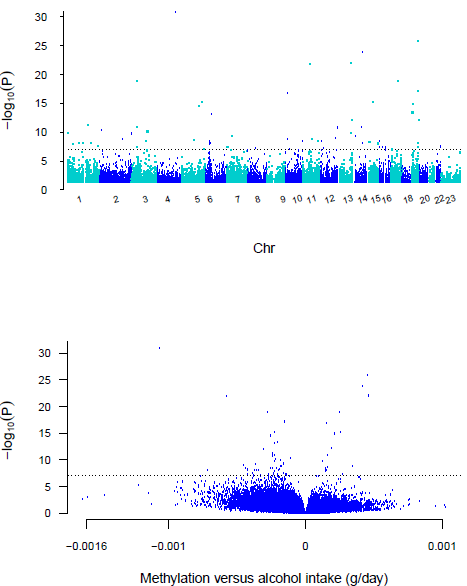




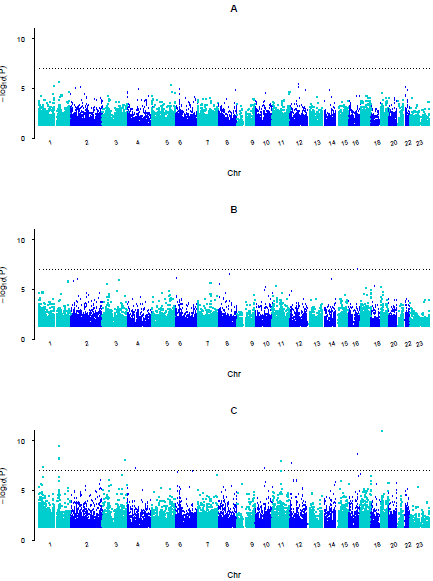
Supplementary Figure 1. **A Methylation Biomarker in Discriminating Drinking Categories.** Four sets of CpGs were selected at s=0.12 (5 CpGs), s=0.08 (23 CpGs), s=“lambda.1se” (78 CpGs), and s="lambda.min” (144 CpGs) using LASSO in the FHS cohort (the training cohort). ROC analysis was performed to classify heavy drinkers versus non-drinkers and heavy drinkers versus light drinkers. “Non-drinkers” were subjects with no alcohol consumption (i.e., g per day=0); “light drinkers” were subjects who consumed 0<g per day≤28 in men and 0<g per day≤14 in women; “at risk-drinkers” were subjects who consumed 28<g per day<42 in men and 14<g per day<28 in women; “heavy drinkers” were subjects who consumed ≥42 g per day in men and ≥28 g per day in women.



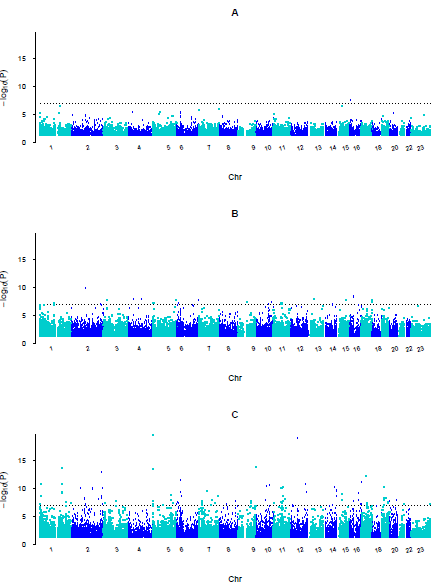
**Supplementary Figure 2.** **Manhattan (Top) and Volcano (Bottom) Plots of Epigenome-wide Association and Meta-analysis of the continuous alcohol Phenotype in African Ancestry Whole Blood Samples**.The DNA methylation proportion was the outcome variable, grams alcohol consumed per day (d/day) was the predictor variable, adjusting for age, sex, body mass index, technical covariates and white blood cell count differentials. The inverse variance weighted random effects model was performed in meta-analysis. The dotted line represents the genome-wide significance level of 1x10-7.



