**Supplementary Materials**

**Description of lifestyle factors and MDD status variables in Generation Scotland**

Body mass index (BMI) was computed using height (cm) and weight (kg) as measured by clinical staff at baseline recruitment. Alcohol intake was self-reported during the pre-clinical questionnaire, which was administered during the weeks before blood draw date. Participants noted whether they were “never”, “former”, or “current” drinkers; current drinkers were asked: “During the past week, please record how many units of alcohol you have had”. Smoking status was self-reported by participants in response to the question, “Have you ever smoked tobacco?”. Answers were recorded as: “Yes, currently smoke”; “Yes, but stopped within the past 12 months”; “Yes, but stopped more than 12 months ago”; and “No, never smoked”. Pack years for smoking behaviour data were calculated by multiplying the number of cigarette packs (20 cigarettes/pack) smoked per day by the number of years a person has smoked.

Baseline MDD status was assessed by trained researchers who administered an adapted version of the axis-I Structured Clinical Interview of the Diagnostic and Statistical Manual, version IV (SCID). This was administered to participants who answered “Yes” to either of the following screening questions: “Have you ever seen anybody for emotional or psychiatric problems?” and “Was there ever a time when you, or someone else, thought you should see someone because of the way you were feeling or acting?”. Participants with no MDD were defined as those individuals who answered “No” to the two screening questions above and did not fulfil criteria for a current or previous MDD diagnosis following the SCID interview.

**R code**

1. **Epigenome-wide association study**: EWAS was performed through a locally-created pipeline; the phenotype input file should contain a column of participant IDs and separate columns for each phenotype; the pipeline also takes as input a covariate file, which should contain a column of participant IDs and separate columns with each covariate. The script below is the main part of the pipeline, indicating the regression model for self-reported antidepressant use EWAS; SSRI use EWAS were performed in a similar fashion.

Input sample: 14832 # initial input file contains 14,832 individuals

Complete phenotype and covariates: 6428 # by merging with DNA methylation data and covariates, there are 6,428 individuals remaining

Probes excluded: 0 # no probes were excluded for this analysis

Model: CpG ~antidep + units + bmi + ever\_smoke + pack\_years + dep\_status + wave + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10 + PC11 + PC12 + PC13 + PC14 + PC15 + PC16 + PC17 + PC18 + PC19 + PC20

Design matrix: CpG ~ (Intercept) + antidep + units + bmi + ever\_smoke + pack\_years + dep\_status + wave3 + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10 + PC11 + PC12 + PC13 + PC14 + PC15 + PC16 + PC17 + PC18 + PC19 + PC20

EWAS sample size: 6428 # final EWAS sample size

EWAS probes: 713522 # final number of EWAS probes

1. **Manhattan plot:** the manhattan plot function was performed using the script from the following website:

<https://github.com/pcgoddard/Burchardlab_Tutorials/wiki/GGplot2-Manhattan-Plot-Function>

gg.manhattan <- function(df, threshold, hlight, col, ylims, title){

# format df

df.tmp <- df %>%

# Compute chromosome size

group\_by(CHR) %>%

summarise(chr\_len=max(MAPINFO)) %>%

# Calculate cumulative position of each chromosome

mutate(tot=cumsum(chr\_len)-chr\_len) %>%

select(-chr\_len) %>%

# Add this info to the initial dataset

left\_join(df, ., by=c("CHR"="CHR")) %>%

# Add a cumulative position of each SNP

arrange(CHR, MAPINFO) %>%

mutate( BPcum=MAPINFO+tot) %>%

# Add highlight and annotation information

mutate( is\_highlight=ifelse(SNP %in% hlight, "yes", "no")) %>%

mutate( is\_annotate=ifelse(P < threshold, "yes", "no"))

# get chromosome center positions for x-axis

axisdf <- df.tmp %>% group\_by(CHR) %>% summarize(center=( max(BPcum) + min(BPcum) ) / 2 )

ggplot(df.tmp, aes(x=BPcum, y=-log10(P))) +

# Show all points

geom\_point(aes(color=as.factor(CHR)), alpha=0.8, size=2) +

scale\_color\_manual(values = rep(col, 22 )) +

# custom X axis:

scale\_x\_continuous( label = axisdf$CHR, breaks= axisdf$center ) +

scale\_y\_continuous(expand = c(0, 0), limits = ylims) + # expand=c(0,0)removes space between plot area and x axis

# add plot and axis titles

ggtitle(paste0(title)) +

labs(x = "Chromosome") +

# add genome-wide sig and sugg lines

geom\_hline(yintercept = -log10(sig)) +

geom\_hline(yintercept = -log10(sugg), linetype="dashed") +

# Add highlighted points

#geom\_point(data=subset(df.tmp, is\_highlight=="yes"), color="orange", size=2) +

# Add label using ggrepel to avoid overlapping

geom\_label\_repel(data=df.tmp[df.tmp$is\_annotate=="yes",], aes(label=as.factor(SNP), alpha=0.7), size=5, force=1.3) +

# Custom the theme:

theme\_bw(base\_size = 22) +

theme(

plot.title = element\_text(hjust = 0.5),

legend.position="none",

panel.border = element\_blank(),

panel.grid.major.x = element\_blank(),

panel.grid.minor.x = element\_blank()

