main_exons

September 8, 2021

1 eQTL boxplot: Enrichment and Overlap of PGC2+CLOZUK

This is script ported from python to fix unknown plotting error.

```
[1]: suppressPackageStartupMessages({
         library(tidyverse)
         library(ggpubr)
     })
```

1.1 Functions

```
[2]: feature = "exons"
```

1.1.1 Cached functions

```
[3]: get_de_df <- function(){
         de_file = paste0("../../differential_expression/_m/", feature,
                          "/diffExpr_szVctl_full.txt")
         return(data.table::fread(de_file))
     }
     memDE <- memoise::memoise(get_de_df)

     get_eqtl_df <- function(){
         eGenes_file = paste0('../../eqtl/caudate/summary_table/_m/',
                              'Brainseq_LIBD_caudate_4features.signifpairs.txt.gz')
         eGenes = data.table::fread(eGenes_file) %>%
             filter(Type == feature_map(feature)) %>%
             arrange(pval_nominal)
         return(eGenes)
     }
     memEQTL <- memoise::memoise(get_eqtl_df)

     get_pheno_df <- function(){
         phenotype_file = paste0('/ceph/projects/v4_phase3_paper/inputs/',
                                 'phenotypes/_m/merged_phenotypes.csv')
         return(data.table::fread(phenotype_file))
     }
     memPHENO <- memoise::memoise(get_pheno_df)
```

```r
get_residualized_df <- function(){
    expr_file = paste0("../../differential_expression/_m/", feature,
                       "/residualized_expression.tsv")
    return(data.table::fread(expr_file) %>% column_to_rownames("V1"))
}
memRES <- memoise::memoise(get_residualized_df)

get_genotypes <- function(){
    traw_file = paste0("/ceph/projects/brainseq/genotype/download/topmed/
 ↪convert2plink/",
                       "filter_maf_01/a_transpose/_m/LIBD_Brain_TopMed.traw")
    traw = data.table::fread(traw_file) %>% rename_with(~ gsub('\\_.*', '', .x))
    return(traw)
}
memSNPs <- memoise::memoise(get_genotypes)

get_gwas_snps <- function(){
    gwas_snp_file = paste0('/ceph/projects/v4_phase3_paper/inputs/sz_gwas/',
                           'pgc2_clozuk/map_phase3/_m/libd_hg38_pgc2sz_snps.tsv')
    gwas_df = data.table::fread(gwas_snp_file) %>% arrange(P)
    return(gwas_df)
}
memGWAS <- memoise::memoise(get_gwas_snps)

get_integration_df <- function(){
    return(inner_join(memGWAS(), memEQTL(),
                      by=c("our_snp_id"="variant_id"),
                      suffix=c("_PGC2", "_eQTL")) %>%
           inner_join(memDE(), by=c("gene_id"="V1")) %>%
           mutate(agree_direction=sign(OR -1) * sign(slope) * sign(t) *␣
 ↪ifelse(pgc2_a1_same_as_our_counted, 1, -1)))
}
memMERGE <- memoise::memoise(get_integration_df)

get_geno_annot <- function(){
    return(memSNPs() %>% select(CHR, SNP, POS, COUNTED, ALT))
}
memANNOT <- memoise::memoise(get_geno_annot)

get_snps_df <- function(){
    return(memSNPs() %>% select("SNP", starts_with("Br")))
}
memGENOTYPES <- memoise::memoise(get_snps_df)

get_snp_df <- function(variant_id, gene_id){
    zz = memANNOT() %>% filter(SNP == variant_id)
```

```r
    xx = memGENOTYPES() %>% filter(SNP == variant_id) %>%
        column_to_rownames("SNP") %>% t %>% as.data.frame %>%
        rownames_to_column("BrNum") %>% mutate(COUNTED=zz$COUNTED, ALT=zz$ALT)␣
↪%>%
        rename("SNP"=all_of(variant_id))
    yy = memRES()[gene_id, ] %>% t %>% as.data.frame %>%
        rownames_to_column("RNum") %>% inner_join(memPHENO(), by="RNum")
    ## Annotated SNPs
    letters = c()
    for(ii in seq_along(xx$COUNTED)){
        a0 = xx$COUNTED[ii]; a1 = xx$ALT[ii]; number = xx$SNP[ii]
        letters <- append(letters, letter_snp(number, a0, a1))
    }
    xx = xx %>% mutate(LETTER=letters, ID=paste(SNP, LETTER, sep="\n"))
    df = inner_join(xx, yy, by="BrNum") %>% mutate_if(is.character, as.factor)
    return(df)
}
memDF <- memoise::memoise(get_snp_df)
```

1.1.2 Simple functions

```r
[4]: feature_map <- function(feature){
    return(list("genes"="Gene", "transcripts"= "Transcript",
                "exons"= "Exon", "junctions"= "Junction")[[feature]])
}

letter_snp <- function(number, a0, a1){
    if(is.na(number)){ return(NA) }
    if( length(a0) == 1 & length(a1) == 1){
        seps = ""; collapse=""
    } else {
        seps = " "; collapse=NULL
    }
    return(paste(paste0(rep(a0, number), collapse = collapse),
                 paste0(rep(a1, (2-number)), collapse = collapse), sep=seps))
}

save_ggplots <- function(fn, p, w, h){
    for(ext in c('.pdf', '.png', '.svg')){
        ggsave(paste0(fn, ext), plot=p, width=w, height=h)
    }
}

get_biomart_df <- function(){
    biomart = data.table::fread("../_h/biomart.csv")
}
memMART <- memoise::memoise(get_biomart_df)
```

```r
get_gene_symbol <- function(gene_id){
    ensemblID = gsub("\\..*", "", gene_id)
    geneid = memMART() %>% filter(ensembl_gene_id == gsub("\\..*", "", gene_id))
    if(dim(geneid)[1] == 0){
        return("")
    } else {
        return(geneid$external_gene_name)
    }
}


plot_simple_eqtl <- function(fn, gene_id, variant_id, eqtl_annot){
    bxp = memDF(variant_id, gene_id) %>%
        ggboxplot(x="ID", y=gene_id, fill="red", add="jitter", xlab="",
                  ylab="Residualized Expression", outlier.shape=NA,
                  add.params=list(alpha=0.5), alpha=0.4,
                  ggtheme=theme_pubr(base_size=20, border=TRUE)) +
        font("xy.title", face="bold") +
        ggtitle(paste(get_gene_symbol(gene_id), gene_id, eqtl_annot, sep='\n'))␣
 ↪+
        theme(plot.title = element_text(hjust = 0.5, face="bold"))
    print(bxp)
    save_ggplots(fn, bxp, 7, 7)
}
```

1.1.3 GWAS plots

```r
get_risk_allele <- function(OR, A1, A2){
    ra = ifelse(OR > 1, A1, A2)
    return(ra)
}

get_df <- function(){
    return(memEQTL() %>% inner_join(memGWAS(), by="variant_id"))
}

get_gwas_ordered_snp_df <- function(variant_id, gene_id,␣
 ↪pgc2_a1_same_as_our_counted, OR){
    df = memDF(variant_id, gene_id)
    if(!pgc2_a1_same_as_our_counted){ # Fix bug with matching alleles!
        if(OR < 1){ df = df %>% mutate(SNP = 2-SNP, ID=paste(SNP, LETTER,␣
 ↪sep="\n")) }
    } else {
        if(OR > 1){ df = df %>% mutate(SNP = 2-SNP, ID=paste(SNP, LETTER,␣
 ↪sep="\n")) }
    }
    return(df)
```

```
}

plot_gwas_eqtl_pheno <- function(fn, gene_id, variant_id,␣
 ↪pgc2_a1_same_as_our_counted, OR, title){
    bxp = get_gwas_ordered_snp_df(variant_id, gene_id,␣
 ↪pgc2_a1_same_as_our_counted, OR) %>%
        mutate_if(is.character, as.factor) %>% filter(Dx %in% c("CTL", "SZ"),␣
 ↪Age > 17) %>%
        ggboxplot(x="ID", y=gene_id, fill="Dx", color="Dx", add="jitter",␣
 ↪xlab=variant_id,
                ylab="Residualized Expression", outlier.shape=NA,
                add.params=list(alpha=0.5), alpha=0.4, legend="bottom",
                ggtheme=theme_pubr(base_size=20, border=TRUE)) +
        font("xy.title", face="bold") + ggtitle(title) +
        theme(plot.title = element_text(hjust = 0.5, face="bold"))
    print(bxp)
    save_ggplots(fn, bxp, 7, 9)
}
```

1.1.4 Exon specific

```
[6]: get_exon_annot <- function(){
    return(data.table::fread("../_h/exons.csv") %>%
            select(exon_id, brainseq_exon_id, gene_name) %>%
            distinct(brainseq_exon_id, .keep_all = TRUE))
}
memEXONS <- memoise::memoise(get_exon_annot)

exon_annotation <- function(gene_id){
    return(filter(memEXONS(), brainseq_exon_id == gene_id)$exon_id)
}
```

1.2 Integration analysis

```
[7]: dir.create(feature)
```

1.2.1 Enrichment

Integrate DEG with PGC2+CLOZUK SNPs

```
[8]: dft = memMERGE() %>% mutate(agree_direction=ifelse(agree_direction == 1, "Yes",
                                            ifelse(agree_direction ==␣
 ↪-1, "No", 0)))
dim(dft)
```

1. 3650186 2. 53

```
[9]: table(dft$agree_direction)
```

```
        O      No     Yes
     3615 1820221 1826350
```

```
[10]: table = matrix(c(sum((dft$P<5e-8)  & (dft$adj.P.Val < 0.05)),
                        sum((dft$P>=5e-8) & (dft$adj.P.Val < 0.05)),
                        sum((dft$P<5e-8)  & (dft$adj.P.Val >= 0.05)),
                        sum((dft$P>=5e-8) & (dft$adj.P.Val >= 0.05))),
                    nrow=2)
      print(table)
      fisher.test(table)
```

```
         [,1]    [,2]
  [1,]   3962   67556
  [2,] 256601 3322067
```

```
        Fisher's Exact Test for Count Data

  data:  table
  p-value < 2.2e-16
  alternative hypothesis: true odds ratio is not equal to 1
  95 percent confidence interval:
   0.734971 0.784211
  sample estimates:
  odds ratio
   0.7592562
```

```
[11]: dft2 = dft %>% filter(P <= 5e-8, `adj.P.Val` < 0.05) %>%
          mutate(eqtl_gwas_dir=sign(OR -1) * sign(slope) *␣
      →ifelse(pgc2_a1_same_as_our_counted, 1, -1),
                 de_dir=sign(t), eqtl_slope=sign(OR␣
      →-1)*sign(slope)*ifelse(pgc2_a1_same_as_our_counted, 1, -1)) %>%
          #rowwise() %>% mutate(risk_allele=get_risk_allele(our_snp_id)) %>%
          select(gene_id, Symbol, our_snp_id, rsid, A1, A2, OR, P, pval_nominal, adj.
      →P.Val, logFC,
                 t, eqtl_slope, de_dir, eqtl_gwas_dir, agree_direction,␣
      →pgc2_a1_same_as_our_counted) %>%
          rename("variant_id"="our_snp_id") %>% mutate_all(list(~na_if(.,""))) %>%
          mutate(Symbol = coalesce(Symbol,gene_id))
      dft2 %>% data.table::fwrite(paste0(feature, "/integration_by_symbol.txt"),␣
      →sep='\t')
      dim(dft2)
```

1. 3964 2. 17

```
[12]: df = dft2 %>% group_by(Symbol) %>% slice(1) %>% arrange(P)
      table(df$agree_direction)
```

```
    No Yes
    11  14
```

[13]: `df`

A grouped_df: 25 × 17

gene_id <chr>	Symbol <chr>	variant_id <chr>	rsid <chr>	A1 <chr>	A2 <chr>	OR <dbl>
e378566	ZSCAN26	chr6:28744470:A:G	rs1233578	A	G	1.2228(
e378375	ZNF391	chr6:27837477:A:C	rs34706883	A	C	1.2622(
e379384	HCG4	chr6:29298706:G:A	rs3117439	G	A	1.2636(
e380557	FLOT1	chr6:31204374:T:C	rs3132510	T	C	1.2068(
e386363	BRD2	chr6:32669525:G:A	rs2395231	G	A	1.2197(
e380196	PPP1R10	chr6:30894646:C:T	rs1264319	C	T	1.1415(
e598887	CNNM2	chr10:102852578:T:A	rs11191419	T	A	1.0875(
e378174	HCG11	chr6:26466161:G:A	rs1977199	G	A	0.9143;
e385196	ATF6B	chr6:31758911:C:A	rs707939	C	A	0.9263;
e805625	ZFYVE21	chr14:103847845:G:A	rs10083370	G	A	1.0780(
e193333	NGEF	chr2:232930111:C:T	rs2944591	C	T	0.9246;
e805477	e805477	chr14:103778246:C:T	rs12890837	C	T	0.9286;
e4306	PLCH2	chr1:2455662:C:T	rs4648845	C	T	0.9264(
e361367	LINC01470	chr5:152797561:A:G	rs111294930	A	G	1.0858(
e251235	PCCB	chr3:136435986:C:T	rs66691851	C	T	1.0680(
e387950	IP6K3	chr6:33773939:A:G	rs4711350	A	G	0.9238(
e378369	ZNF204P	chr6:27487359:G:A	rs6456793	G	A	1.0680(
e251474	SLC35G2	chr3:136789166:C:A	rs1280622	C	A	0.9404(
e880446	HIRIP3	chr16:30004701:C:T	rs12928610	C	T	1.0571(
e353080	REEP2	chr5:138439892:C:G	rs982085	C	G	1.0571(
e226781	PPM1M	chr3:52304836:T:C	rs9311474	T	C	1.0558(
e1136648	ZC3H7B	chr22:41357599:G:A	rs11090045	G	A	0.9410;
e678063	B3GAT1	chr11:134426490:C:T	rs893949	C	T	1.0548(
e1037535	ZNF14	chr19:19633270:T:C	rs11878202	T	C	1.0619(
e897871	SLC7A6	chr16:68251944:A:G	rs1975802	A	G	0.9329;