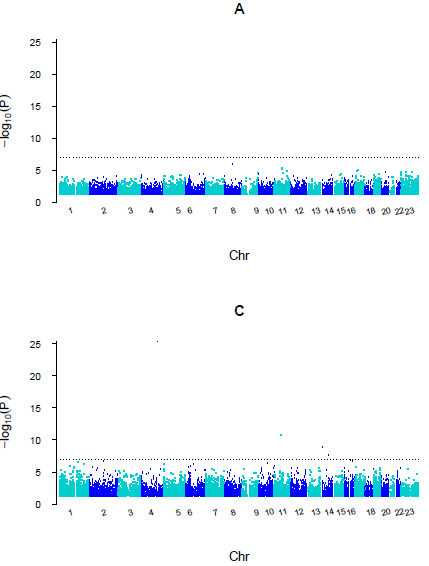
**Supplementary Figure 3.** **Manhattan (A) and Volcano (B) Plots of Epigenome-wide Association of the continuous alcohol Phenotype in CD14+ Monocyte Samples.** The DNA methylation proportion was the outcome variable, grams alcohol consumed per day (d/day) was the predictor variable, adjusting for age, sex, body mass index, technical covariates and white blood cell count differentials. The inverse variance weighted random effects model was performed in meta-analysis. The dotted line represents the genome-wide significance level of 1x10-7

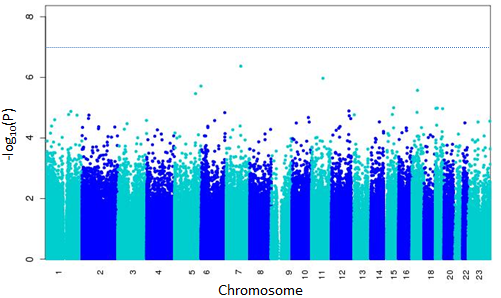


**Supplementary Figure 4. Manhattan Plots of Epigenome-wide Association of Alcohol as a Categorical Phenotype in European Ancestry Whole Blood samples.** A, Light drinkers (>0 g per day ≤28 in men and ≤14 in women, n=5,539) versus non-drinkers (n=3,334); B. At-risk drinkers (>28 g per day<42 in men and >14, n=1,005) versus non-drinkers; and C. Heavy drinkers (g per day ≥42 in men and ≥28 in women, n=694) versus non-drinkers.

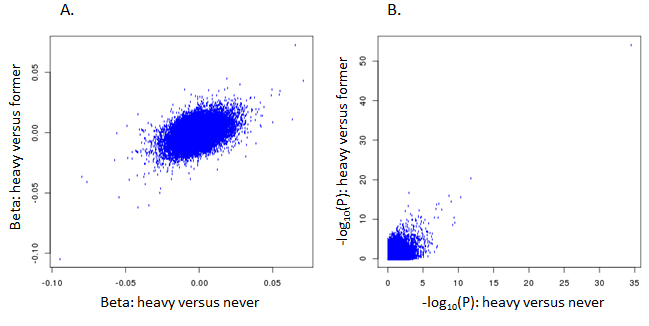


**Supplementary Figure 5. Manhattan Plots of Epigenome-wide Association of the Categorical Alcohol Phenotype in African Ancestry Whole Blood Samples**: light drinkers (0.1-28 g per day in men and 0.1-14 g per day in women) (A), at-risk drinkers (28-42 g per day in men and 14-28 g per day in women) (B), and heavy drinkers (g per day ≥42 men and ≥28 in women) (C) versus non-drinkers (g per day=0). The dotted line represents the genome-wide significance level of 1x10-7.