)

}

# Variables

mypalette <- c("dark grey", "black", "orangered3") # chr color palette

mysnps <- subset(selfrep\_mod3, p\_corrected < 0.05)

# snps to highlight

sig = 7.007492e-08 # bonferroni threshold line

sugg = 0.05 # suggestive threshold line

# Run script on self-rep; 3-month; 6-month; 9-month; and 12-month analyses

# Change "CHR" to numbers

selfrep\_mod3$CHR = sub("chr", "", selfrep\_mod3$CHR)

selfrep\_mod3$CHR = as.numeric(selfrep\_mod3$CHR)

colnames(selfrep\_mod3)[1]<-"SNP"

colnames(selfrep\_mod3)[10]<-"P"

selfrep\_mod3$MAPINFO = as.numeric(selfrep\_mod3$MAPINFO)

# RUN PLOT

tiff("C:/path\_to\_file/selfrep\_manh.tiff",

width=3200, height=2000, res = 200)

gg.manhattan(selfrep\_mod3,threshold=7.007492e-08, hlight=mysnps, col=mypalette,

ylims=c(0,12), title="Self-reported antidepressant use in GS (N=6,428)")

dev.off()

1. **DNAm risk scores – biglasso and risk score calculation**

**LASSO regression (used for training dataset)**

Install.packages(“glmnet”)

install.packages("methods")

library(glmnet)

library(methods)

beta<- read.table("/path/to/DNAm")

phenotype\_file <- read.table("/path/to/phenotype\_file” ,header=T)

# subset methylation data according to whether they have the predictor of interest

pheno <- phenotype\_file[!is.na(phenotype\_file$pheno),]

a = which(colnames(clean\_beta) %in% pheno$ID)

clean\_beta1 = clean\_beta[,a]

rm(clean\_beta)

tclean\_beta = t(clean\_beta1)

rm(clean\_beta1)

ids = rownames(tclean\_beta)

dep1 = pheno[match(ids, pheno$ID),]

y = dep1$pheno

# remove columns with NAs as glmnet can't use missing data

x1 = tclean\_beta[,apply(tclean\_beta, 2, function(tar) !any(is.na(tar)))]

rm(tclean\_beta)

set.seed(1.234)

# Cross-validation: find the best shrinkage value - the alpha=1 means it's a LASSO model

lasso.cv <- cv.glmnet(x1, y, alpha=1, nfolds=10)

save(lasso.cv, “/lasso/output/")

lambda.min = lasso.cv$lambda.min

m = glmnet(x1,y,lambda=lambda.min)

# create dataframe with weights for each probe

test = data.frame(coef.name = dimnames(coef(m))[[1]], coef.value = matrix(coef(m)))

# only take predictors where the co-efficient value is not 0

coef = test[test$coef.value!=0,]

save(coef, “coef/file/")

**DNAm risk score calculation (used for testing dataset)**

## Create antidep scores in wave 3 GS

## Read in raw methylation dataset (colnames should be participant IDs; rownames should be CpG sites)

data <- read.table("/path/to/DNAm”)

## Read in list of participants you want to calculate risk scores for (ID column: participant ID; antidepressant use column: 0/1 for non-users/users)

dep\_test <- readRDS("/path/to/phenotype\_file”)

a = which(colnames(data) %in% dep\_test$ID)

meth = data[,a]

rm(data)

dat = meth

rm(meth)

meth = t(dat)

rm(dat)

meth1 = as.data.frame(meth)

meth1$id = as.character(rownames(meth1))

## Read in LASSO coefficients, which contain coef.name (CpG site); coef.value (CpG weight) --> this was derived in training dataset

load("/path/to/coefficient\_file”)

a = which(names(meth1) %in% coef$coef.name)

meth2 = meth1[,a]

meth3 = t(meth2)

probes <- intersect(coef$coef.name, rownames(meth3))

rownames(coef) = coef$coef.name

b = meth3[probes,]

p = coef[probes,]

for (i in probes) {

b[i,]= b[i,]\*p[i,"coef.value"]

}

predicted\_dep=colSums(b) + coef[1,2]

pred\_dep = as.data.frame(predicted\_dep)

pred\_dep$ID = rownames(pred\_dep)

dep = merge(dep\_test, pred\_dep, by="ID")

## Save this dataset; this should have: participant ID; antidepressant use phenotype; DNAm risk score as columns

save(dep, “/path/to/risk\_score\_file”)