1.2.2 Plot with PGC2 risk allele

[14]:
```r
for(num in seq_along(df$gene_id)){
    variant_id = df$variant_id[num]
    gene_id = df$gene_id[num]
    exon_id = exon_annotation(gene_id)
    gene_name = df$Symbol[num]
    gencode_id = filter(memDE(), V1 == gene_id)$gencodeID
    pgc2_a1_same_as_our_counted = df$pgc2_a1_same_as_our_counted[num]
    OR = df$OR[num]; A1 = df$A1[num]; A2 = df$A2[num]
    fn = paste0(feature, "/eqtl_gwas_", gsub("\\.", "_", gene_name))
    de_annot = paste('DE adj.P.Val:', signif(df$adj.P.Val[num], 2))
    eqtl_annot = paste("eQTL nominal p-value:", signif(df$pval_nominal[num], 2))
    gwas_annot = paste("SZ GWAS pvalue:", signif(df$P[num], 2))
```

```
    risk_annot = paste("SZ risk allele:", get_risk_allele(OR, A1, A2))
    title = paste(get_gene_symbol(gencode_id), exon_id, gwas_annot,
                  risk_annot, eqtl_annot, de_annot, sep='\n')
    plot_gwas_eqtl_pheno(fn, gene_id, variant_id, pgc2_a1_same_as_our_counted,␣
 ↪OR, title)
    #print(title)
}
```
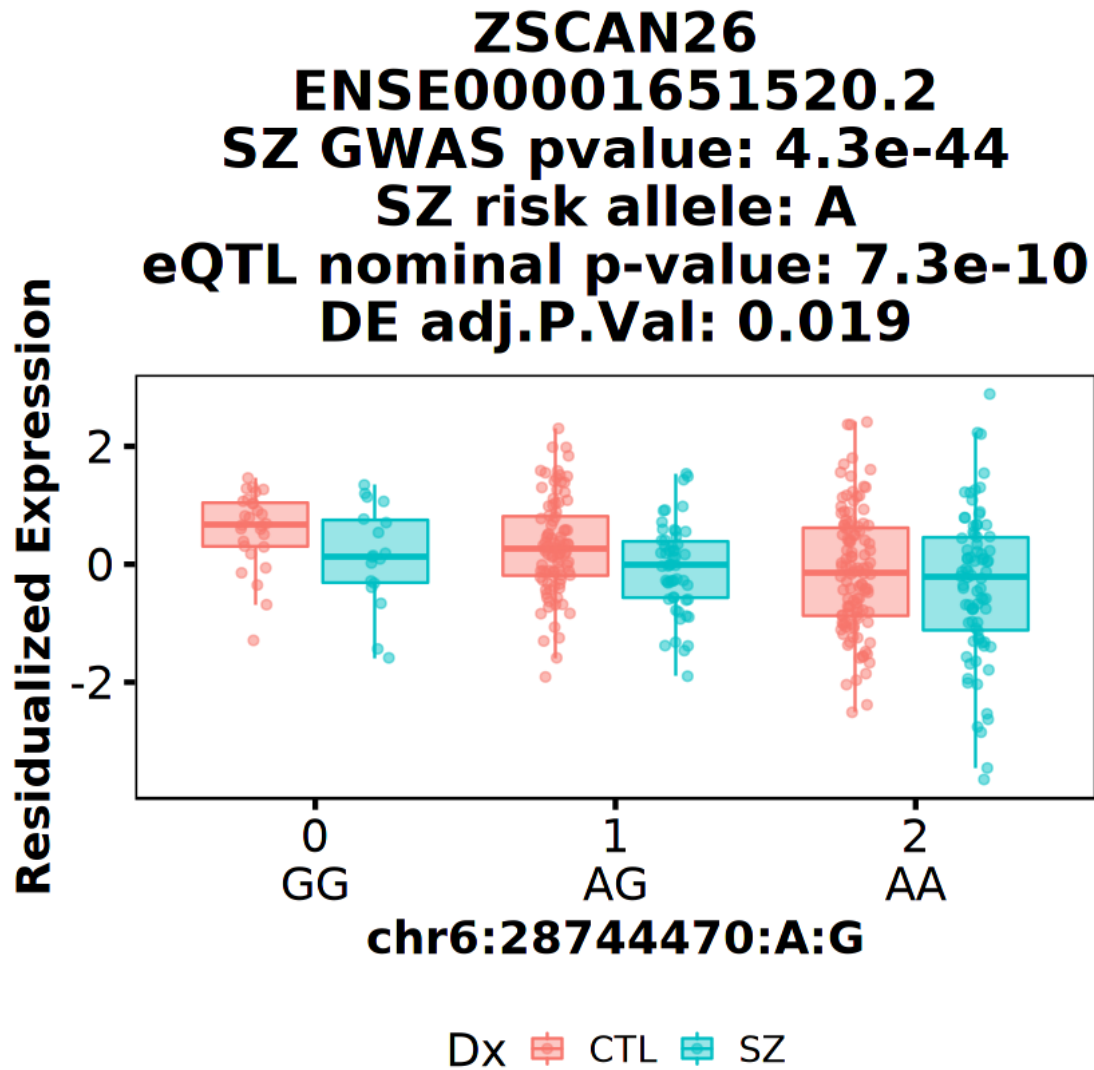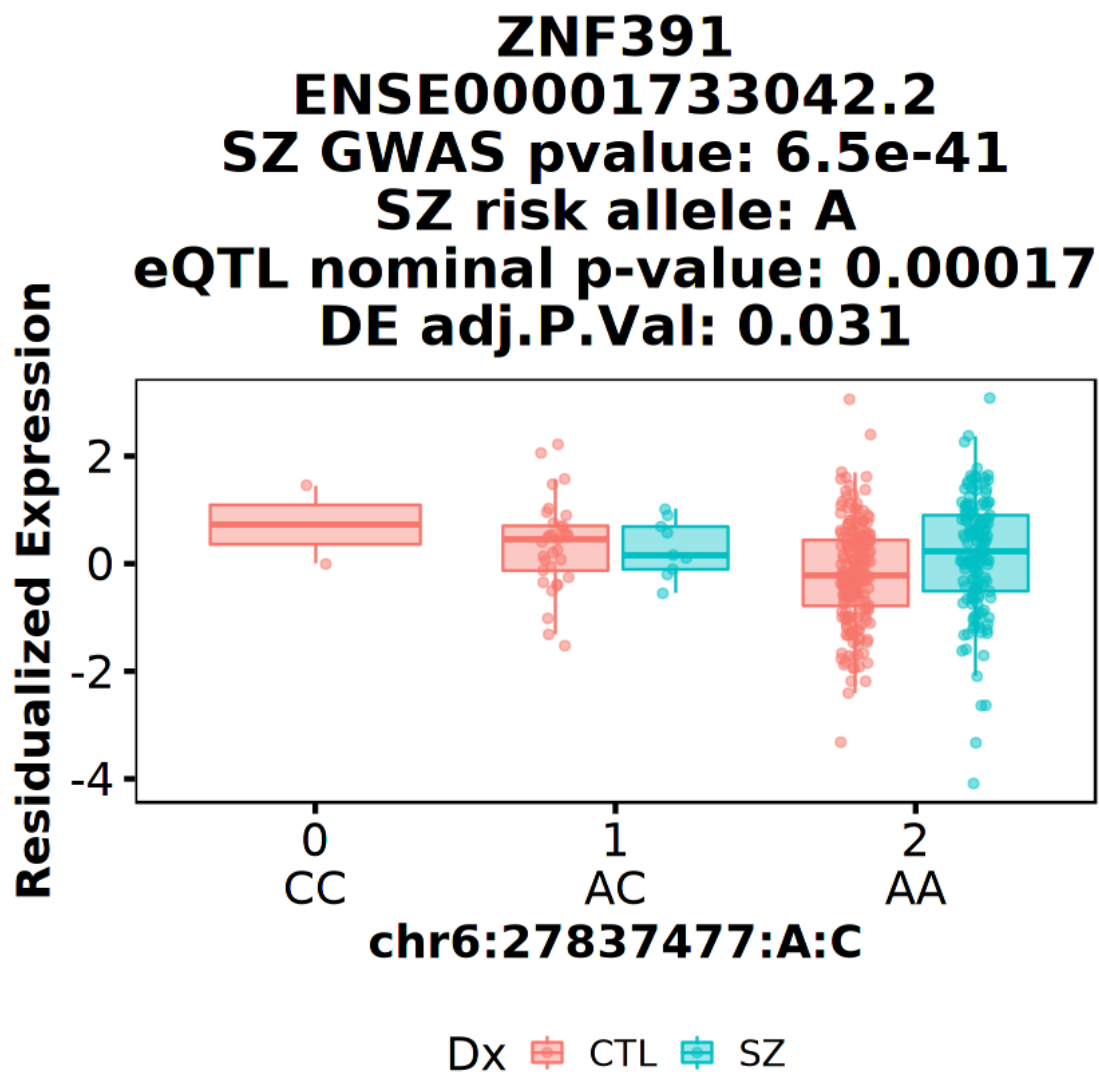
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
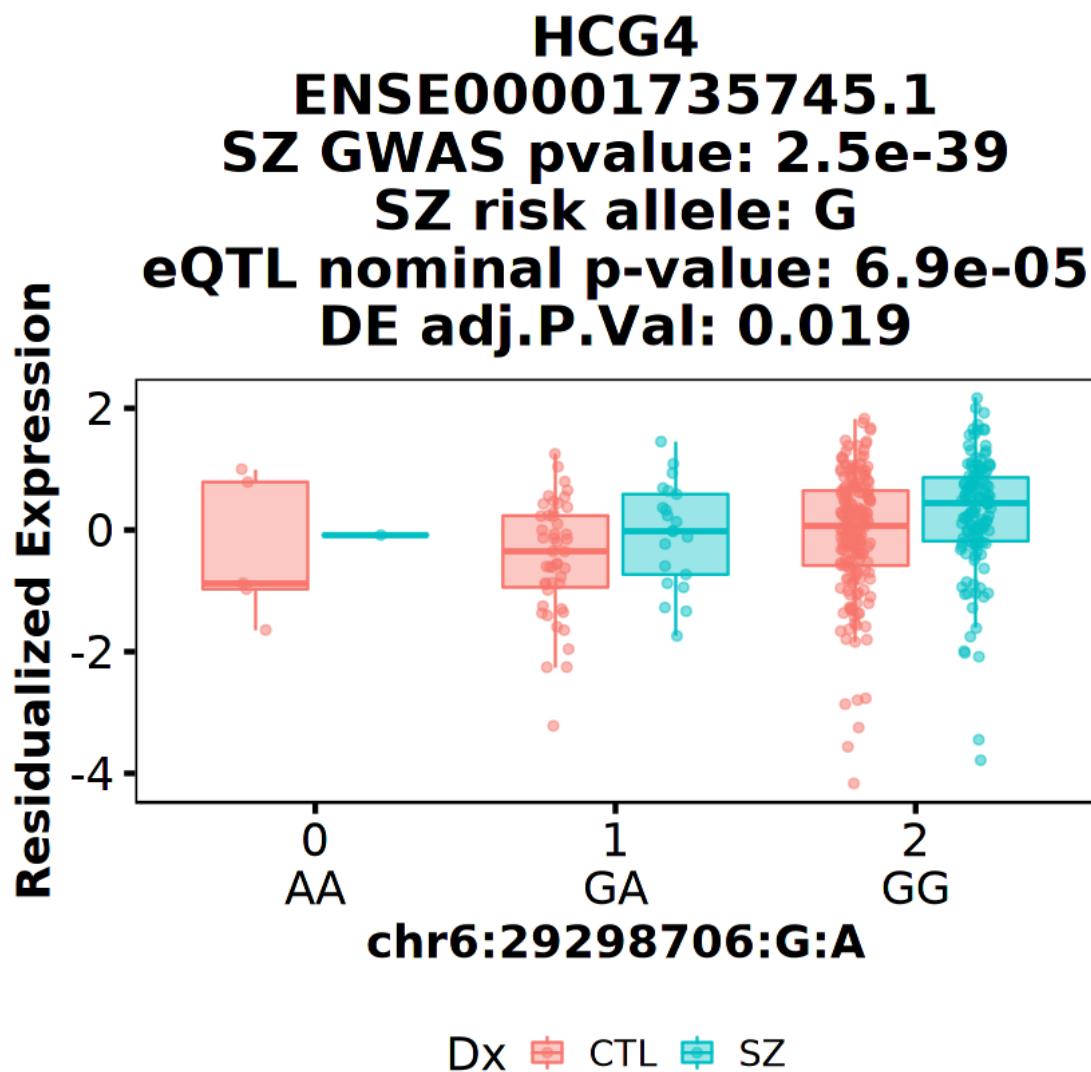Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
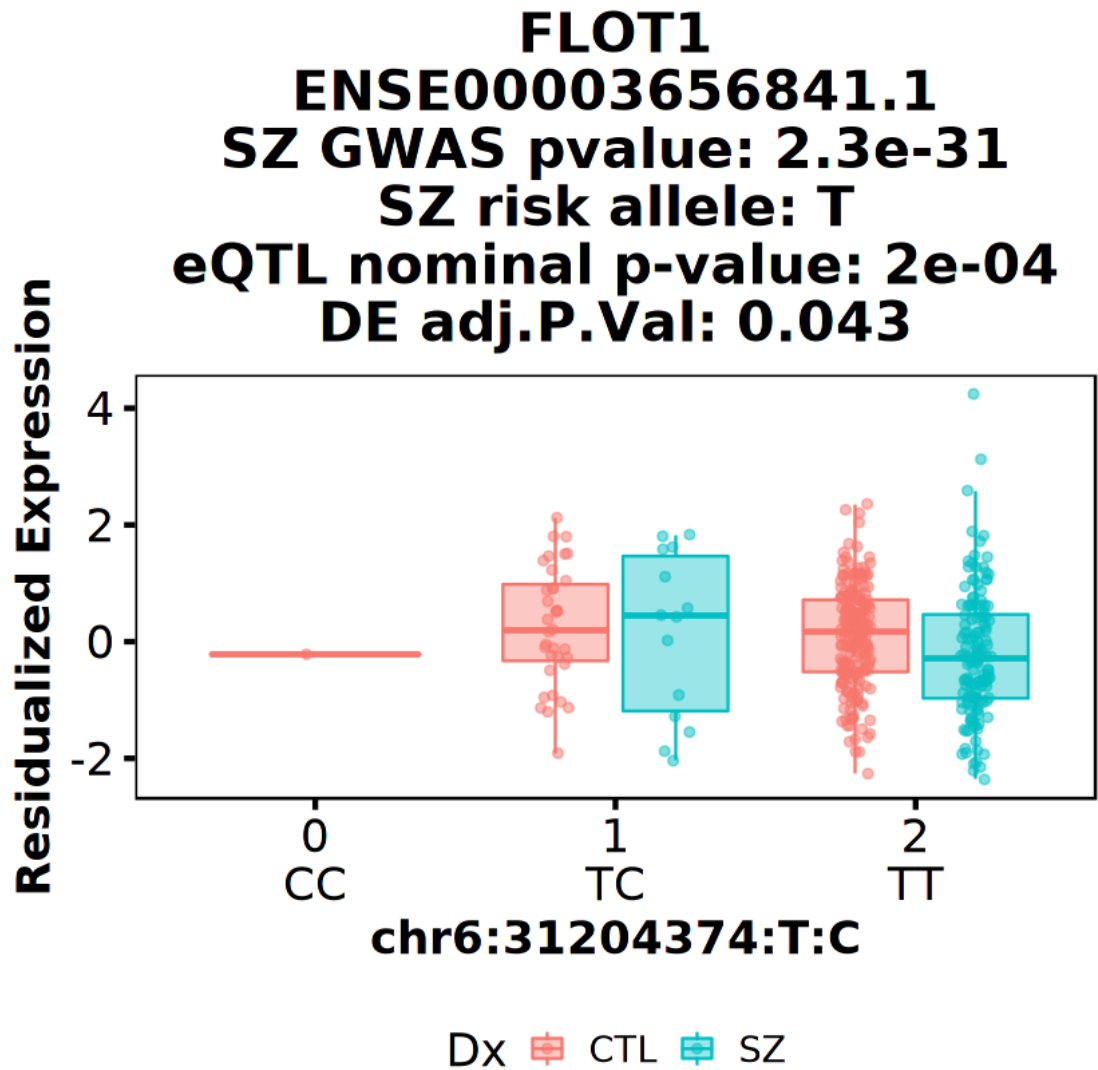
ZSCAN26
ENSE00001651520.2
SZ GWAS pvalue: 4.3e-44
SZ risk allele: A
eQTL nominal p-value: 7.3e-10
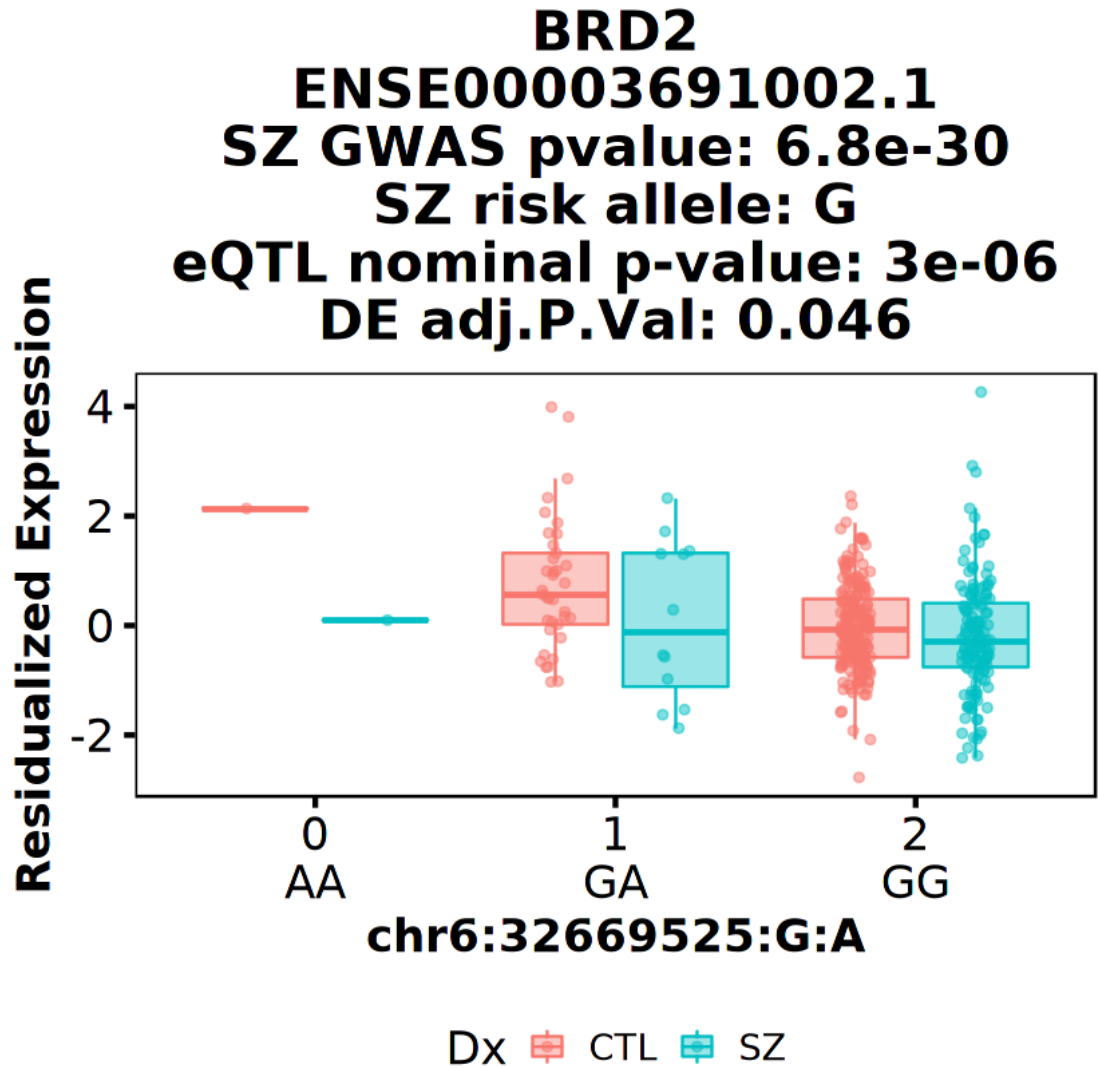**DE adj.P.Val: 0.019**

```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

ZNF391
ENSE00001733042.2
SZ GWAS pvalue: 6.5e-41
SZ risk allele: A
eQTL nominal p-value: 0.00017
DE adj.P.Val: 0.031