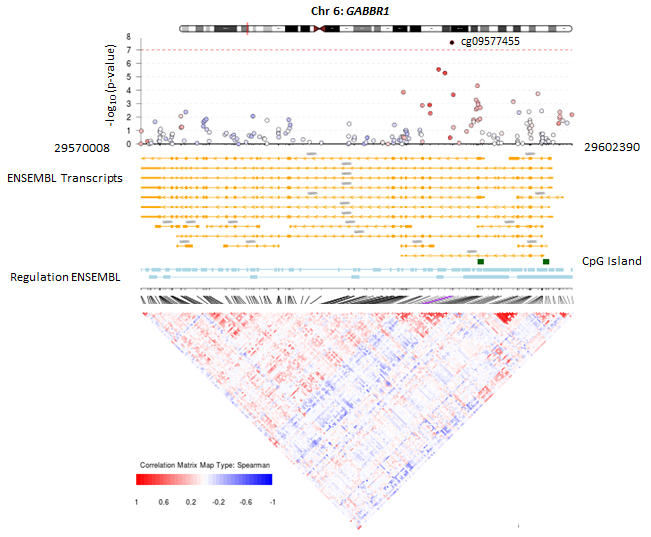
 **Supplementary Figure 6. Manhattan Plots of Epigenome-wide Association of the Categorical Alcohol Phenotype in Monocyte Derived DNA Samples**: light drinkers (0.1-28 g per day in men and 0.1-14 g per day in women) (A) and (C) versus non-drinkers (g per day=0). The dotted line represents the genome-wide significance level of 1x10-7.



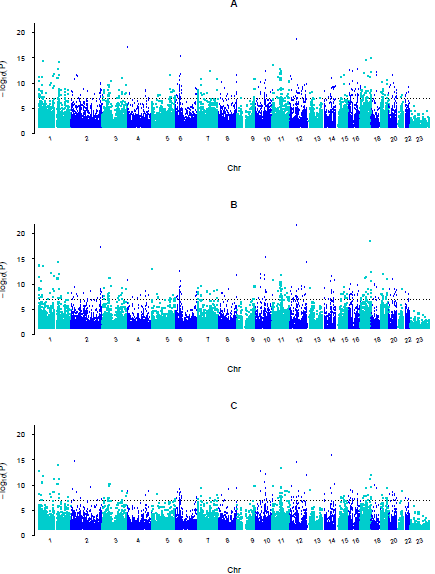
**Supplementary Figure 7. Epigenome-wide Association Analysis of the Former Drinkers versus Never Drinkers.** In Framingham Heart Study, all current non-drinkers (n=693, see Table 1) were evaluated for their alcohol consumption at any prior examinations and were classified into “never” drinkers and “former” drinkers. “Never” drinkers were individuals who self-reported no alcohol consumption at any examination; “former” drinkers were individuals who had alcohol consumption at any examination. DNA methylation was used as the outcome variable and the binary trait “former” drinkers versus “never” drinkers as the independent variables adjusting age, sex, BMI, technical covariates and WBC.



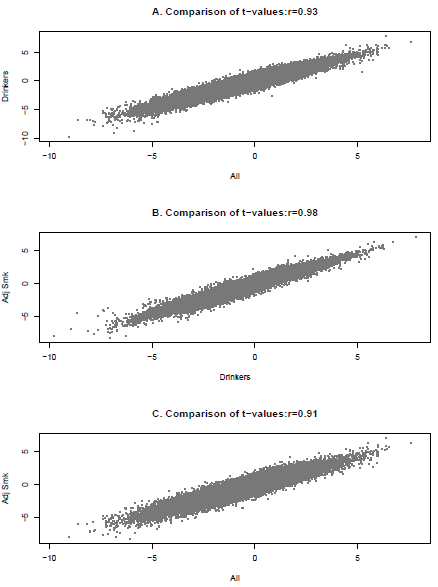
**Supplementary Figure 8. Comparisons of Regression Coefficients (A) and P-values (B) between Epigenome-wide Association Analysis of the Current Heavy Drinkers versus Never Drinkers and the Current Heavy Drinkers versus Former Drinkers.** In Framingham Heart Study, all current non-drinkers (n=693, see Table 1) were evaluated for their alcohol consumption at any prior examinations and were classified into “never” drinkers and “former” drinkers. “Never” drinkers were individuals who self-reported no alcohol consumption at any examination; “former” drinkers were individuals who had alcohol consumption at any examination. DNA methylation was used as the outcome variable and the binary traits the current “heavy” drinkers versus “former” drinkers (x-axis) and the current “heavy” drinkers versus “former” drinkers (y-axis) as the independent variables adjusting age, sex, BMI, technical covariates and WBC. For genome-wide methylation loci, the correlation was 0.32 for regression coefficients and 0.20 for –log (P values); for loci with *P* value<1x10-7 (n=92) in either “heavy” versus “never” or “heavy” versus “former”, the correlation was 0.91 for regression coefficients and 0.88 for –log (*P* values) (supplementary Figure 8). These results indicate that DNA methylation levels were not significantly different between “never” drinkers and “former” drinkers.



