main_genes

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 = "genes"
```

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)
```

```
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_snp_df <- function(variant_id, gene_id){
    zz = get_geno_annot() %>% filter(SNP == variant_id)
    xx = get_snps_df() %>% 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

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

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

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

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)

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

[5]:
```
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)
}
```

1.2 Integration analysis

[6]:
```
dir.create(feature)
```

1.2.1 Enrichment

Integrate DEG with PGC2+CLOZUK SNPs

[7]:
```
dft = memMERGE() %>% mutate(agree_direction=ifelse(agree_direction == 1, "Yes",␣
↪ifelse(agree_direction == -1, "No", 0)))
dim(dft)
```

1. 1616724 2. 53

[8]:
```
table(dft$agree_direction)
```

```
     0     No    Yes
  1873 798695 816156
```

[9]:
```
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,]   3020   33767
[2,] 168656 1411281
```

```
        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:
 0.7205791 0.7770428
sample estimates:
odds ratio
 0.7483872
```

[10]:
```
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. 3020 2. 17

[11]:
```
df = dft2 %>% group_by(gene_id) %>% slice(1) %>% arrange(P)
table(df$agree_direction)
```

```
 No Yes
 15  20
```

[12]:
```
df
```

A grouped_df: 35 × 17

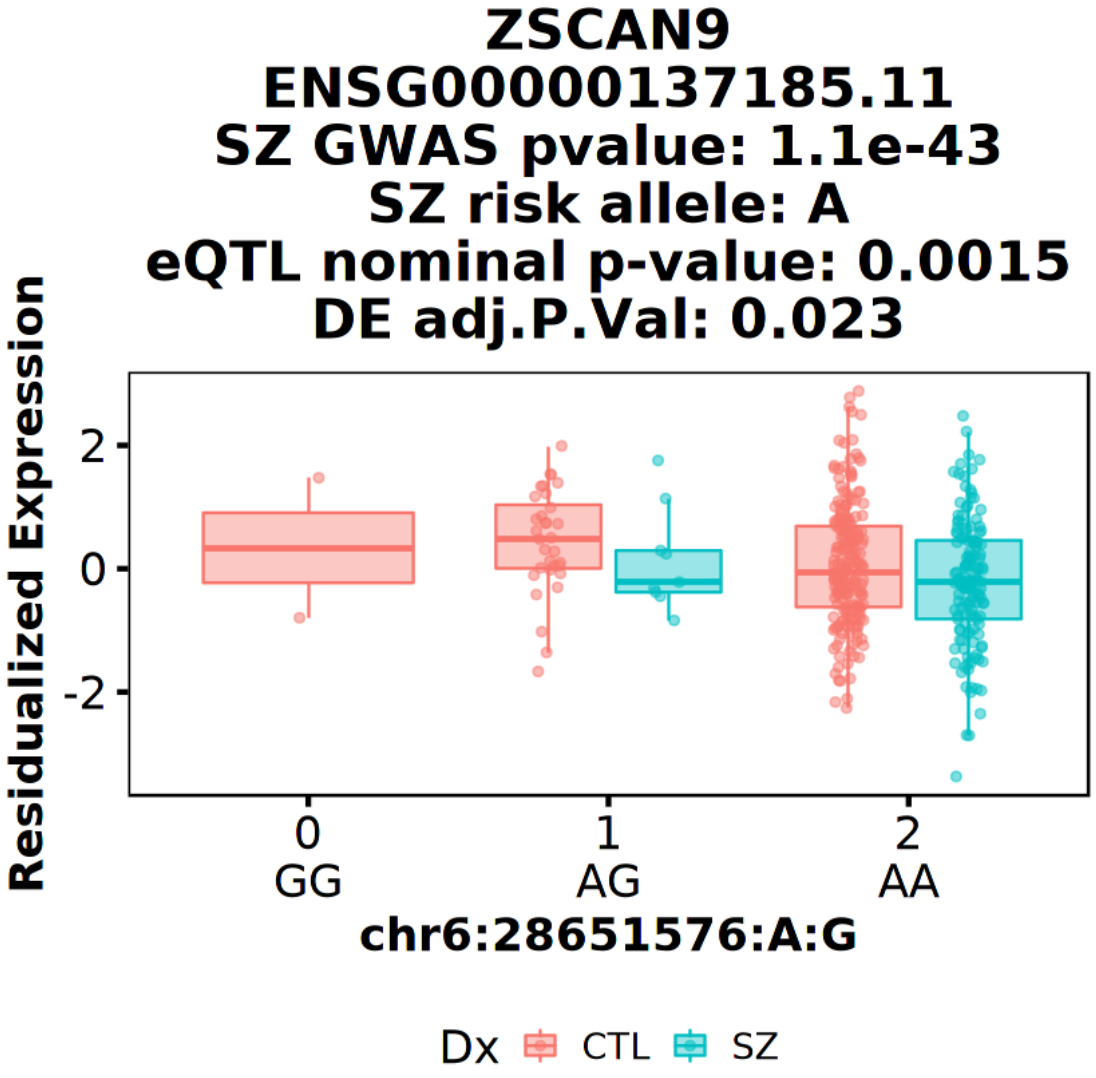
gene_id <chr>	Symbol <chr>	variant_id <chr>	rsid <chr>	A1 <c
ENSG00000137185.11	ZSCAN9	chr6:28651576:A:G	rs34724414	A
ENSG00000124613.8	ZNF391	chr6:27424023:C:T	rs34071253	C
ENSG00000176998.4	HCG4	chr6:29315895:C:T	rs3129682	C