```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

HCG4
ENSE00001735745.1
SZ GWAS pvalue: 2.5e-39
SZ risk allele: G
eQTL nominal p-value: 6.9e-05
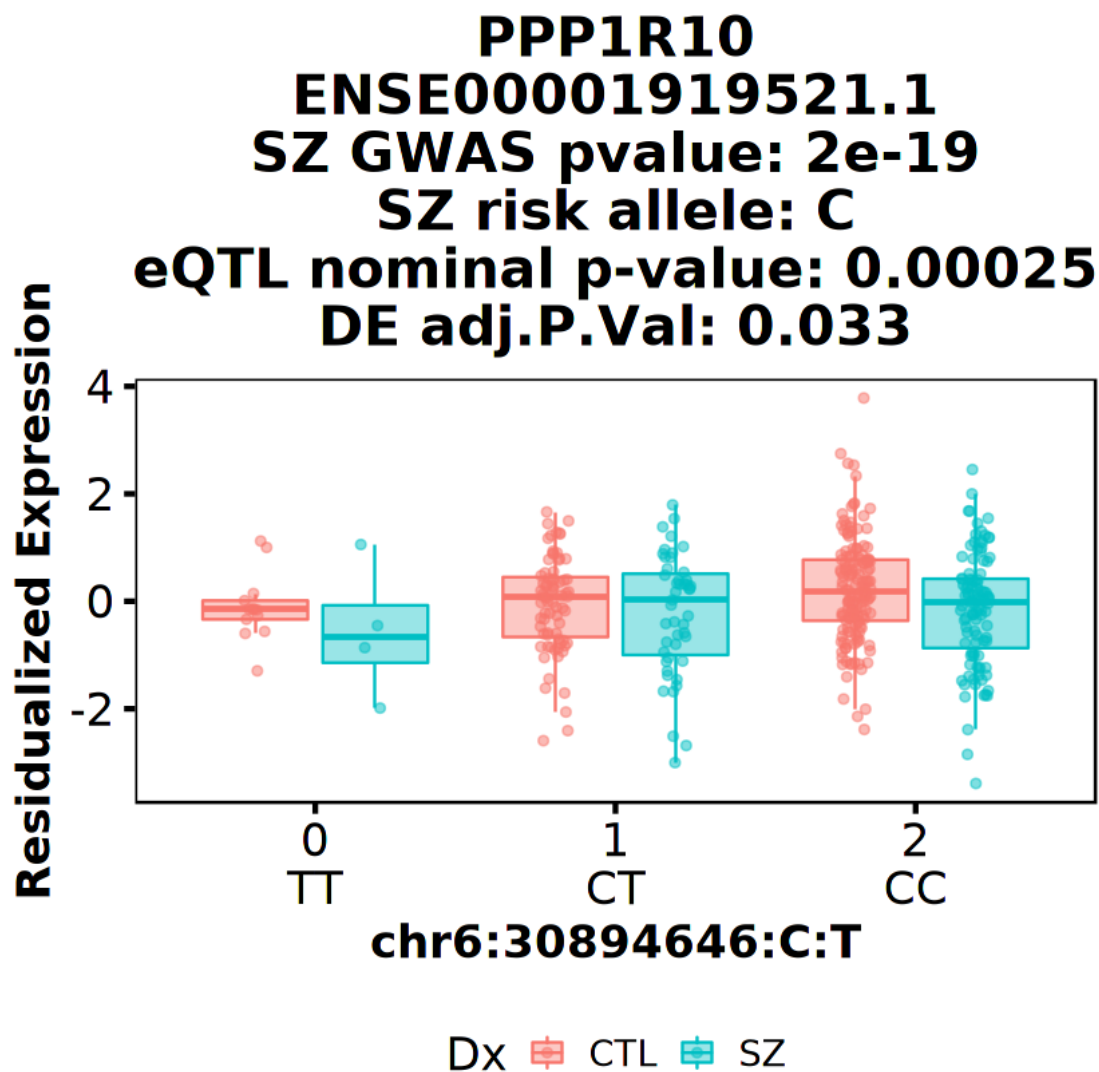**DE adj.P.Val: 0.019**

```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

FLOT1
ENSE00003656841.1
SZ GWAS pvalue: 2.3e-31
SZ risk allele: T
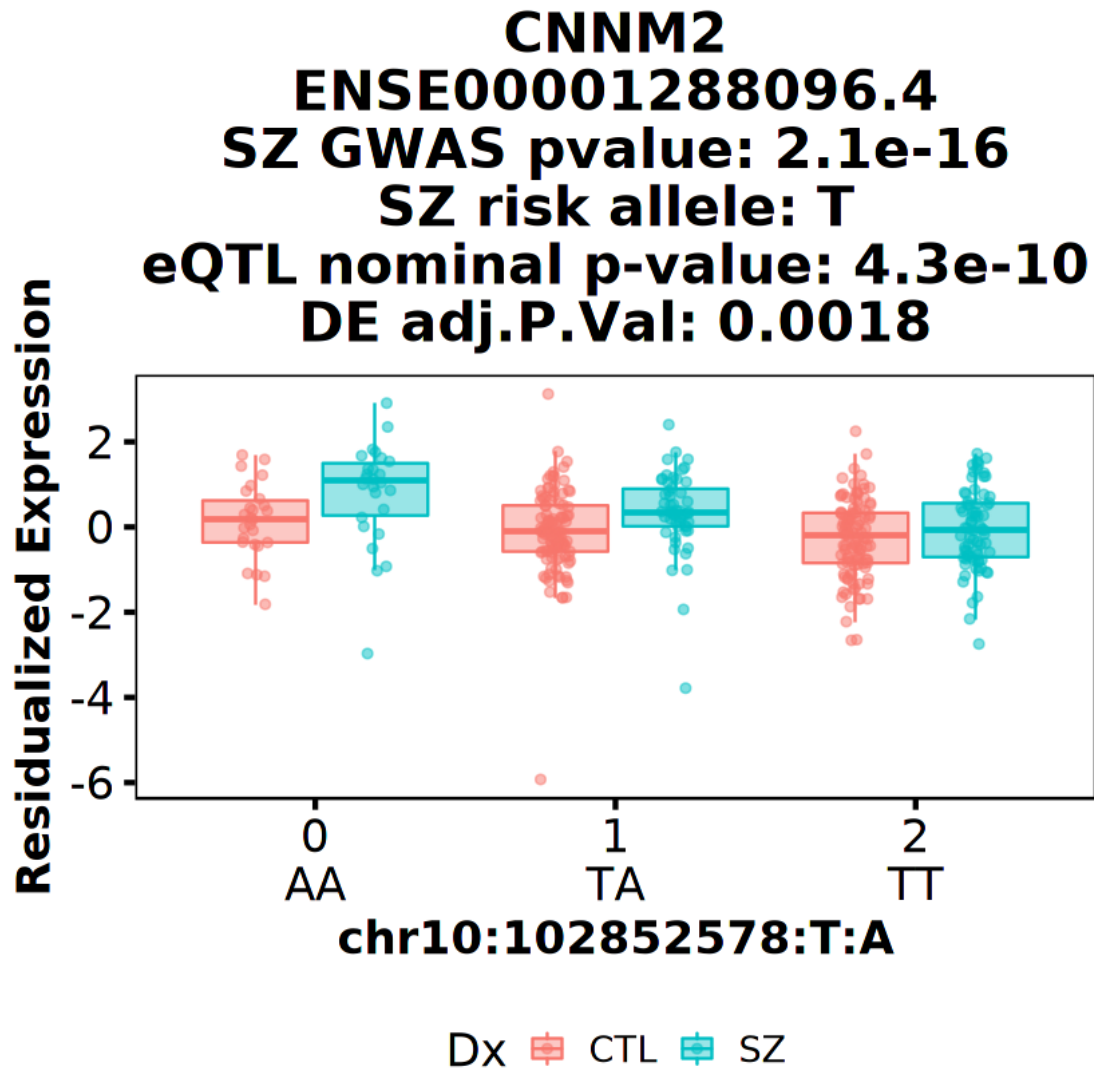## eQTL nominal p-value: 2e-04
DE adj.P.Val: 0.043

Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."

BRD2
ENSE00003691002.1
SZ GWAS pvalue: 6.8e-30
SZ risk allele: G
eQTL nominal p-value: 3e-06
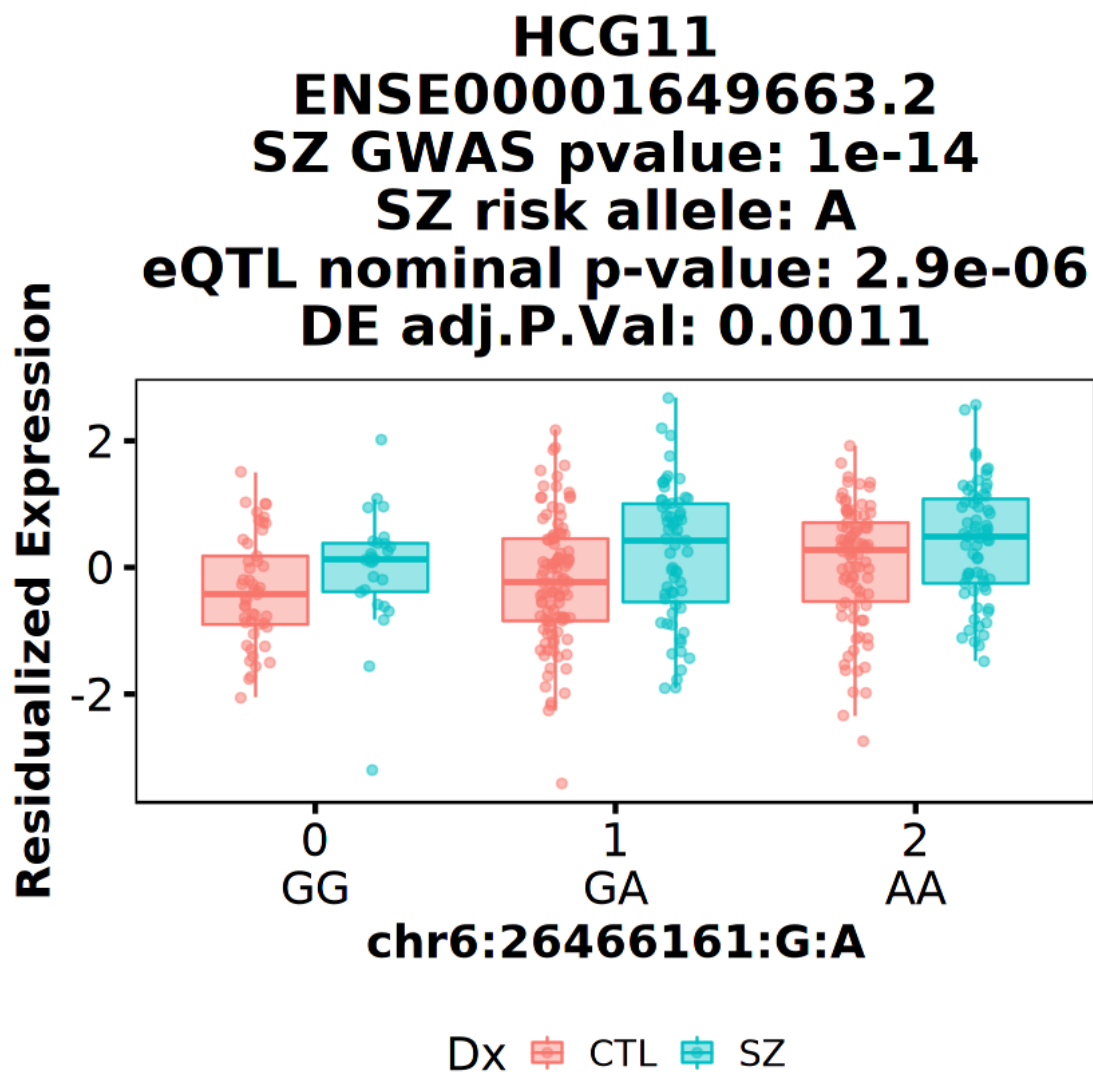## DE adj.P.Val: 0.046


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

PPP1R10
ENSE00001919521.1
SZ GWAS pvalue: 2e-19
SZ risk allele: C
eQTL nominal p-value: 0.00025
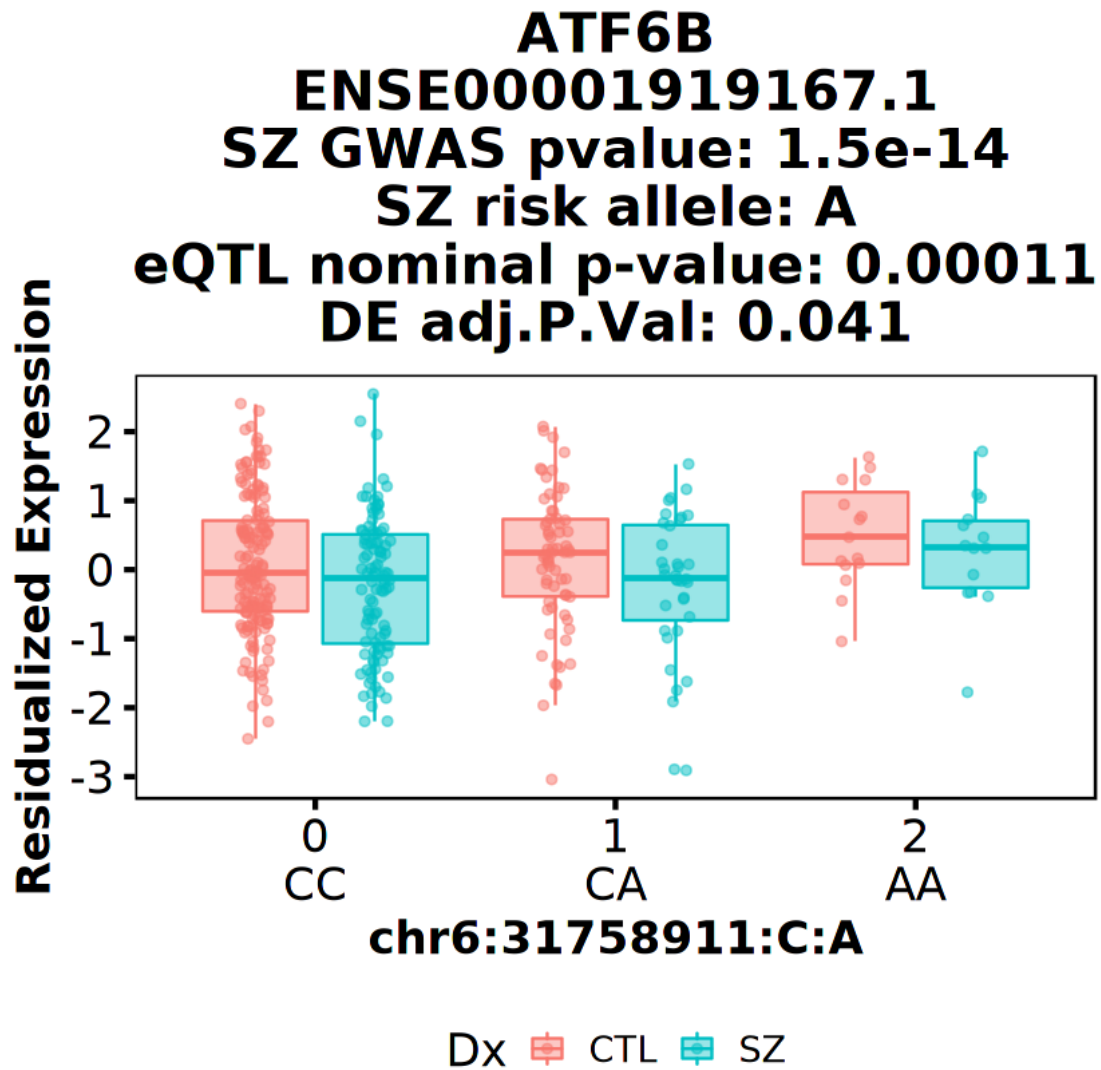**DE adj.P.Val: 0.033**