|  |  |
| --- | --- |
| **CpG site** | **Beta** |
| cg03527086 | -0.004212405 |
| cg07986378 | -0.020272215 |
| cg20379007 | -0.089534147 |
| cg15696038 | 0.021687729 |
| cg19312404 | 0.010117578 |
| cg14215464 | -0.009402831 |
| cg24659858 | 0.002745261 |
| cg17494199 | -0.000933754 |
| cg24162270 | 0.010325688 |
| cg02315513 | 0.010006061 |
| cg24996979 | -0.05186749 |
| cg20858400 | -0.01063382 |
| cg01180441 | -0.003359784 |
| cg04895205 | 0.019848384 |
| cg26923045 | -0.009201063 |
| cg15774391 | -0.006088868 |
| cg13153808 | -0.00206376 |
| cg24174557 | -0.013570738 |
| cg16183741 | -0.010937547 |
| cg02790691 | -0.002922424 |
| cg00530593 | 0.006247074 |
| cg19915582 | 0.001650308 |
| cg11766468 | 0.009234013 |
| cg03636183 | -0.030474026 |
| cg00545759 | -0.000773749 |
| cg01604411 | -0.01739118 |
| cg07626482 | -0.06551365 |
| cg05603985 | -0.112459989 |
| cg16694480 | 0.045996054 |
| cg17465569 | -0.001741683 |
| cg05063806 | -0.009961271 |
| cg04609640 | 0.020506055 |
| cg09935388 | -0.004991975 |
| cg15134919 | -0.010941227 |
| cg19693031 | -0.00497332 |
| cg10117369 | 0.002004 |
| cg07380021 | -0.009811928 |
| cg15743533 | -0.012943329 |
| cg15348679 | 0.008388729 |
| cg09988805 | 0.006476494 |
| cg09902029 | -0.005004094 |
| cg16639998 | -0.006535687 |
| cg06798115 | -0.002743076 |
| cg20496896 | 0.009960566 |
| cg18902742 | -0.001149112 |
| cg17470397 | -0.005131391 |
| ch.4.194519R | -0.02003258 |
| cg16265557 | -0.003987442 |
| cg27324327 | -0.004702437 |
| cg17108430 | -0.013116569 |
| cg27324117 | 0.038234925 |
| cg18304448 | 0.000541316 |
| cg10841124 | 0.00605048 |
| cg08294028 | -0.004832679 |
| cg17508274 | -0.000326059 |
| cg03321829 | -0.003000215 |
| cg20900927 | 0.004271275 |
| cg25798600 | 0.005927247 |
| cg01787834 | -0.007071672 |
| cg06152926 | 0.001857273 |
| cg20160695 | 0.021406791 |
| cg24805875 | -0.007495492 |
| cg23528705 | -0.008169313 |
| cg11165752 | -0.00057982 |
| cg02183564 | 0.005727186 |
| cg07168526 | -0.005590957 |
| cg17440987 | 0.007992899 |
| cg27282922 | -0.001564032 |
| cg17637107 | -0.013347462 |
| cg13566278 | -0.002232548 |
| cg01692968 | -0.007056327 |
| cg13810489 | 0.007524041 |
| cg10310274 | 0.024670046 |
| cg02407308 | 0.002439814 |
| cg13732465 | -0.002694819 |
| cg17443080 | -0.037955157 |

**Table 1.** MRS CpG sites and weights.

|  |  |
| --- | --- |
| **CpG site** | **Beta** |
| cg06339248 | -0.000943258 |
| cg21709803 | -0.016357319 |
| cg09912988 | -0.00652509 |
| cg02585417 | -0.001346409 |
| cg11152667 | 2.52E-05 |
| cg12778065 | 0.01892422 |
| cg19549714 | -0.000880584 |
| cg19839825 | 0.008791186 |
| cg22797514 | -0.009421663 |
| cg10967587 | -0.015833629 |
| cg20435485 | -0.005177475 |
| cg15796978 | 0.006324823 |
| cg12194939 | -0.005430253 |
| cg06115614 | -0.043073434 |
| cg05781504 | 0.002694968 |
| cg03217966 | -0.000699748 |
| cg25711963 | 0.000407187 |
| cg08647951 | -0.006016662 |
| cg07530264 | 0.044305071 |
| cg13604445 | -0.005637856 |

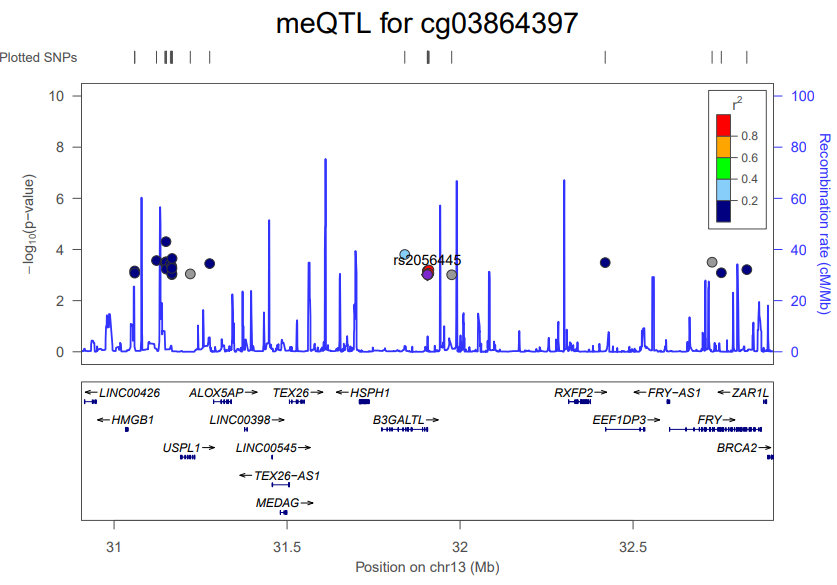
**Table 2.** MRS-ns CpG sites and weights.