ENSG00000244731.7	C4A	chr6:31793436:G:A	rs2607014	G
ENSG00000137312.14	FLOT1	chr6:30765081:C:T	rs3094118	C
ENSG00000204256.12	BRD2	chr6:32669525:G:A	rs2395231	G
ENSG00000204789.4	ZNF204P	chr6:27722160:T:C	rs2056923	T
ENSG00000148842.17	CNNM2	chr10:102852578:T:A	rs11191419	T
ENSG00000204463.12	BAG6	chr6:31358852:C:T	rs2523594	C
ENSG00000228223.2	HCG11	chr6:26466161:G:A	rs1977199	G
ENSG00000066248.14	NGEF	chr2:232925939:G:T	rs1878287	G
ENSG00000204356.12	NELFE	chr6:31868665:A:G	rs542418	A
ENSG00000112511.17	PHF1	chr6:33427422:G:A	rs9461856	G
ENSG00000204469.12	PRRC2A	chr6:31624151:A:G	rs2260050	A
ENSG00000114054.13	PCCB	chr3:136569563:G:A	rs7432375	G
ENSG00000166166.12	TRMT61A	chr14:103809712:C:T	rs12880821	C
ENSG00000149527.17	PLCH2	chr1:2455662:C:T	rs4648845	C
ENSG00000249484.8	LINC01470	chr5:152797561:A:G	rs111294930	A
ENSG00000171045.14	TSNARE1	chr8:142276955:C:A	rs72687362	C
ENSG00000198911.11	SREBF2	chr22:41938949:C:G	rs1047997	C
ENSG00000161896.11	IP6K3	chr6:33773939:A:G	rs4711350	A
ENSG00000105894.11	PTN	chr7:137398534:A:G	rs12530512	A
ENSG00000196850.5	PPTC7	chr12:110537085:T:C	rs10849900	T
ENSG00000205981.6	DNAJC19	chr3:181110020:A:G	rs35212830	A
ENSG00000147535.16	PLPP5	chr8:38294926:A:G	rs12386976	A
ENSG00000028116.16	VRK2	chr2:57944085:A:C	rs1518393	A
ENSG00000149929.15	HIRIP3	chr16:29998953:A:G	rs4788207	A
ENSG00000104765.15	BNIP3L	chr8:26411675:C:T	rs1042992	C
ENSG00000120733.13	KDM3B	chr5:138439892:C:G	rs982085	C
ENSG00000132563.15	REEP2	chr5:138439892:C:G	rs982085	C
ENSG00000253553.5	ENSG00000253553.5	chr8:88310775:A:C	rs3844198	A
ENSG00000164088.17	PPM1M	chr3:52253452:G:T	rs11717383	G
ENSG00000100403.11	ZC3H7B	chr22:41357599:G:A	rs11090045	G
ENSG00000228944.1	ENSG00000228944.1	chr7:24695385:G:C	rs112509803	G
ENSG00000105708.8	ZNF14	chr19:19629920:C:G	rs880090	C