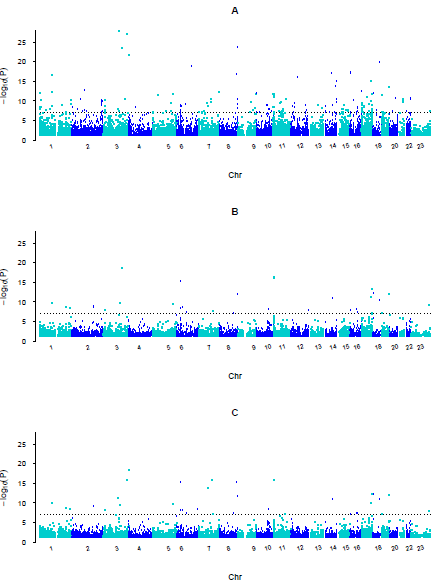
**Supplementary Figure 9**. **The GABA B Receptor Subunit 1 (*GABBR1*):** the associations of the 178 CpGs, genomic and regulatory features, and correlation of methylation. This region is loaded at 6p22.1 spanning 32.4 kb on chromosome 6.



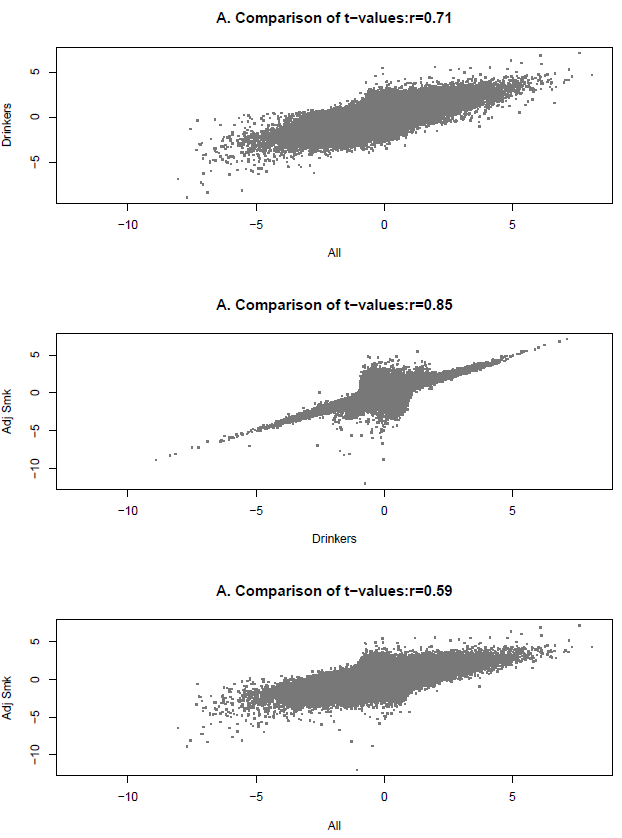
**Supplementary Figure 10. Manhattan Plots of Epigenome-wide Association of the Continuous Alcohol Phenotype in European Ancestry Whole Blood Samples**: in all individuals (A), drinkers (B), and drinkers in models adjusted for smoking (C). The dotted line represents the genome-wide significance level of 1x10-7.