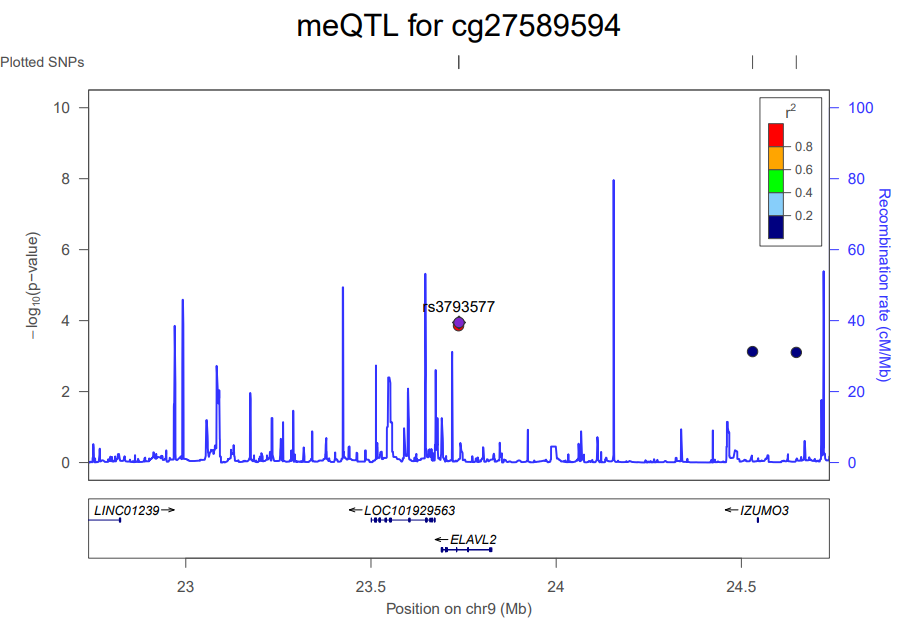
|  |  |
| --- | --- |
| **CpG site** | **Beta** |
| cg00574958 | -0.00326 |
| cg03713997 | -0.0018 |
| cg24996979 | -0.05164 |
| cg10724690 | -0.02684 |
| cg08022502 | -0.00162 |
| cg02753404 | 0.011948 |
| cg06673794 | 0.013664 |
| cg15443822 | -0.01321 |
| cg25912016 | 0.007045 |
| cg12889746 | -0.01679 |
| cg06653773 | -0.02097 |
| cg13388467 | 0.003218 |
| cg18149561 | -0.0904 |
| cg01387972 | -0.00907 |
| cg03636183 | -0.02859 |
| cg13716638 | -0.00198 |
| cg07626482 | -0.00068 |
| cg24603481 | -0.01121 |
| cg03725309 | -0.03741 |
| cg19693031 | -0.01664 |
| cg05164968 | 0.117529 |
| cg19863073 | -0.00632 |
| cg25119002 | -0.00902 |
| cg03721387 | -0.00546 |
| cg13820699 | -0.00699 |
| cg17887427 | -0.0224 |
| cg04586023 | -0.00459 |
| cg25868769 | -0.00762 |
| cg26337070 | -0.00025 |
| cg16639998 | -0.01056 |
| cg01940273 | -0.02708 |
| cg16265557 | -0.00746 |
| cg05179645 | -0.00099 |
| cg25075684 | 0.008939 |
| cg05423942 | -0.00356 |