```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

CNNM2
ENSE00001288096.4
SZ GWAS pvalue: 2.1e-16
SZ risk allele: T
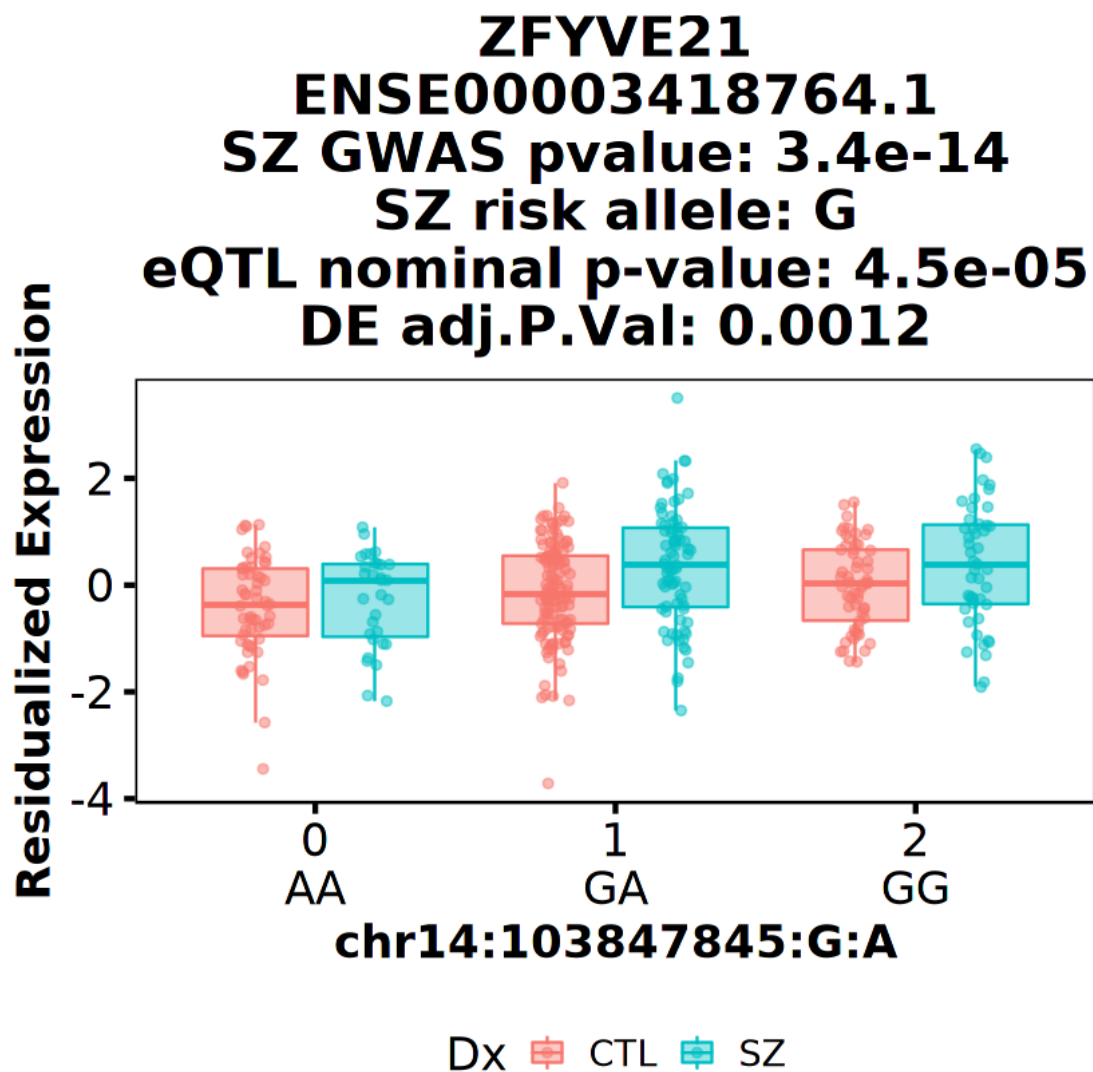# eQTL nominal p-value: 4.3e-10
DE adj.P.Val: 0.0018


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

HCG11
ENSE00001649663.2
SZ GWAS pvalue: 1e-14
SZ risk allele: A
eQTL nominal p-value: 2.9e-06
DE adj.P.Val: 0.0011
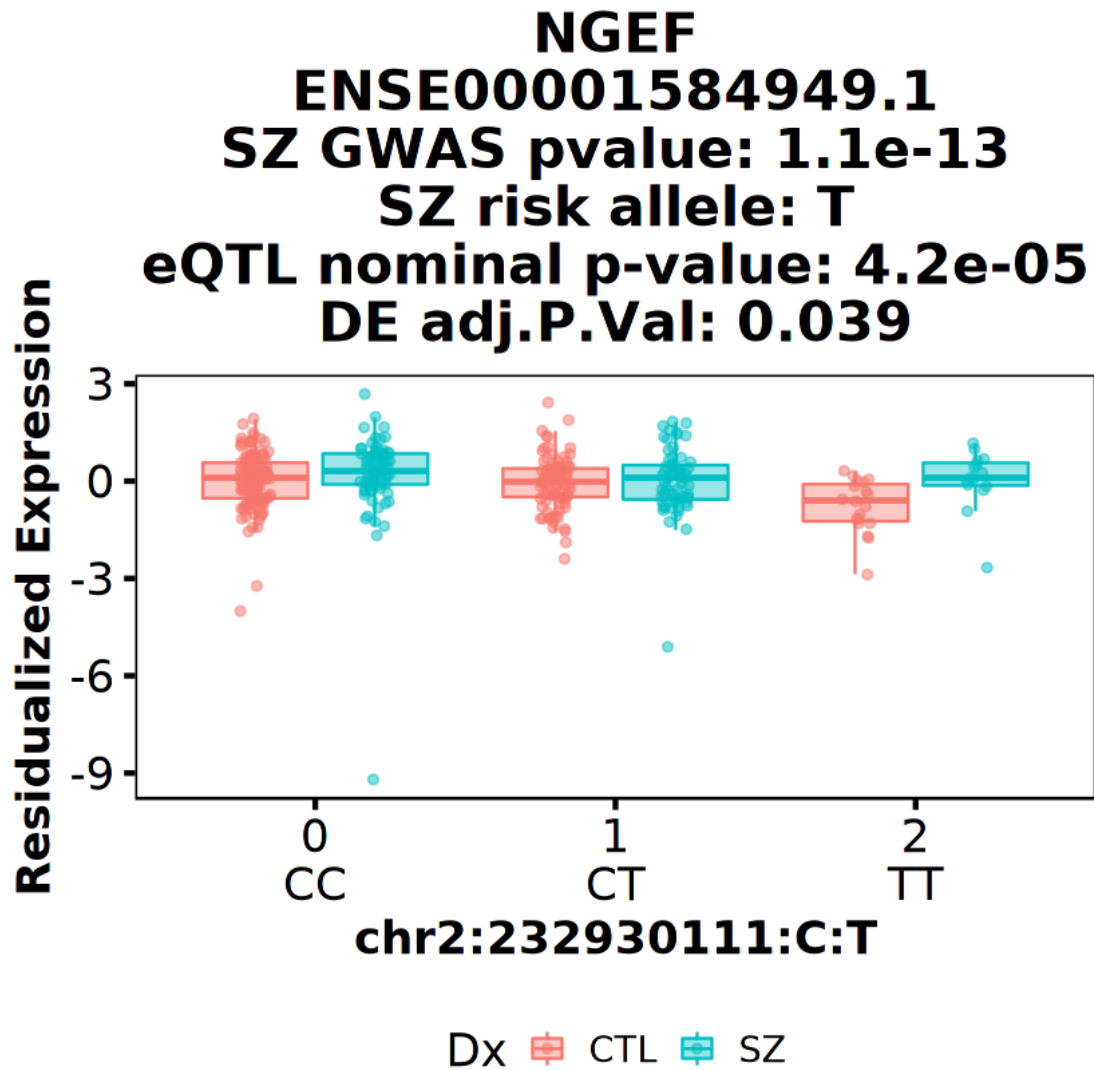
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."

**ATF6B
ENSE00001919167.1
SZ GWAS pvalue: 1.5e-14
SZ risk allele: A
eQTL nominal p-value: 0.00011
DE adj.P.Val: 0.041**

chr6:31758911:C:A

Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."