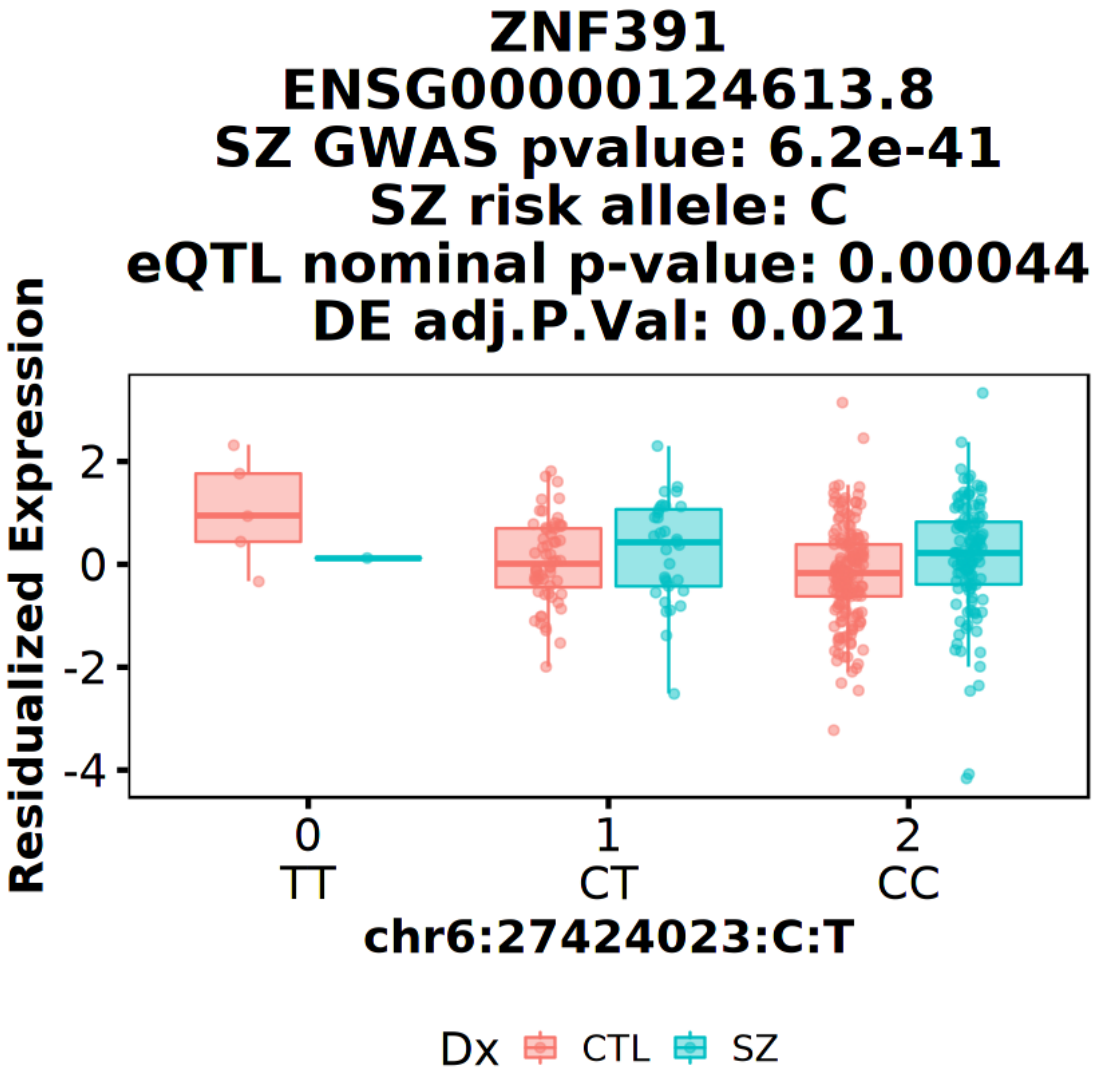
1.2.2 Plot with PGC2 risk allele

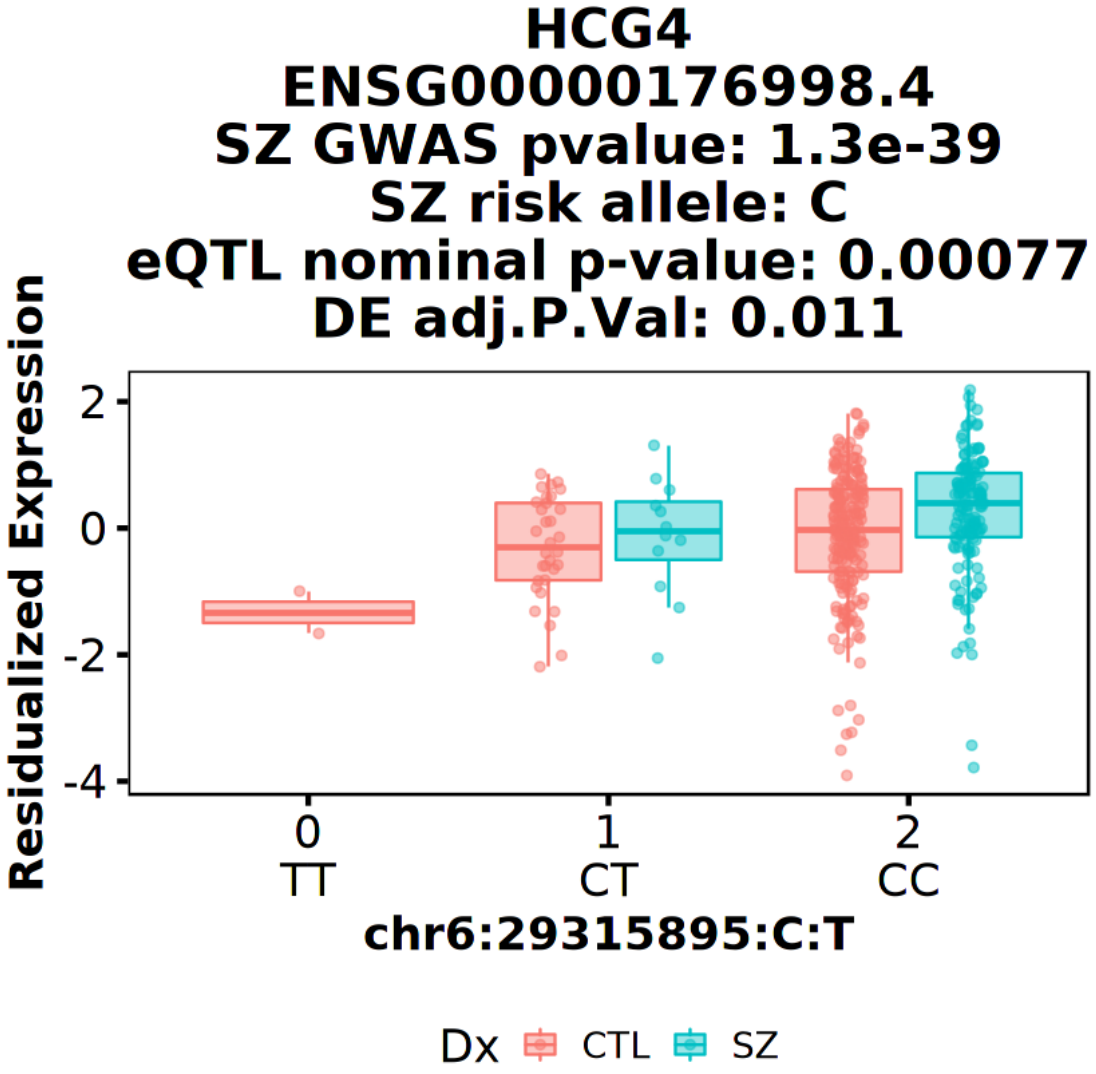
[13]:
```
for(num in seq_along(df$gene_id)){
    variant_id = df$variant_id[num]
    gene_id = df$gene_id[num]
    gene_name = df$Symbol[num]
    pgc2_a1_same_as_our_counted = df$pgc2_a1_same_as_our_counted[num]
    OR = df$OR[num]; A1 = df$A1[num]; A2 = df$A2[num]
    fn = paste0(feature, "/eqtl_gwas_", gsub("\\.", "_", gene_name))
    de_annot = paste('DE adj.P.Val:', signif(df$adj.P.Val[num], 2))
```