**Table 3.** MRS-control CpG sites and weights.

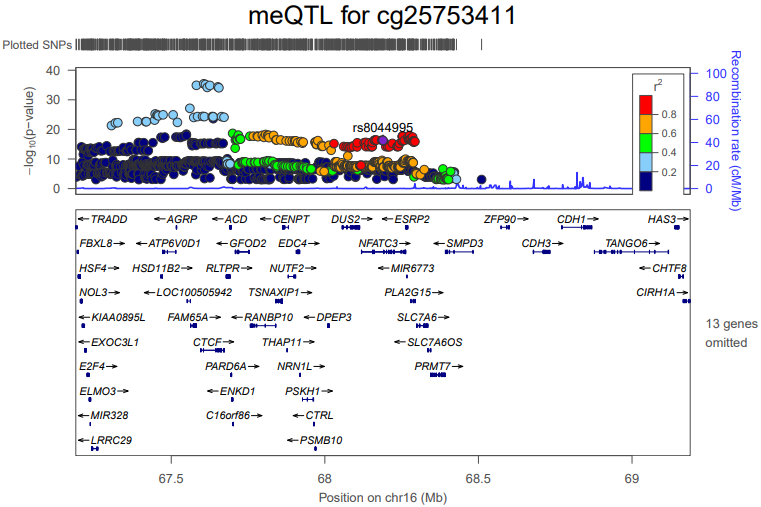
**LocusZoom plots**

**MDD**

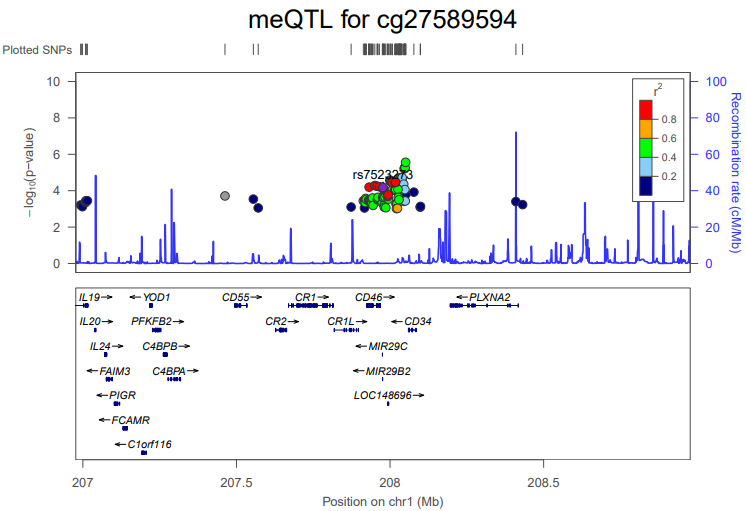
**Figure 1. Regional visualisation plot centred on the MDD-associated variant, rs2056445, in association with cg03864397, located on chromosome 2 (p = 0.0009) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg03864397 at p < 1 x 10‑3 +/- 1,000 KB from rs2056445 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.

**Figure 2. Regional visualisation plot centred on the MDD-associated variant, rs3793577, in association with cg27589594, located on chromosome 17 (p = 0.0001) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg27589594 at p < 1 x 10-3 +/- 1,000 KB from rs3793577 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.

**Schizophrenia**

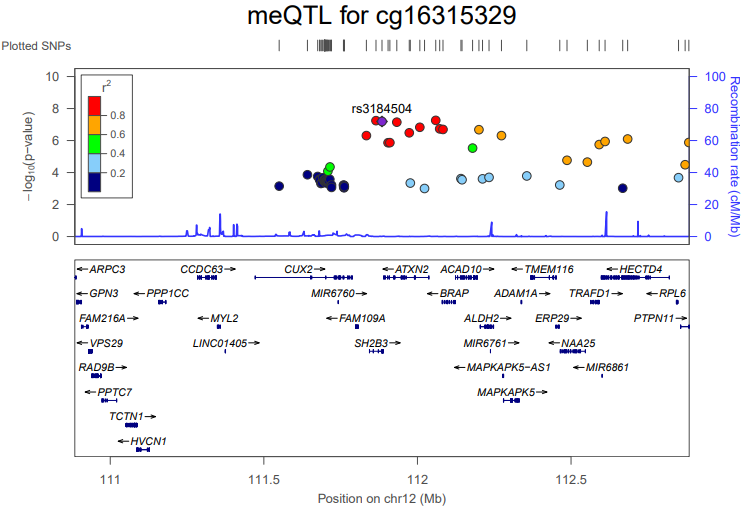
****

**Figure 3. Regional visualisation plot centred on the schizophrenia-associated variant, rs8044995, in association with cg25753411, located on chromosome 6 (p = 5.2x10-17) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg25753411at p < 1 x 10‑3 +/- 1,000 KB from rs8044995are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.

****

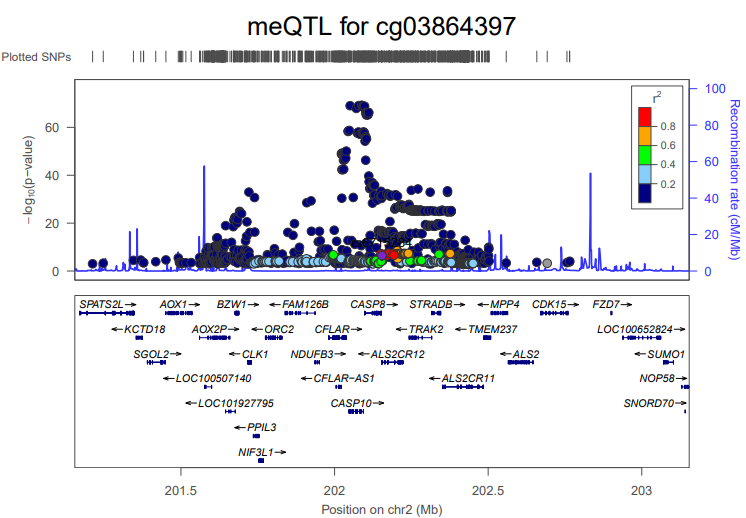
**Figure 4. Regional visualisation plot centred on the schizophrenia-associated variant, rs7523273, in association with cg27589594, located on chromosome 17 (p = 6.62x10-5) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg27589594 at p < 1 x 10‑3 +/- 1,000 KB from rs7523273 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.