ZFYVE21
ENSE00003418764.1
SZ GWAS pvalue: 3.4e-14
SZ risk allele: G
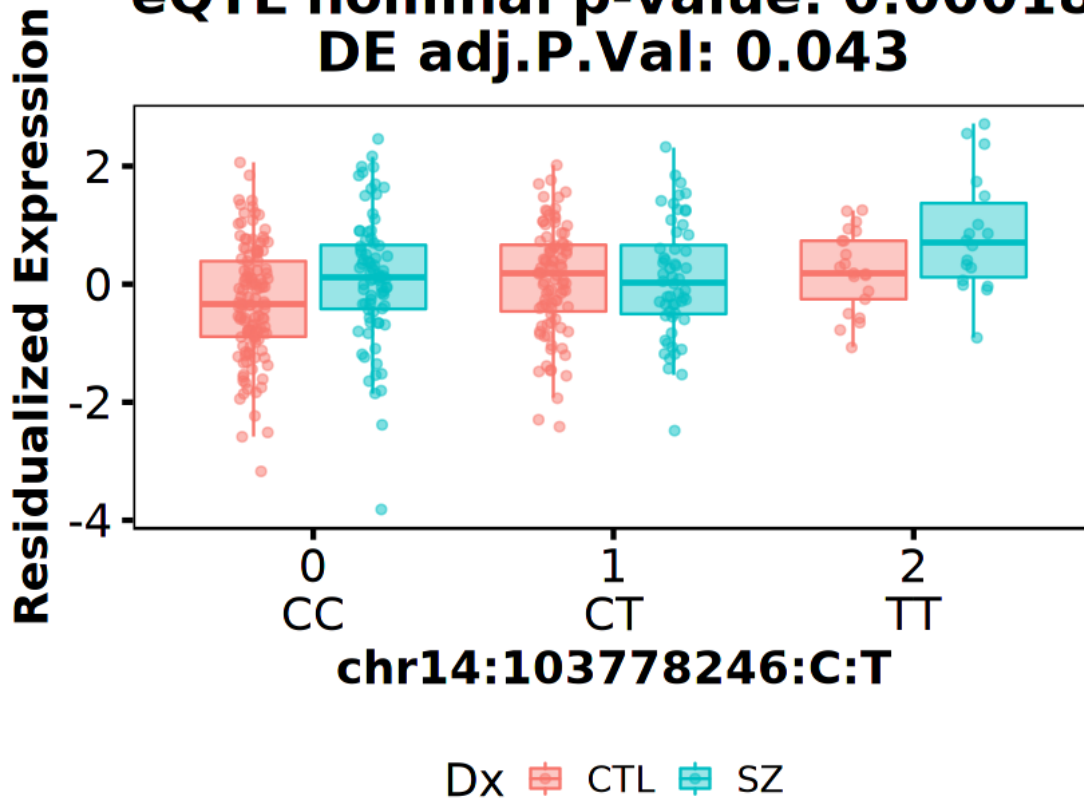# eQTL nominal p-value: 4.5e-05
DE adj.P.Val: 0.0012


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

**NGEF
ENSE00001584949.1
SZ GWAS pvalue: 1.1e-13
SZ risk allele: T
eQTL nominal p-value: 4.2e-05
DE adj.P.Val: 0.039**
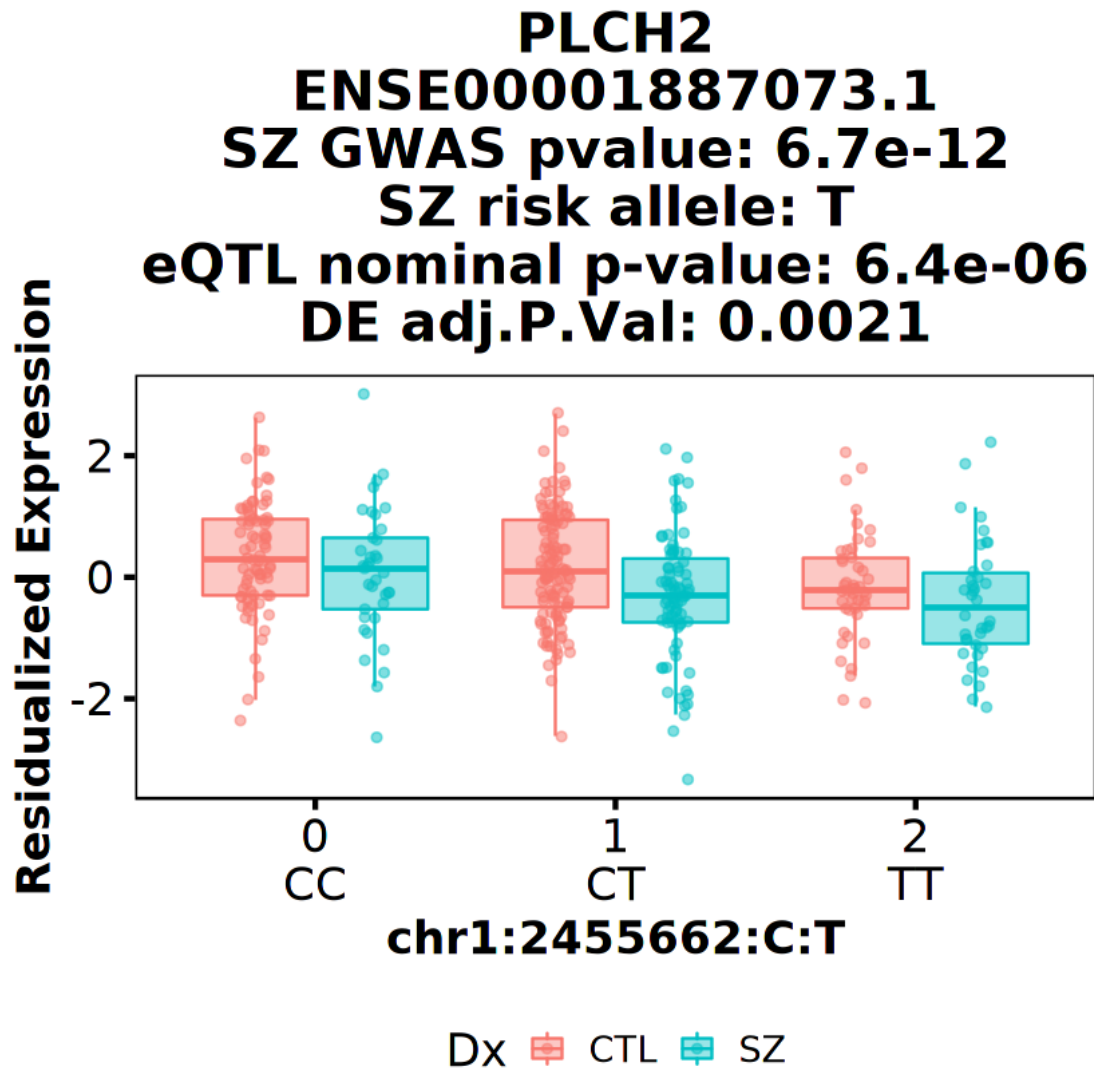
chr2:232930111:C:T