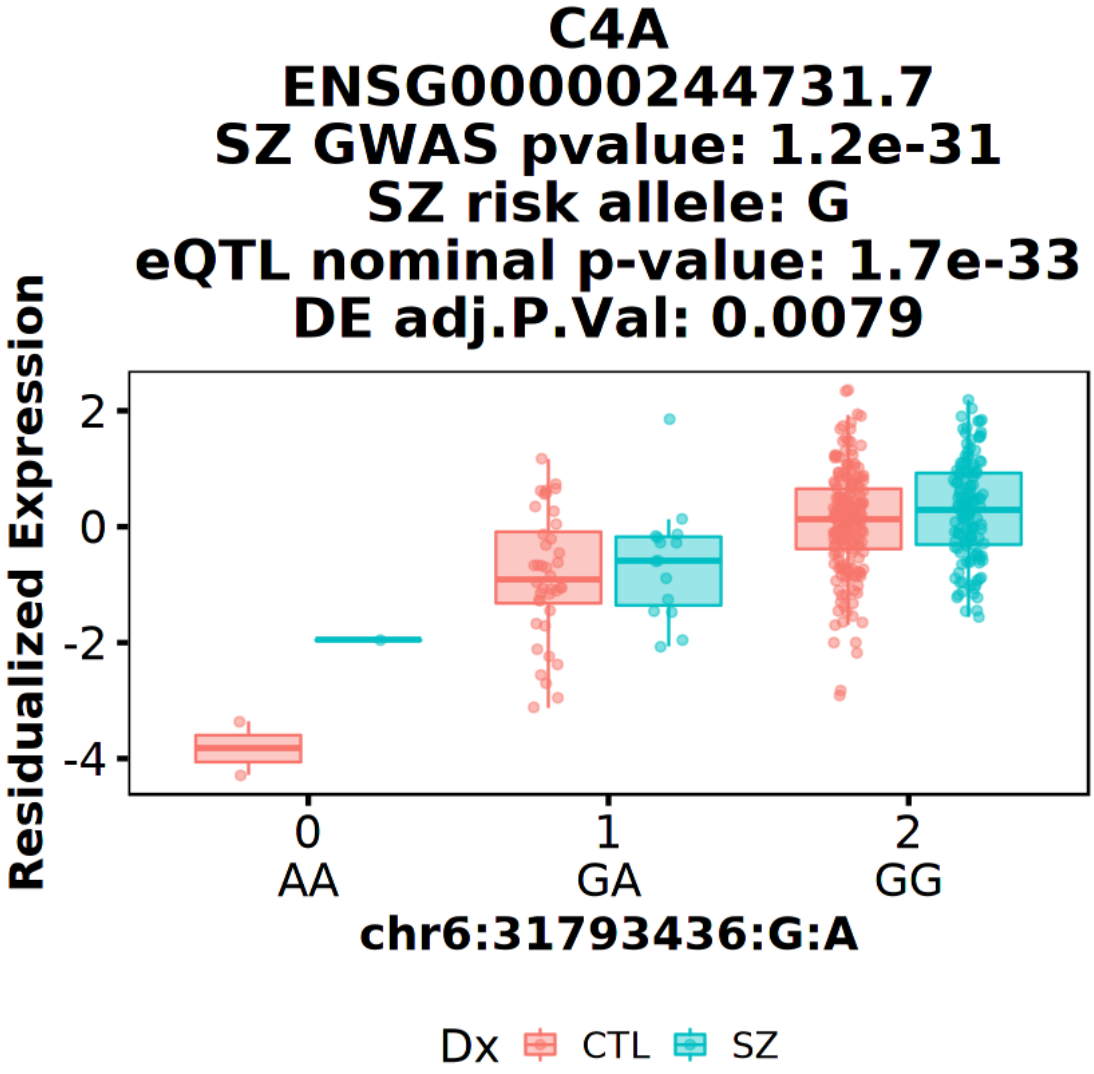
```
    eqtl_annot = paste("eQTL nominal p-value:", signif(df$pval_nominal[num], 2))
    gwas_annot = paste("SZ GWAS pvalue:", signif(df$P[num], 2))
    risk_annot = paste("SZ risk allele:", get_risk_allele(OR, A1, A2))
    title = paste(get_gene_symbol(gene_id), gene_id, gwas_annot,
                  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)
}
```