**Cardiovascular disease (CAD)**

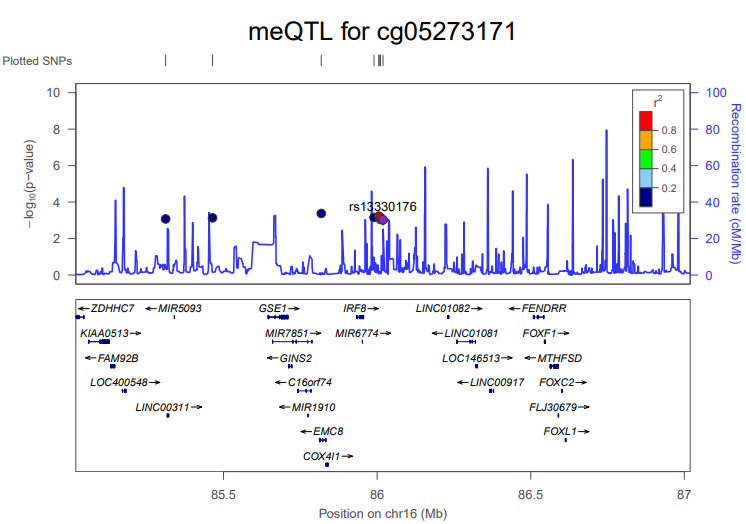
****

**Figure 5. Regional visualisation plot centred on the CAD-associated variant, rs3184504, in association with cg16315329, located on chromosome 3 (p = 6.43x10-8) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg16315329 at p < 1 x 10‑3 +/- 1,000 KB from rs3184504 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.

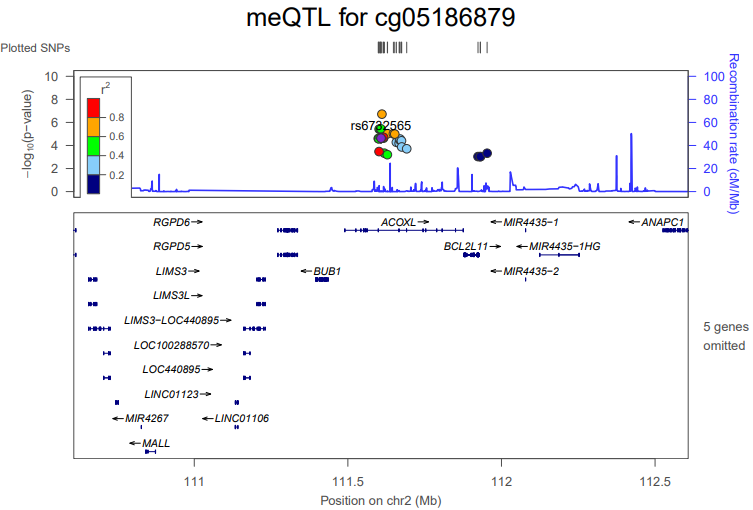
**Rheumatoid arthritis (RA)**

****

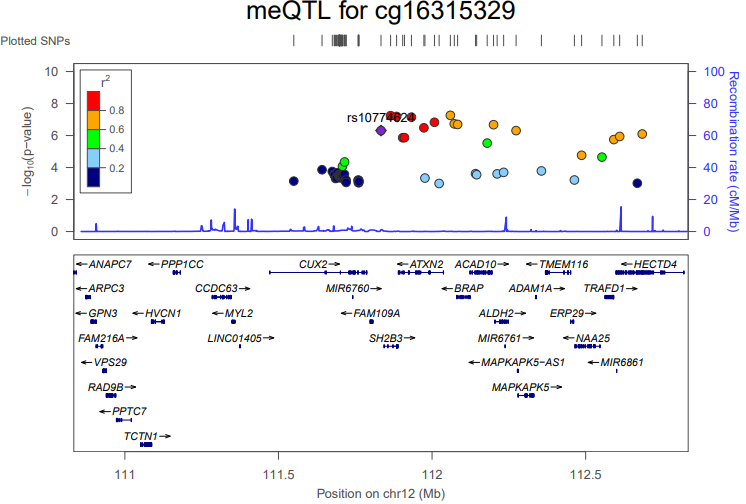
**Figure 6. Regional visualisation plot centred on the RA-associated variant, rs6715284, in association with cg03864397, located on chromosome 2 (p = 3.39x10-7) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg03864397 at p < 1 x 10‑3 +/- 1,000 KB from rs6715284 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.

****

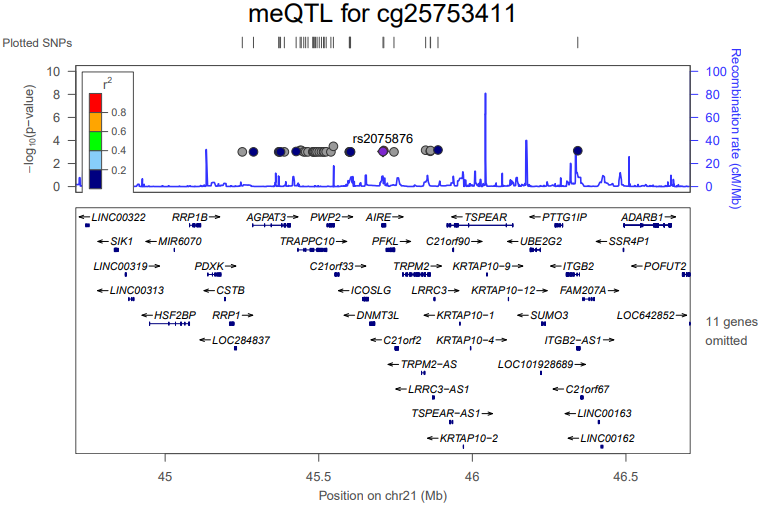
**Figure 7. Regional visualisation plot centred on the RA-associated variant, rs13330176, in association with cg05273171, located on chromosome 2 (p = 0.0009) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg05273171 at p < 1 x 10‑3 +/- 1,000 KB from rs13330176 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.