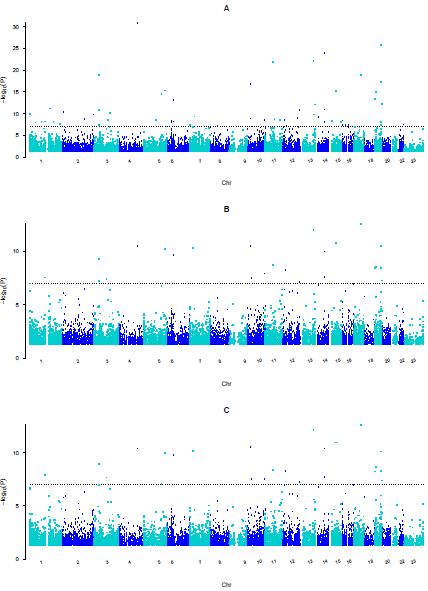
**Supplementary Figure 11. Comparison of t-values in Epigenome-wide Association of the Continuous Alcohol Phenotype in European Ancestry Whole Blood Samples**. A. All individuals (x-axis) versus drinkers (y-axis); B. drinkers (x-axis) versus drinkers adjusted for smoking (y-axis); C. all individuals (x-axis) versus drinkers adjusted for smoking (y-axis).



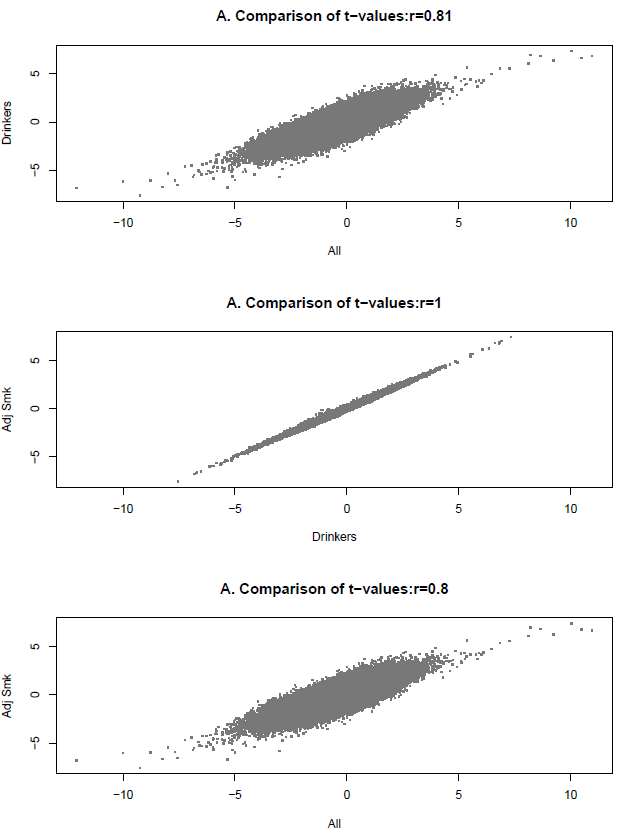
**Supplementary Figure 12. Manhattan Plots of Epigenome-wide Association of the Continuous Alcohol Phenotype in African Ancestry Whole Blood Samples**: in all individuals (A), drinkers (B), and drinkers in models adjusted for smoking (C). The dotted line represents the genome-wide significance level of 1x10-7.