Dx ⊟ CTL ⊟ SZ

```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

AL049840.5
ENSE00003251427.1
SZ GWAS pvalue: 5.6e-13
SZ risk allele: T
eQTL nominal p-value: 0.00018
DE adj.P.Val: 0.043


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```
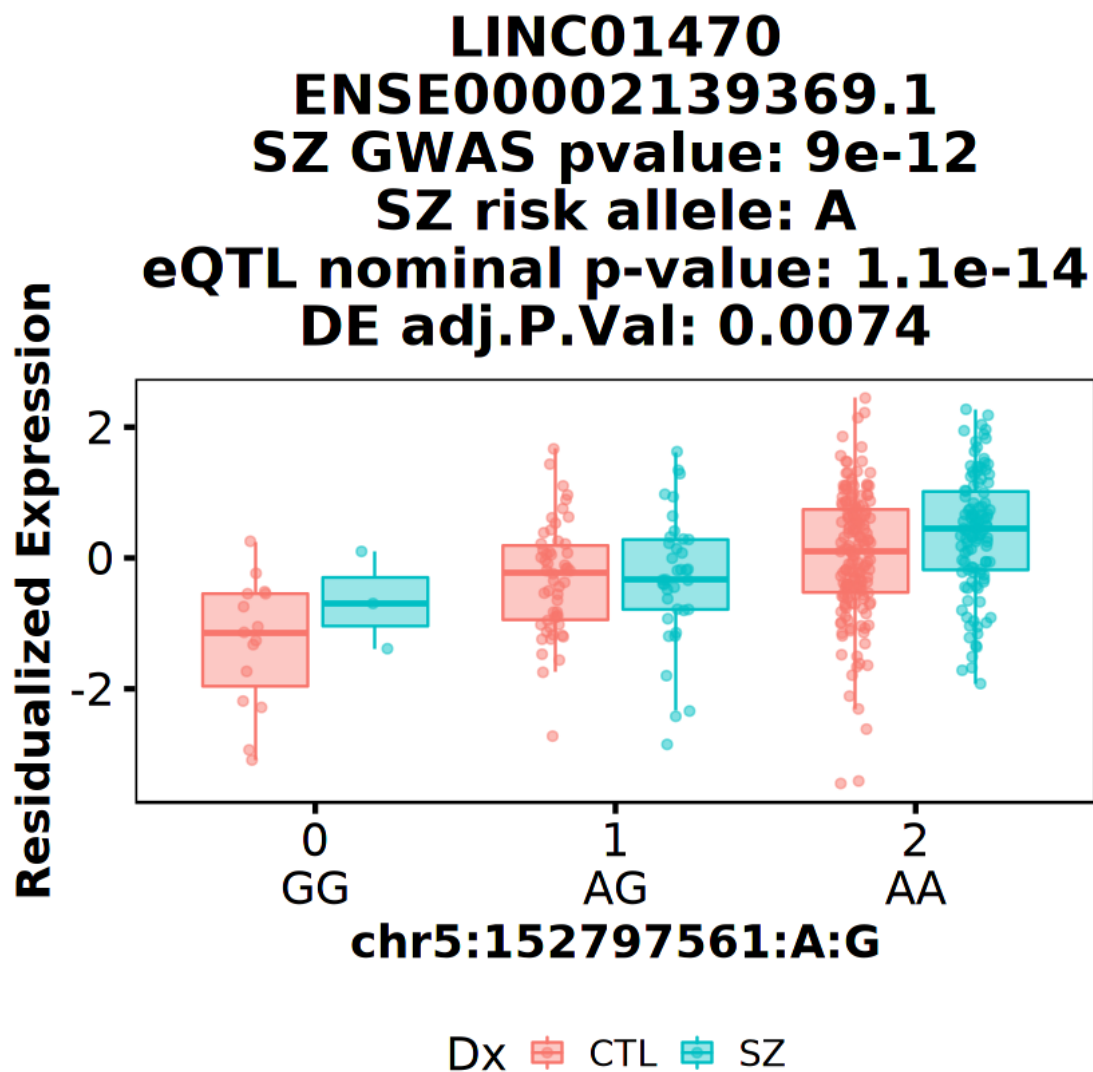
PLCH2
ENSE00001887073.1
SZ GWAS pvalue: 6.7e-12
SZ risk allele: T
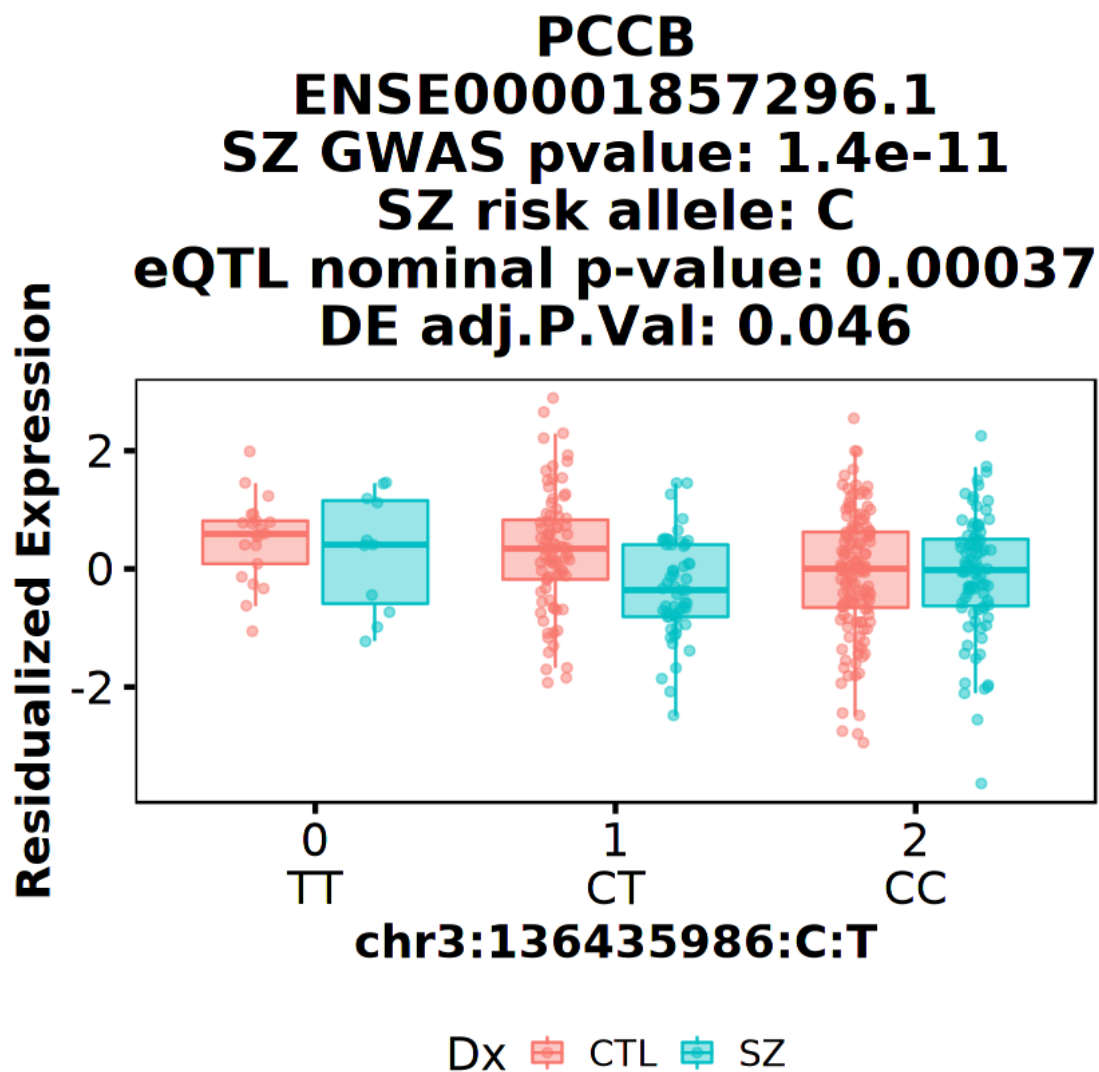## eQTL nominal p-value: 6.4e-06
DE adj.P.Val: 0.0021


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

LINC01470
ENSE00002139369.1
SZ GWAS pvalue: 9e-12
SZ risk allele: A
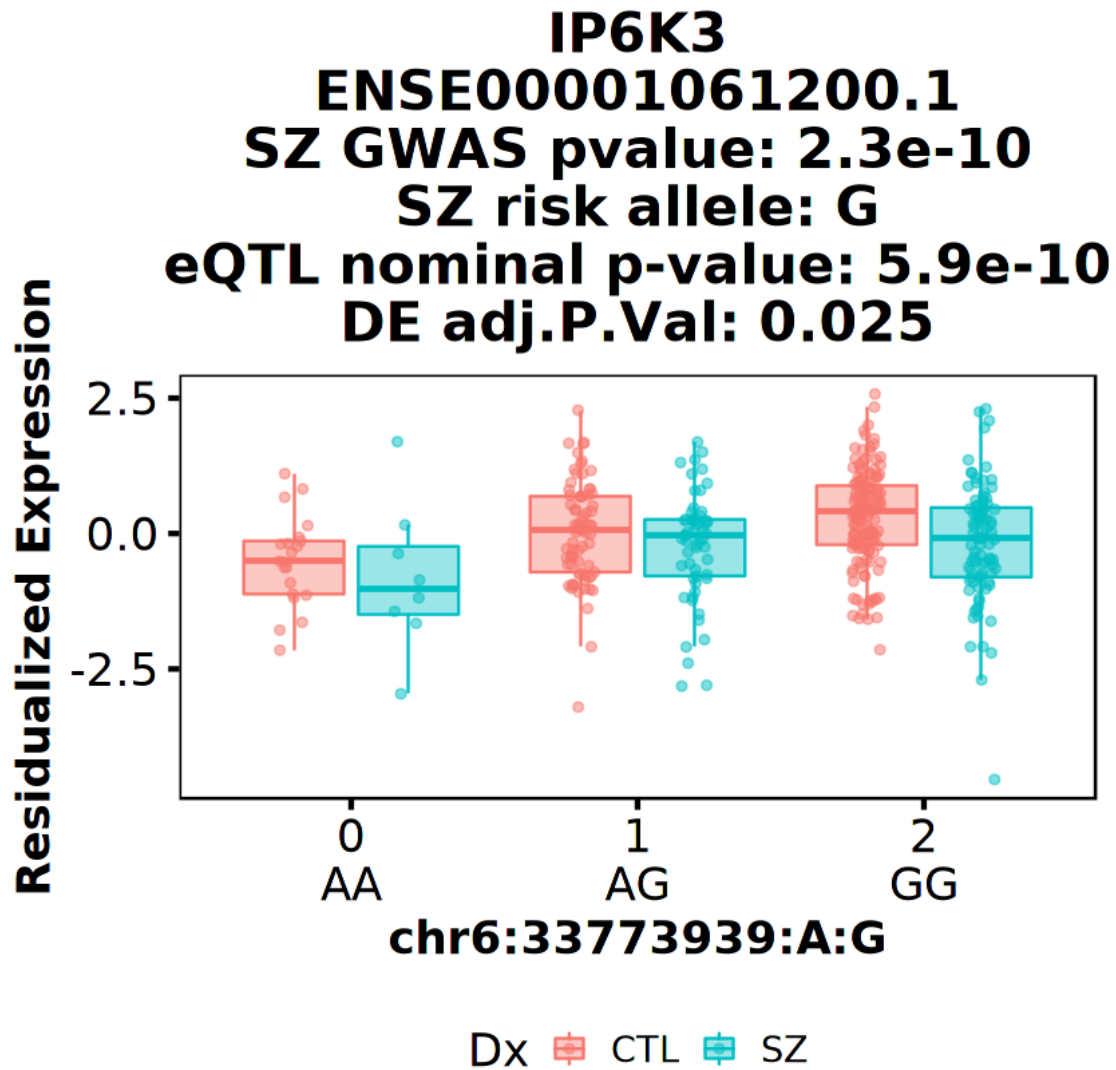# eQTL nominal p-value: 1.1e-14
DE adj.P.Val: 0.0074


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

PCCB
ENSE00001857296.1
SZ GWAS pvalue: 1.4e-11
SZ risk allele: C
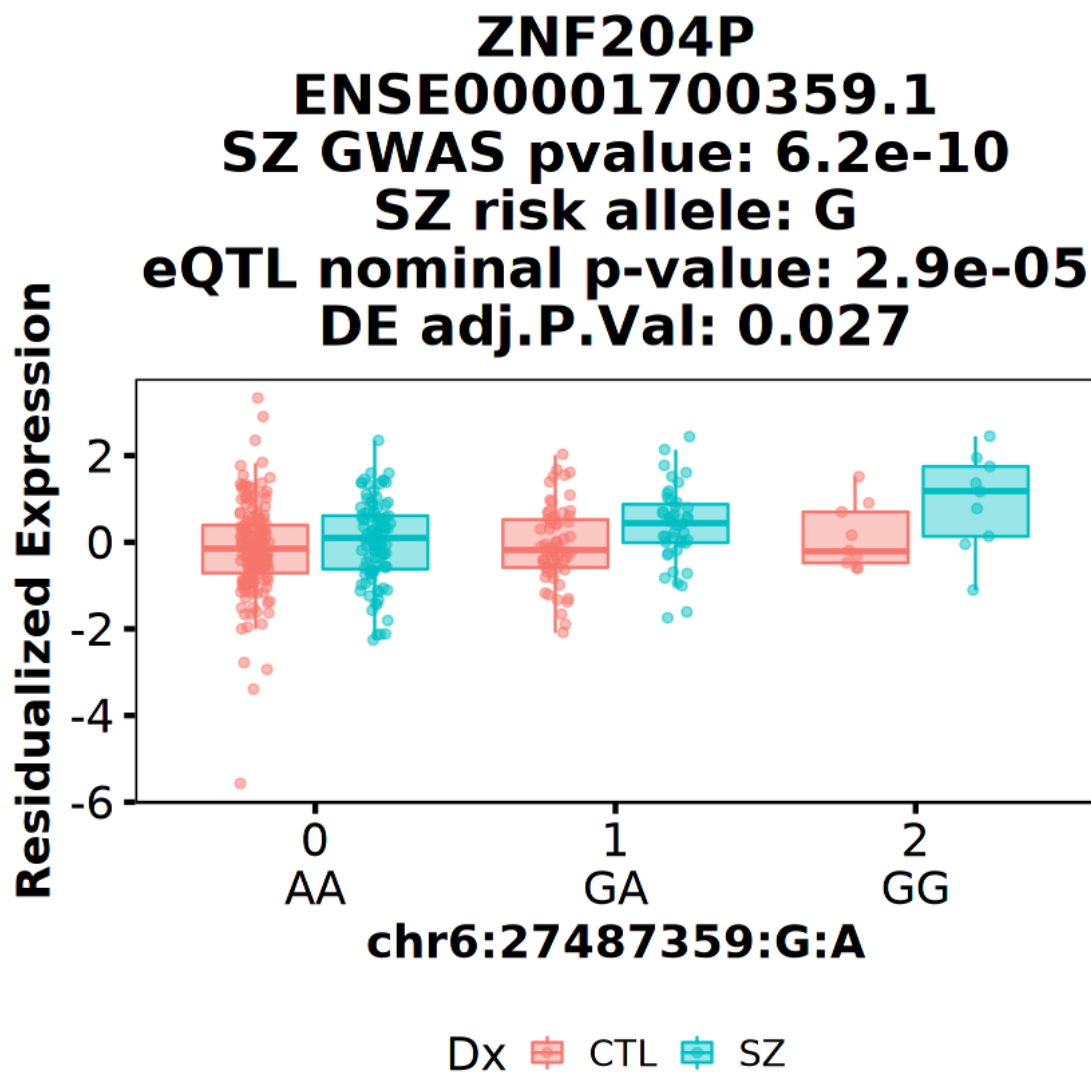## eQTL nominal p-value: 0.00037
DE adj.P.Val: 0.046


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

IP6K3
ENSE00001061200.1
SZ GWAS pvalue: 2.3e-10
SZ risk allele: G
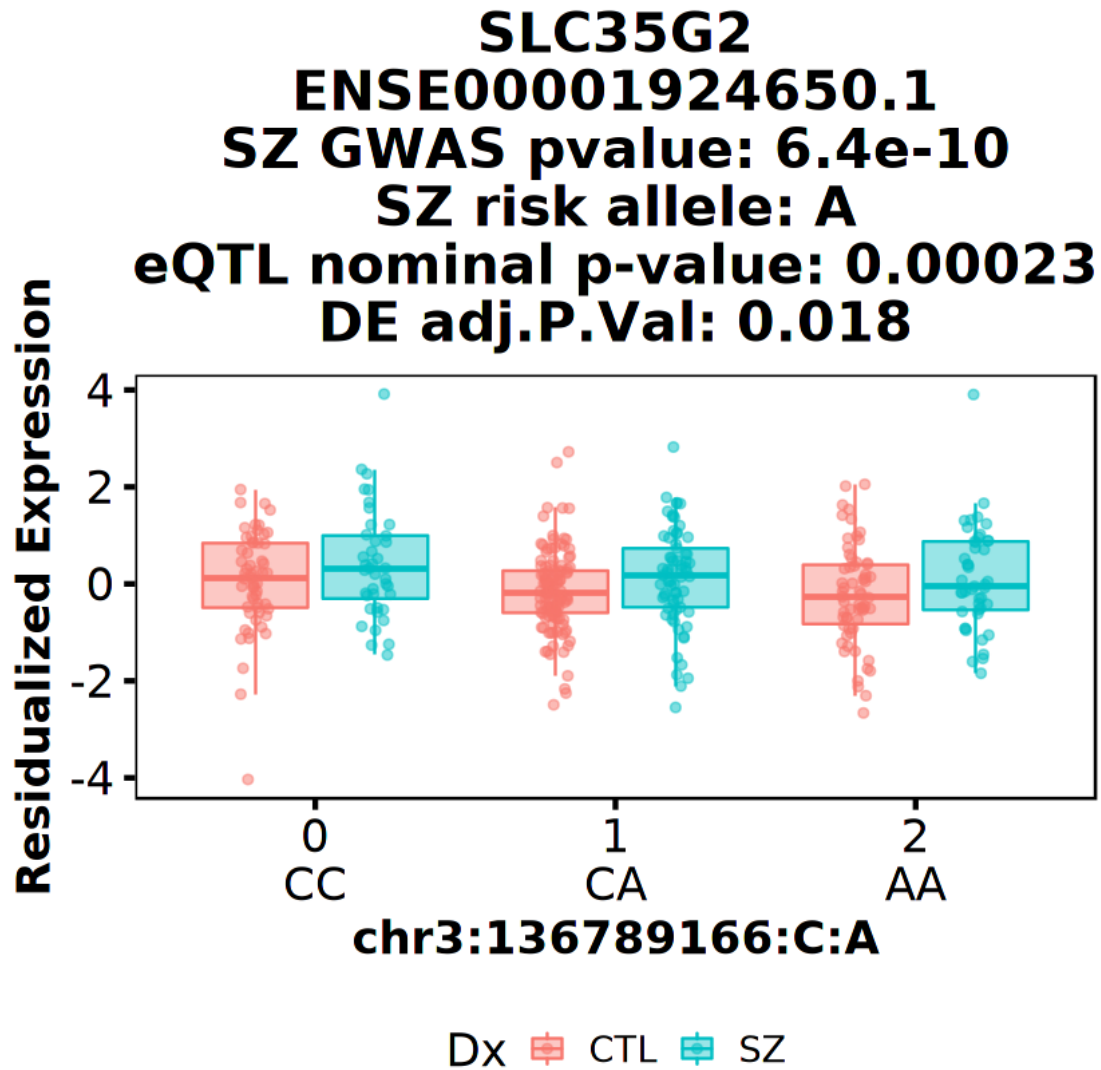## eQTL nominal p-value: 5.9e-10
DE adj.P.Val: 0.025


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

ZNF204P
ENSE00001700359.1
SZ GWAS pvalue: 6.2e-10
SZ risk allele: G
eQTL nominal p-value: 2.9e-05
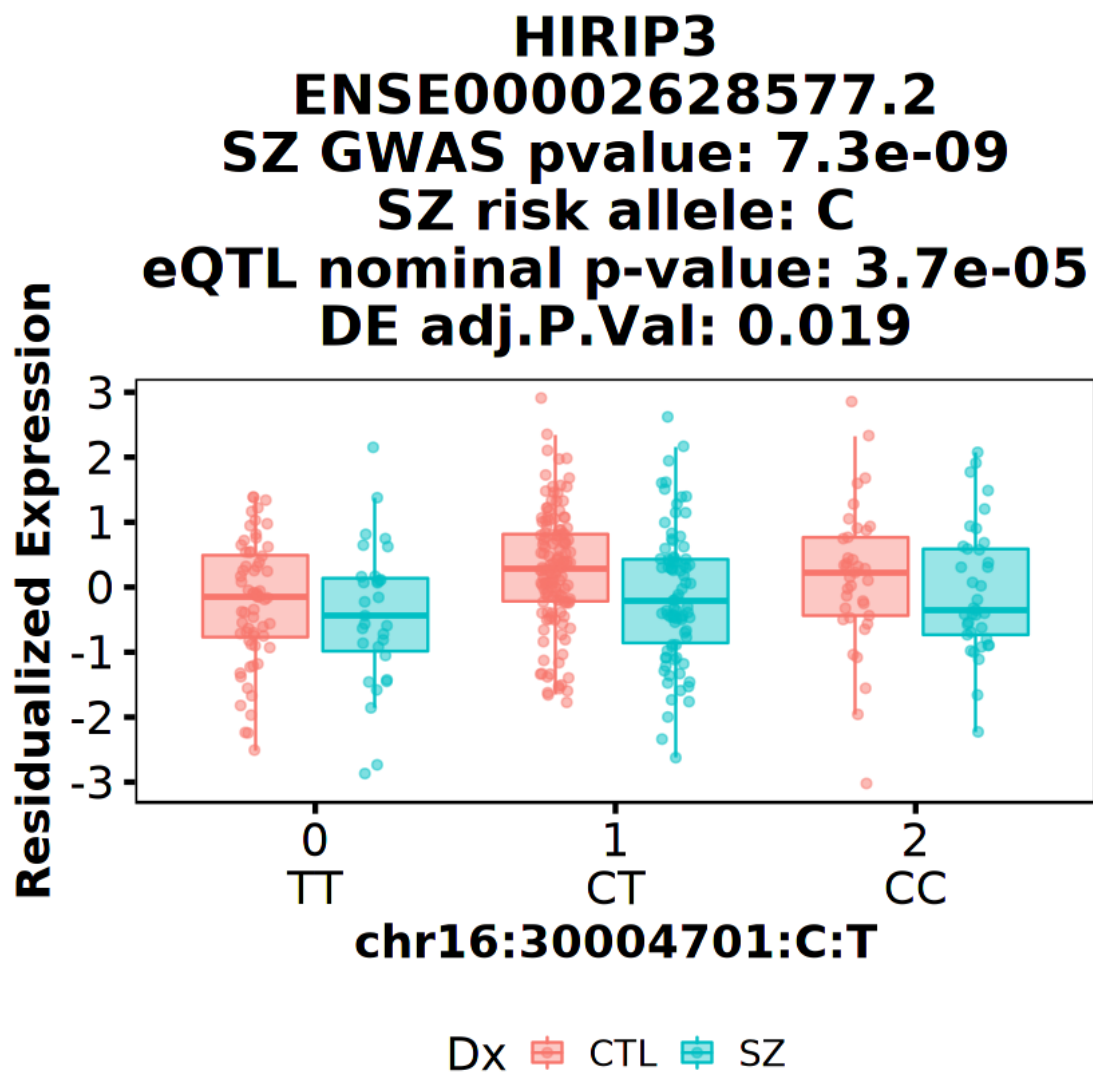**DE adj.P.Val: 0.027**