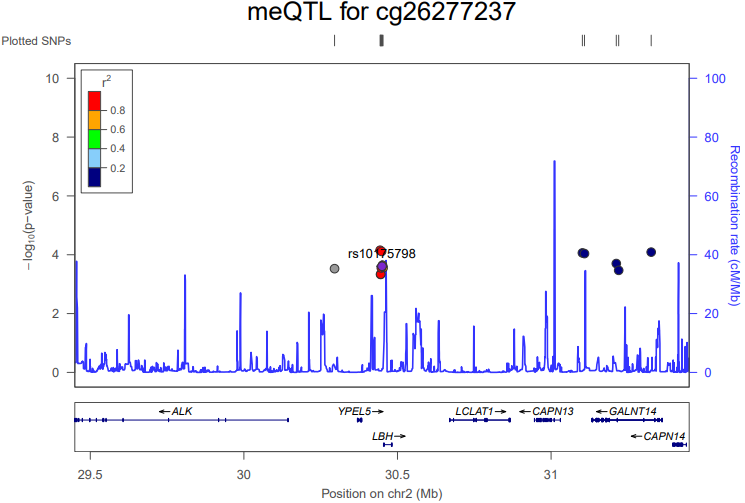
**Figure 8. Regional visualisation plot centred on the RA-associated variant, rs6732565, in association with cg05186879, located on chromosome 3 (p = 2.31x10-5) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg05186879 at p < 1 x 10‑3 +/- 1,000 KB from rs6732565 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.



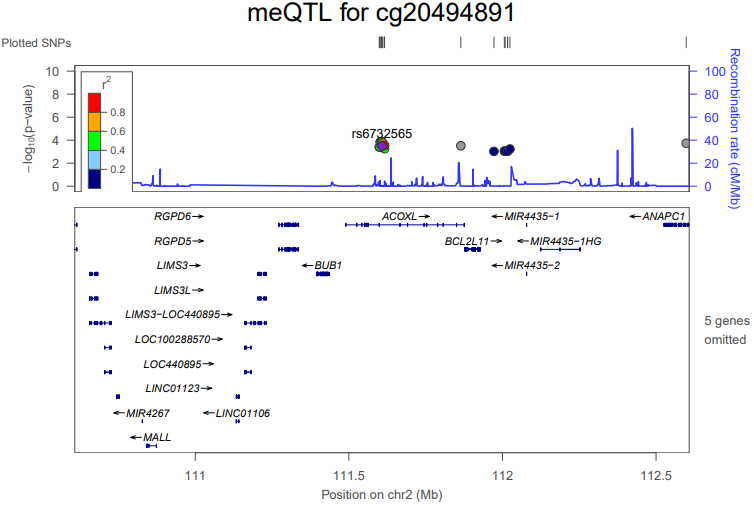
**Figure 9. Regional visualisation plot centred on the RA-associated variant, rs10774624, in association with cg16315329, located on chromosome 3 (p = 4.89x10-7) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg16315329 at p < 1 x 10‑3 +/- 1,000 KB from rs10774624 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.



**Figure 10. Regional visualisation plot centred on the RA-associated variant, rs2075876, in association with cg25753411, located on chromosome 6 (p = 0.0009) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg25753411 at p < 1 x 10‑3 +/- 1,000 KB from rs2075876 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.



**Figure 11. Regional visualisation plot centred on the RA-associated variant, rs10175798, in association with cg26277237, located on chromosome 9 (p = 0.0002) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg26277237 at p < 1 x 10‑3 +/- 1,000 KB from rs10175798 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.