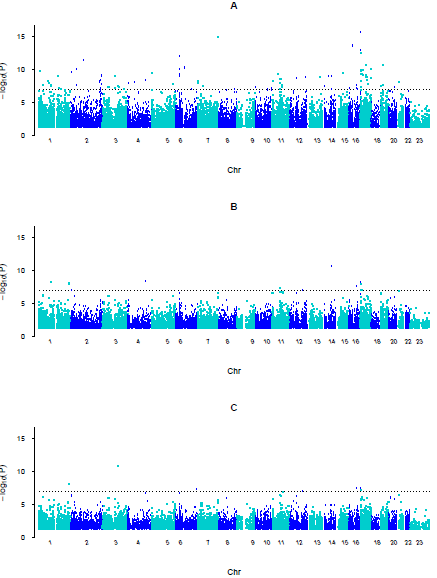
**Supplementary Figure 13. Comparison of t-values in Epigenome-wide Association of the Continuous Alcohol Phenotype in African Ancestry Whole Blood Samples**. A. All individuals (x-axis) versus drinkers (y-axis); B. drinkers (x-axis) versus drinkers adjusted for smoking (y-axis); C. all individuals (x-axis) versus drinkers adjusted for smoking (y-axis).



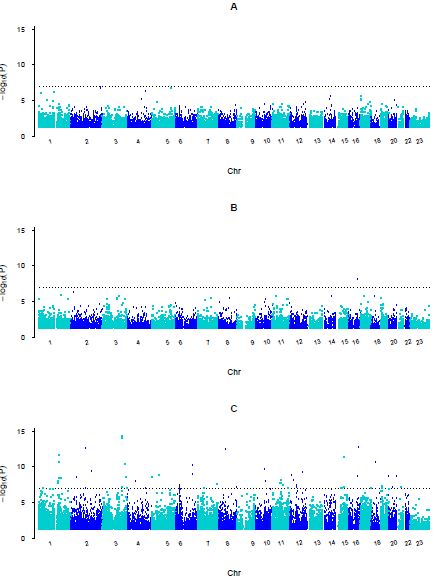
**Supplementary Figure 14. Manhattan Plots of Epigenome-wide Association of the Continuous Alcohol Phenotype in Monocyte Samples:** in all individuals (A), drinkers (B), and drinkers in models adjusted for smoking (C). The dotted line represents the genome-wide significance level of 1x10-7.



**Supplementary Figure 15. Comparison of t-values in Epigenome-wide Association of the Continuous Alcohol Phenotype in CD14+ Monocyte Samples.** A. All individuals (x-axis) versus drinkers (y-axis); B. drinkers (x-axis) versus drinkers adjusted for smoking (y-axis); C. all individuals (x-axis) versus drinkers adjusted for smoking (y-axis).



**Supplementary Figure 16. Manhattan Plots of Epigenome-wide Association of the continuous alcohol Phenotype in all Pooled Samples (A), drinkers (B), and drinkers in models adjusted for smoking (C).**  The dotted line represents the genome-wide significance level of 1x10-7.



**Supplementary Figure 17. Manhattan Plots of Epigenome-wide Association of as a Categorical Phenotype in all Pooled Samples** A, Light drinkers (>0 g per day ≤28 in men and ≤14 in women, n=6,224) versus non-drinkers (n=5,712); B. At-risk drinkers (>28 g per day<42 in men and >14, n=1,074) versus non-drinkers; and C. Heavy drinkers (g per day ≥42 in men and ≥28 in women, n=1,116) versus non-drinkers.

**Reference**

1. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009 Jun;41(6):677-87.

2. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007 Jul;39(7):906-13.

3. Huan T, Liu C, Joehanes R, Zhang X, Chen BH, Johnson AD, et al. A systematic heritability analysis of the human whole blood transcriptome. Hum Genet. 2015 Mar;134(3):343-58.

4. Joehanes R, Ying S, Huan T, Johnson AD, Raghavachari N, Wang R, et al. Gene expression signatures of coronary heart disease. Arterioscler Thromb Vasc Biol. 2013 Jun;33(6):1418-26.

5. Huan T, Zhang B, Wang Z, Joehanes R, Zhu J, Johnson AD, et al. A systems biology framework identifies molecular underpinnings of coronary heart disease. Arterioscler Thromb Vasc Biol. 2013 Jun;33(6):1427-34.

6. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012 May 8;13(1):86.

7. Schurmann C, Heim K, Schillert A, Blankenberg S, Carstensen M, Dorr M, et al. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. PloS one. 2012;7(12):e50938.

8. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003 Jan 22;19(2):185-93.

9. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008 Jul 1;24(13):1547-8.

10. Philibert RA, Beach SR, Brody GH. Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. Epigenetics. 2012 Nov;7(11):1331-8.

11. Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect. 2012 Oct;120(10):1425-31.

12. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011 Jun;6(6):692-702.

13. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989 Apr;129(4):687-702.

14. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol. 2012;13(6):R44.

15. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol. 1991 Feb;1(3):263-76.

16. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014 May 15;30(10):1363-9.

17. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15(12):503.

18. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.

19. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007 Sep;3(9):1724-35.

20. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A, et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. Br J Cancer. 1999 Jul;80 Suppl 1:95-103.

21. Brennan A, Meng Y, Holmes J, Hill-McManus D, Meier PS. Potential benefits of minimum unit pricing for alcohol versus a ban on below cost selling in England 2014: modelling study. BMJ. [Comparative Study Research Support, Non-U.S. Gov't]. 2014;349:g5452.

22. Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health. 1951 Mar;41(3):279-81.

23. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol. 1979 Sep;110(3):281-90.

24. Huan T, Zhang B, Wang Z, Joehanes R, Zhu J, Johnson AD, et al. A systems biology framework identifies molecular underpinnings of coronary heart disease. Arterioscler Thromb Vasc Biol. 2013 Jun;33(6):1427-34.

25. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. 2013;14:293.

26. Gillespie CF, Bradley B, Mercer K, Smith AK, Conneely K, Gapen M, et al. Trauma exposure and stress-related disorders in inner city primary care patients. Gen Hosp Psychiatry. 2009 Nov-Dec;31(6):505-14.

27. Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, et al. Subsystems contributing to the decline in ability to walk: Bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. J Am Geriatr Soc. 2000 Dec;48(12):1618-25.

28. 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 Feb;8(2):203-9.

29. Pisani P, Faggiano F, Krogh V, Palli D, Vineis P, Berrino F. Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. Int J Epidemiol. 1997;26 Suppl 1:S152-60.

30. Holle R, Happich M, Lowel H, Wichmann HE. KORA--a research platform for population based health research. Gesundheitswesen. 2005 Aug;67 Suppl 1:S19-25.

31. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. Int J Epidemiol. 2012 Dec;41(6):1576-84.

32. Deary IJ, Gow AJ, Taylor MD, Corley J, Brett C, Wilson V, et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. BMC Geriatr. 2007;7:28.

33. Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. J Pers Soc Psychol. 2004 Jan;86(1):130-47.

34. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. Genome Res. 2014 Nov;24(11):1725-33.

35. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, et al. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol. 2002 Nov 1;156(9):871-81.

36. Liu Y, Ding J, Reynolds LM, Lohman K, Register TC, De La Fuente A, et al. Methylomics of gene expression in human monocytes. Hum Mol Genet. 2013 Dec 15;22(24):5065-74.

37. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.

38. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes Immun. 2005 Jun;6(4):319-31.

39. Triche TJ, Jr., Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res. 2013 Apr;41(7):e90.

40. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013 Jan 15;29(2):189-96.

41. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. Arterioscler Thromb Vasc Biol. 2005 Nov;25(11):2368-75.

42. Hofman A, Murad SD, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, et al. The Rotterdam Study: 2014 objectives and design update. European Journal of Epidemiology. 2013 Nov;28(11):889-926.

43. Touleimat N, Tost J. Complete pipeline for Infinium (R) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012 Jun;4(3):325-41.

44. Davis S, Du P, Bilke S, Triche TJ, Bootwalla M. Methylumi: Handle Illumina methylation data. R package version 2.14.0. 2015.

45. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort Profile: TwinsUK and healthy ageing twin study. International journal of epidemiology. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. 2013 Feb;42(1):76-85.

46. Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics : official journal of the DNA Methylation Society. 2014;9(10):1382-96.