```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

SLC35G2
ENSE00001924650.1
SZ GWAS pvalue: 6.4e-10
SZ risk allele: A
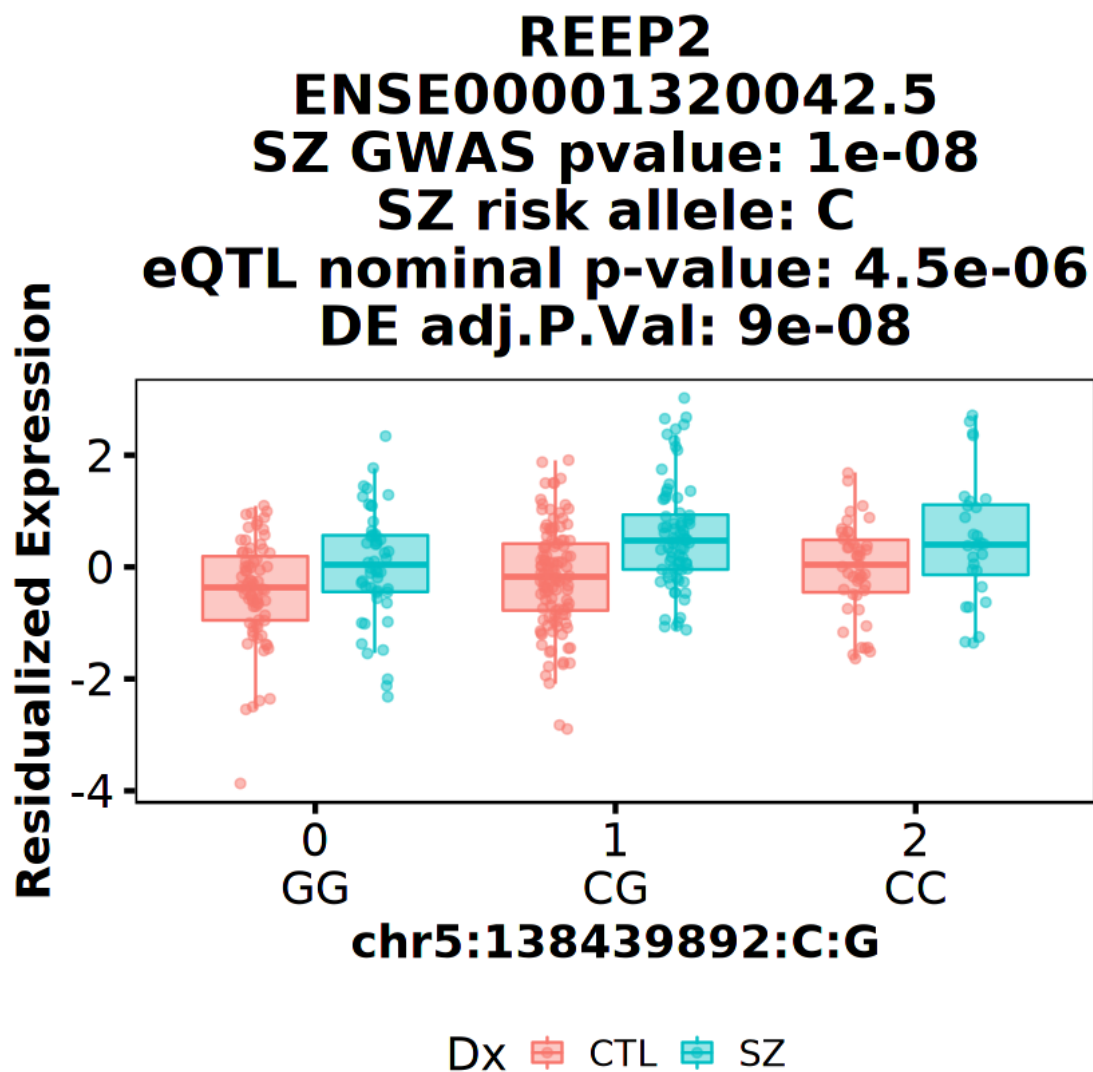# eQTL nominal p-value: 0.00023
DE adj.P.Val: 0.018


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

HIRIP3
ENSE00002628577.2
SZ GWAS pvalue: 7.3e-09
SZ risk allele: C
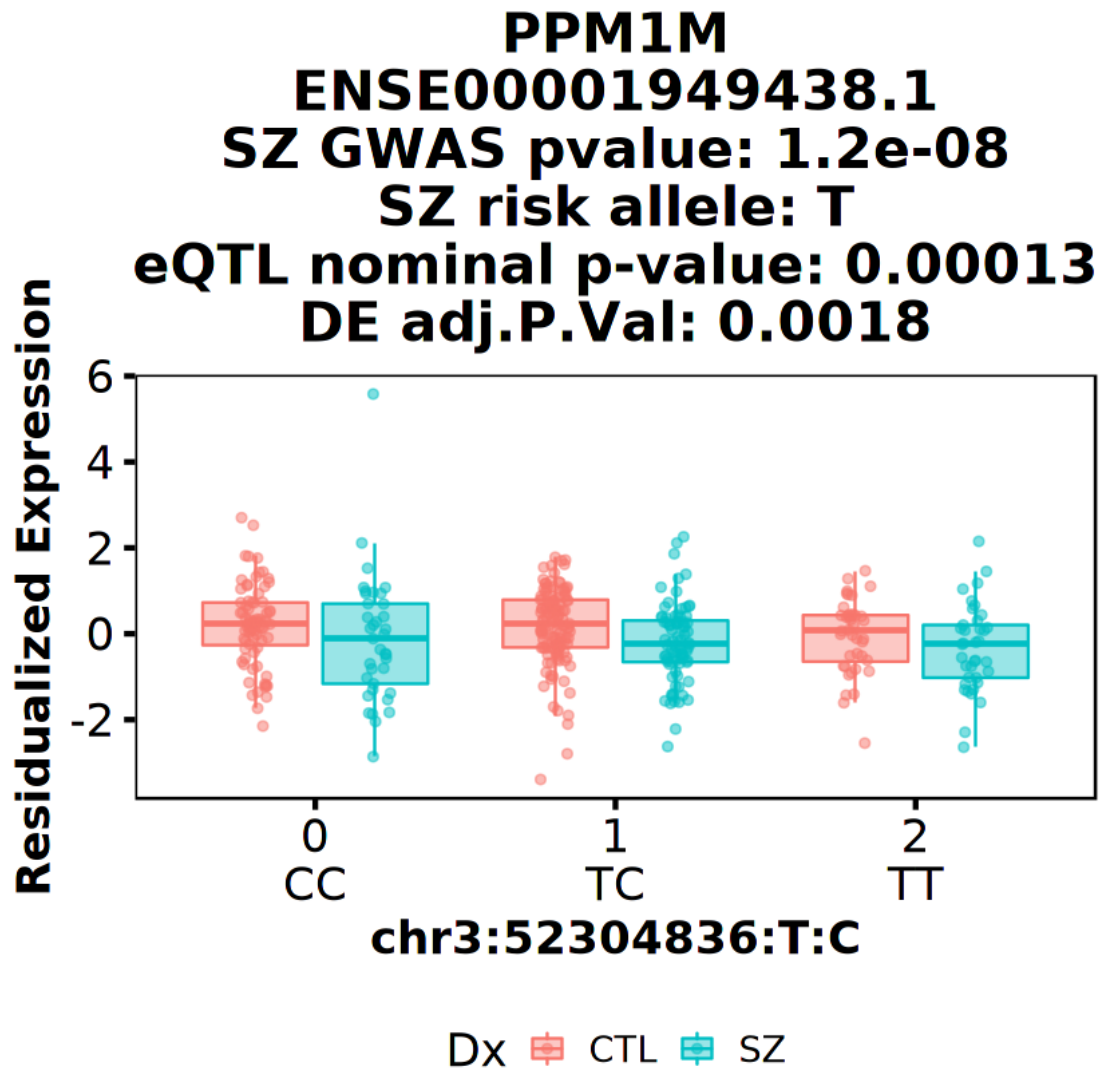eQTL nominal p-value: 3.7e-05
DE adj.P.Val: 0.019

Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."

REEP2
ENSE00001320042.5
SZ GWAS pvalue: 1e-08
SZ risk allele: C
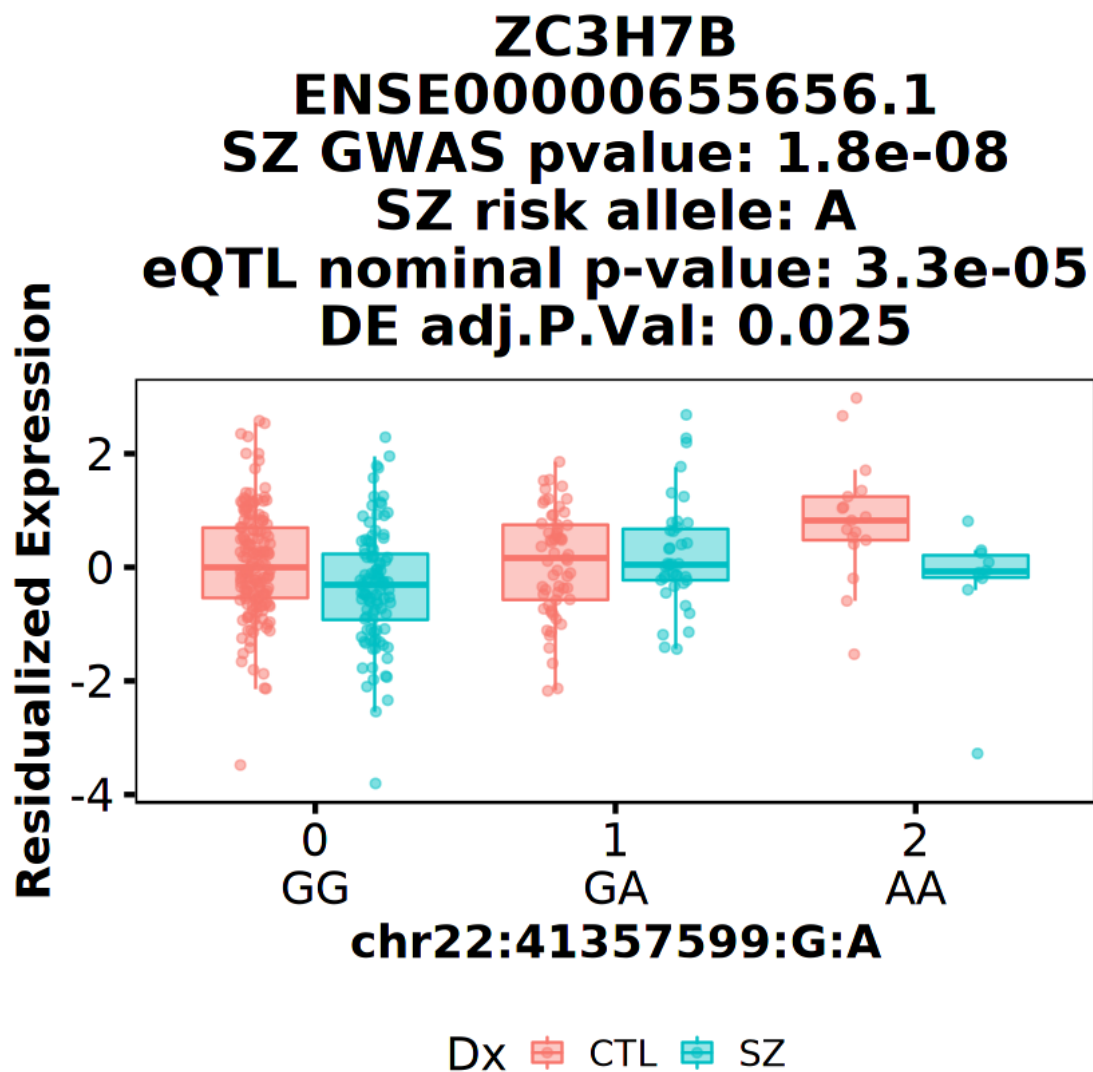## eQTL nominal p-value: 4.5e-06
DE adj.P.Val: 9e-08


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

PPM1M
ENSE00001949438.1
SZ GWAS pvalue: 1.2e-08
SZ risk allele: T
eQTL nominal p-value: 0.00013
DE adj.P.Val: 0.0018