**Figure 12. Regional visualisation plot centred on the RA-associated variant, rs6732565, in association with cg20494891, located on chromosome 15 (p = 0.0003) (N = 5,101 [wave1 DNAm]).** P-values were obtained from an mQTL analysis associating 638,737 CpG sites with SNPs in the Generation Scotland. SNPs associated with cg20494891 at p < 1 x 10‑3 +/- 1,000 KB from rs6732565 are depicted in the current plot. Recombination rates used in the plots are based on the European 1000 Genomes panel from November 2014.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **12-month interval** | | | | | | | |
| **CpG site** | **Gene** | **Chrom** | **β** | **P-value** | **P-corr** | **CpG site information** | **Gene information** |
| cg00477015 | *-* | 1 | -0.010 | 4.61× 10−8 | 0.033 | - | - |
| cg05954120 | *TMEM79* | 1 | -0.010 | 4.93× 10−8 | 0.0352 | HIV infection, gestational age, human liver in subjects with type 2 diabetes, lipid-related phenotypes | Hemoglobin measurement; mean corpuscular haemoglobin concentration; Contributes to epidermal integrity and skin barrier function |
| cg27277419 | *UBR4* | 1 | -0.011 | 4.36× 10−8 | 0.031 | - | Immunoglobulin isotype switching measurement; related pathways: innate immune system, MHC mediated antigen processing and presentation |
| cg02652457 | *NR2C2* | 3 | -0.029 | 3.42× 10−8 | 0.024 | - | Diseases: premature aging, lateral myocardial infarction; related pathways: gene expression, pathogenesis of cardiovascular disease |
| cg24182240 | *-* | 6 | -0.023 | 1.00× 10−8 | 0.007 | - | - |
| cg03924915 | *TRIM38* | 6 | -0.06 | 3.05× 10−9 | 0.002 | - | Well-being; height; Related pathways: innate immune system; interferon gamma signalling |
| cg22528063 | *LOC100129034* | 9 | -0.017 | 2.15× 10−8 | 0.015 | - | - |
| cg20549160 | *PML* | 15 | -0.012 | 1.92× 10−8 | 0.0137 | - | Height; erythrocyte and leukocyte count; insomnia measurement |
| cg04582760 | *UNC45A* | 15 | -0.020 | 3.27× 10−8 | 0.023 | - | Regulatory component of progesterone receptor; essential for normal cell proliferation and accumulation of myosin during muscle cell development |
| cg08227615 | *GNB5* | 15 | -0.013 | 6.82× 10−8 | 0.049 | - | Intelligence; uveal melanoma; Diseases: intellectual developmental disorder with cardiac arrhythmia; language delay and ADHD; cognitive impairment |
| cg06356163 | *ARHGAP17* | 16 | -0.051 | 3.76× 10−8 | 0.027 | - | Appendicular lean mass; blood, urinary, metabolite measurement; chronic kidney disease |

**Table 4.** CpG sites significantly associated with dispensed SSRI prescriptions within 12 months prior to the blood draw date (N=12 CpG sites), along with gene annotations, chromosome, standardised effect size, nominal and multiple comparison-corrected p-values. Background information for each CpG site and gene was extracted from EWAS (<http://www.ewascatalog.org/>) and GWAS (<https://www.ebi.ac.uk/gwas/>) catalogue databases.



**Figure 13.** Manhattan plot showing EWAS of dispensed SSRI prescriptions within 12 months prior to blood draw date. The black line defines the threshold for epigenome-wide significance (p ≤ 7.007 x 10-8) and the dotted line defines the threshold for nominal significance (p ≤ 0.05). CpG sites associated with dispensed SSRI prescriptions within this time interval are indicated above the epigenome-wide significance line. The 10 CpG sites shown to be associated with self-reported antidepressant use are indicated on the plot, between the nominal and epigenome-wide significance lines.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **EWAS** | **CpG site** | **Gene** | **Chrom** | **β** | **P-value** |
| **12-month interval** | cg03864397 | CASP10 | 2 | -0.017 | 0.011 |
| cg05186879 | MAPKAPK3; CISH | 3 | -0.019 | 0.0001 |
| cg05273171 | PNKD; TMBIM1 | 2 | -0.013 | 4.62 x 10-6 |
| cg05603985 | SKI | 1 | -0.018 | 6.23 x 10-5 |
| cg09511513 | ATP6V1B2 | 8 | -0.026 | 0.028 |
| cg16315329 | MAPKAPK3; CISH | 3 | -0.023 | 0.009 |
| cg20494891 | MYO1E | 15 | -0.018 | 0.002 |
| cg25753411 | HEBP2 | 6 | -0.031 | 1.41 x 10-5 |
| cg26277237 | KANK1 | 9 | 0.027 | 3.09 x 10-7 |
| cg27589594 | SLC5A10 | 17 | -0.027 | 4.56 x 10-7 |

**Table 5.** Results for CpG sites significantly associated with self-reported antidepressant use in the EWAS using dispensed SSRI prescriptions within 12 months prior to the blood draw date. The table incudes gene annotations, chromosome, standardised effect size, and p-value (all p-values are nominally significant).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **EWAS** | **CpG site** | **Gene** | **Chrom** | **β** | **P-value** |
| **Self-reported AD use: 12-month interval** | cg00477015 | - | 1 | 0.001 | 0.6\* |
| cg02652457 | NR2C2 | 3 | -0.008 | 0.046 |
| cg03924915 | TRIM38 | 6 | -0.023 | 0.004 |
| cg04582760 | UNC45A | 15 | -0.008 | 0.004 |
| cg04971406 | - | 2 | 0.002 | 0.026 |
| cg05954120 | TMEM79 | 1 | -0.004 | 0.005 |
| cg06356163 | ARHGAP17 | 16 | -0.035 | 1.76 x 10-6 |
| cg08227615 | GNB5 | 15 | -0.009 | 2.82 x 10-6 |
| cg20549160 | PML | 15 | -0.008 | 1.03 x 10-5 |
| cg22528063 | LOC100129034 | 9 | -0.006 | 0.01 |
| cg24182240 | - | 6 | 0.003 | 0.292\* |
| cg27277419 | UBR4 | 1 | -0.005 | 0.002 |

**Table 6.** Results for CpG sites significantly associated with dispensed SSRI prescriptions within 12 months prior to the blood draw date in the EWAS fitting self-reported antidepressant use as predictor; AD: antidepressant. The table incudes gene annotations, chromosome, standardised effect size, and p-value (all p-values are nominally significant). \* = indicates associations that are not nominally significant for self-reported antidepressant use.