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```

ZC3H7B
ENSE00000655656.1
SZ GWAS pvalue: 1.8e-08
SZ risk allele: A
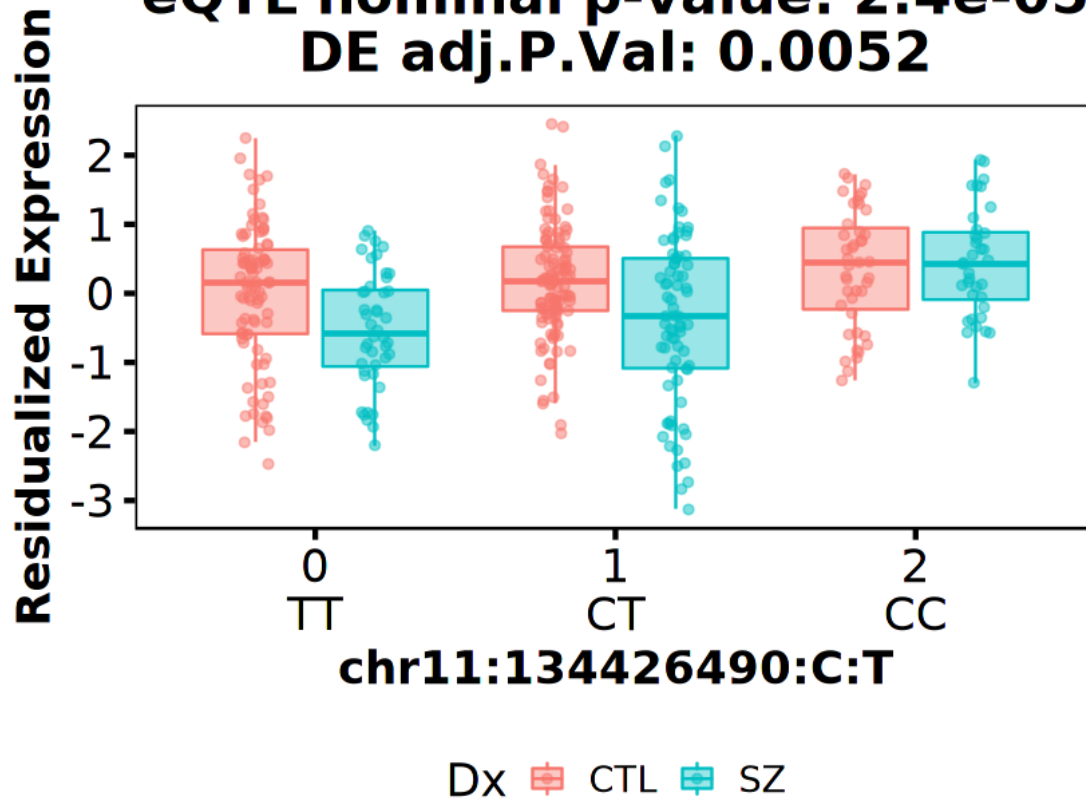## eQTL nominal p-value: 3.3e-05
DE adj.P.Val: 0.025


```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."
```
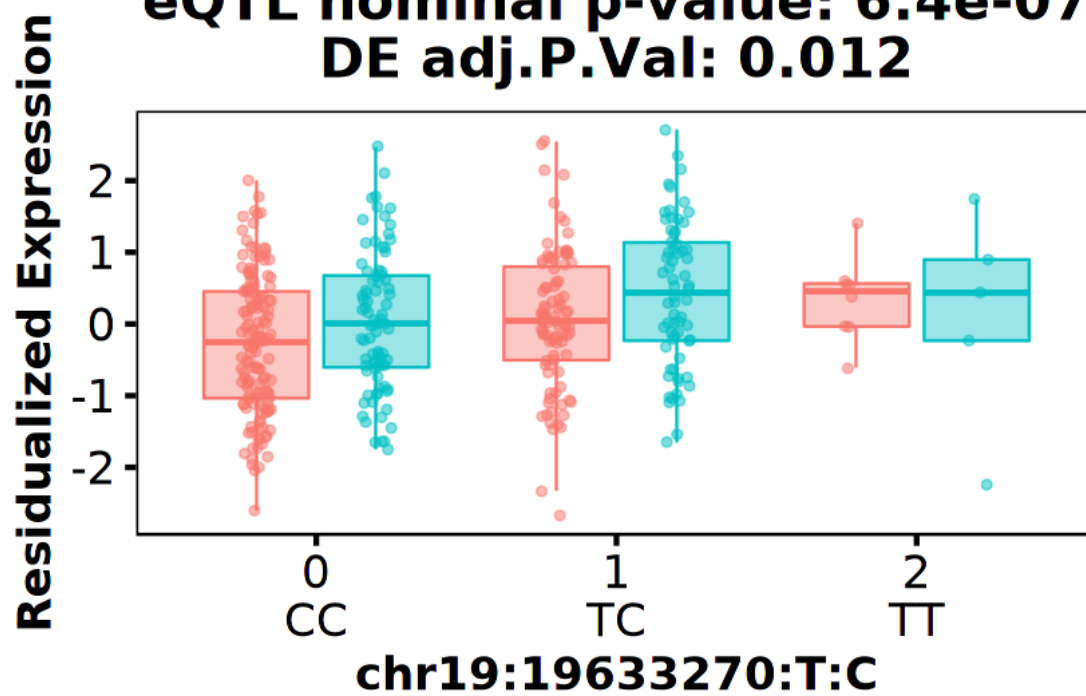
B3GAT1
ENSE00002197011.1
SZ GWAS pvalue: 3e-08
SZ risk allele: C
eQTL nominal p-value: 2.4e-05
DE adj.P.Val: 0.0052

ZNF14
ENSE00001407032.2
SZ GWAS pvalue: 3.1e-08
SZ risk allele: T
eQTL nominal p-value: 6.4e-07
DE adj.P.Val: 0.012

chr19:19633270:T:C

Dx ▉ CTL ▉ SZ

SLC7A6
ENSE00001616168.2
SZ GWAS pvalue: 3.6e-08
SZ risk allele: G
eQTL nominal p-value: 0.00013
DE adj.P.Val: 0.024
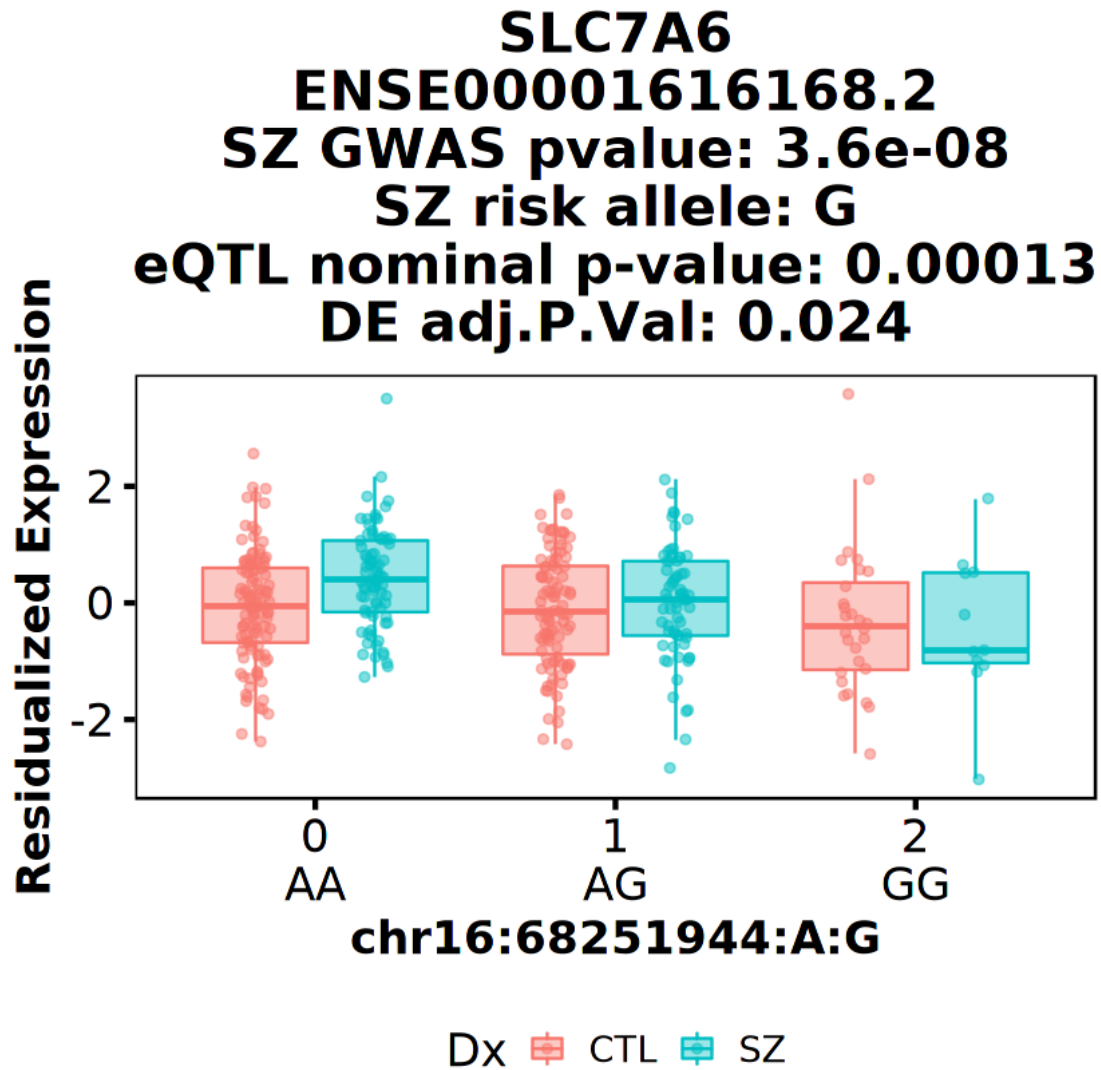
1.3 Session Info

```
[15]: Sys.time()
      proc.time()
      options(width = 120)
      sessioninfo::session_info()
```

```
[1] "2021-09-08 11:31:59 EDT"

     user     system    elapsed
10734.485   5689.653   2323.851

  Session info
 setting   value
```

```
version   R version 4.0.3 (2020-10-10)
os        Arch Linux
system    x86_64, linux-gnu
ui        X11
language  (EN)
collate   en_US.UTF-8
ctype     en_US.UTF-8
tz        America/New_York
date      2021-09-08

 Packages
package      * version  date       lib source
abind          1.4-5    2016-07-21 [1] CRAN (R 4.0.2)
assertthat     0.2.1    2019-03-21 [1] CRAN (R 4.0.2)
backports      1.2.1    2020-12-09 [1] CRAN (R 4.0.2)
base64enc      0.1-3    2015-07-28 [1] CRAN (R 4.0.2)
broom          0.7.9    2021-07-27 [1] CRAN (R 4.0.3)
cachem         1.0.6    2021-08-19 [1] CRAN (R 4.0.3)
Cairo          1.5-12.2 2020-07-07 [1] CRAN (R 4.0.2)
car            3.0-11   2021-06-27 [1] CRAN (R 4.0.3)
carData        3.0-4    2020-05-22 [1] CRAN (R 4.0.2)
cellranger     1.1.0    2016-07-27 [1] CRAN (R 4.0.2)
cli            3.0.1    2021-07-17 [1] CRAN (R 4.0.3)
colorspace     2.0-2    2021-06-24 [1] CRAN (R 4.0.3)
crayon         1.4.1    2021-02-08 [1] CRAN (R 4.0.3)
curl           4.3.2    2021-06-23 [1] CRAN (R 4.0.3)
data.table     1.14.0   2021-02-21 [1] CRAN (R 4.0.3)
DBI            1.1.1    2021-01-15 [1] CRAN (R 4.0.2)
dbplyr         2.1.1    2021-04-06 [1] CRAN (R 4.0.3)
digest         0.6.27   2020-10-24 [1] CRAN (R 4.0.2)
dplyr        * 1.0.7    2021-06-18 [1] CRAN (R 4.0.3)
ellipsis       0.3.2    2021-04-29 [1] CRAN (R 4.0.3)
evaluate       0.14     2019-05-28 [1] CRAN (R 4.0.2)
fansi          0.5.0    2021-05-25 [1] CRAN (R 4.0.3)
farver         2.1.0    2021-02-28 [1] CRAN (R 4.0.3)
fastmap        1.1.0    2021-01-25 [1] CRAN (R 4.0.2)
forcats      * 0.5.1    2021-01-27 [1] CRAN (R 4.0.2)
foreign        0.8-80   2020-05-24 [2] CRAN (R 4.0.3)
fs             1.5.0    2020-07-31 [1] CRAN (R 4.0.2)
generics       0.1.0    2020-10-31 [1] CRAN (R 4.0.2)
ggplot2      * 3.3.5    2021-06-25 [1] CRAN (R 4.0.3)
ggpubr       * 0.4.0    2020-06-27 [1] CRAN (R 4.0.2)
ggsignif       0.6.2    2021-06-14 [1] CRAN (R 4.0.3)
glue           1.4.2    2020-08-27 [1] CRAN (R 4.0.2)
gtable         0.3.0    2019-03-25 [1] CRAN (R 4.0.2)
haven          2.4.3    2021-08-04 [1] CRAN (R 4.0.3)
hms            1.1.0    2021-05-17 [1] CRAN (R 4.0.3)
htmltools      0.5.2    2021-08-25 [1] CRAN (R 4.0.3)
```

```
httr            1.4.2    2020-07-20 [1] CRAN (R 4.0.2)
IRdisplay       1.0      2021-01-20 [1] CRAN (R 4.0.2)
IRkernel        1.2      2021-05-11 [1] CRAN (R 4.0.3)
jsonlite        1.7.2    2020-12-09 [1] CRAN (R 4.0.2)
labeling        0.4.2    2020-10-20 [1] CRAN (R 4.0.2)
lifecycle       1.0.0    2021-02-15 [1] CRAN (R 4.0.3)
lubridate       1.7.10   2021-02-26 [1] CRAN (R 4.0.3)
magrittr        2.0.1    2020-11-17 [1] CRAN (R 4.0.2)
memoise         2.0.0    2021-01-26 [1] CRAN (R 4.0.2)
modelr          0.1.8    2020-05-19 [1] CRAN (R 4.0.2)
munsell         0.5.0    2018-06-12 [1] CRAN (R 4.0.2)
openxlsx        4.2.4    2021-06-16 [1] CRAN (R 4.0.3)
pbdZMQ          0.3-5    2021-02-10 [1] CRAN (R 4.0.3)
pillar          1.6.2    2021-07-29 [1] CRAN (R 4.0.3)
pkgconfig       2.0.3    2019-09-22 [1] CRAN (R 4.0.2)
purrr         * 0.3.4    2020-04-17 [1] CRAN (R 4.0.2)
R.methodsS3     1.8.1    2020-08-26 [1] CRAN (R 4.0.3)
R.oo            1.24.0   2020-08-26 [1] CRAN (R 4.0.3)
R.utils         2.10.1   2020-08-26 [1] CRAN (R 4.0.3)
R6              2.5.1    2021-08-19 [1] CRAN (R 4.0.3)
Rcpp            1.0.7    2021-07-07 [1] CRAN (R 4.0.3)
readr         * 2.0.1    2021-08-10 [1] CRAN (R 4.0.3)
readxl          1.3.1    2019-03-13 [1] CRAN (R 4.0.2)
repr            1.1.3    2021-01-21 [1] CRAN (R 4.0.2)
reprex          2.0.1    2021-08-05 [1] CRAN (R 4.0.3)
rio             0.5.27   2021-06-21 [1] CRAN (R 4.0.3)
rlang           0.4.11   2021-04-30 [1] CRAN (R 4.0.3)
rstatix         0.7.0    2021-02-13 [1] CRAN (R 4.0.3)
rstudioapi      0.13     2020-11-12 [1] CRAN (R 4.0.2)
rvest           1.0.1    2021-07-26 [1] CRAN (R 4.0.3)
scales          1.1.1    2020-05-11 [1] CRAN (R 4.0.2)
sessioninfo     1.1.1    2018-11-05 [1] CRAN (R 4.0.2)
stringi         1.7.4    2021-08-25 [1] CRAN (R 4.0.3)
stringr       * 1.4.0    2019-02-10 [1] CRAN (R 4.0.2)
svglite         2.0.0    2021-02-20 [1] CRAN (R 4.0.3)
systemfonts     1.0.2    2021-05-11 [1] CRAN (R 4.0.3)
tibble        * 3.1.4    2021-08-25 [1] CRAN (R 4.0.3)
tidyr         * 1.1.3    2021-03-03 [1] CRAN (R 4.0.3)
tidyselect      1.1.1    2021-04-30 [1] CRAN (R 4.0.3)
tidyverse     * 1.3.1    2021-04-15 [1] CRAN (R 4.0.3)
tzdb            0.1.2    2021-07-20 [1] CRAN (R 4.0.3)
utf8            1.2.2    2021-07-24 [1] CRAN (R 4.0.3)
uuid            0.1-4    2020-02-26 [1] CRAN (R 4.0.2)
vctrs           0.3.8    2021-04-29 [1] CRAN (R 4.0.3)
withr           2.4.2    2021-04-18 [1] CRAN (R 4.0.3)
xml2            1.3.2    2020-04-23 [1] CRAN (R 4.0.2)
zip             2.2.0    2021-05-31 [1] CRAN (R 4.0.3)
```

```
[1]  /home/jbenja13/R/x86_64-pc-linux-gnu-library/4.0
[2]  /usr/lib/R/library
```