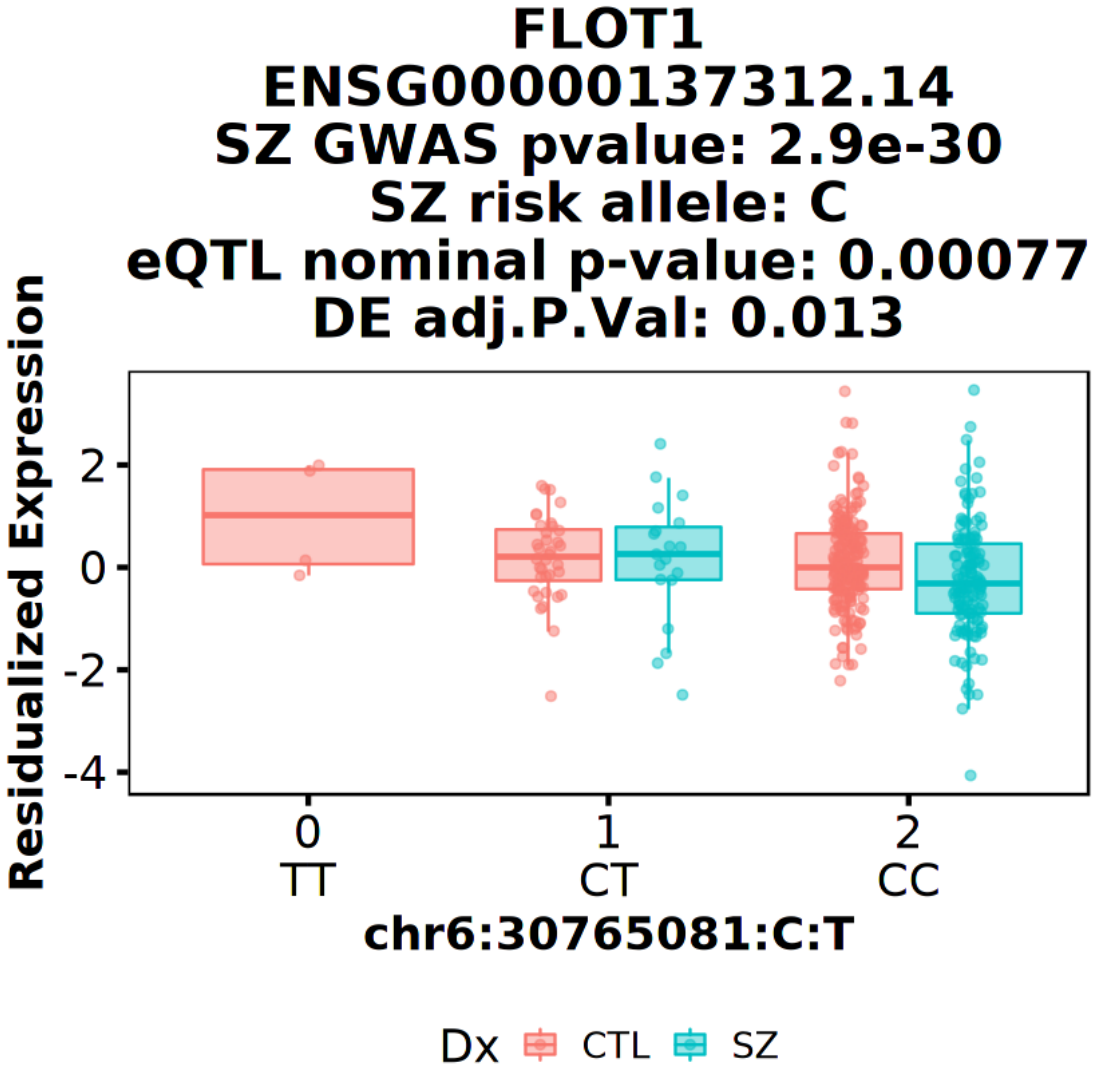
```
Warning message in data.table::fread(expr_file):
"Detected 393 column names but the data has 394 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file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

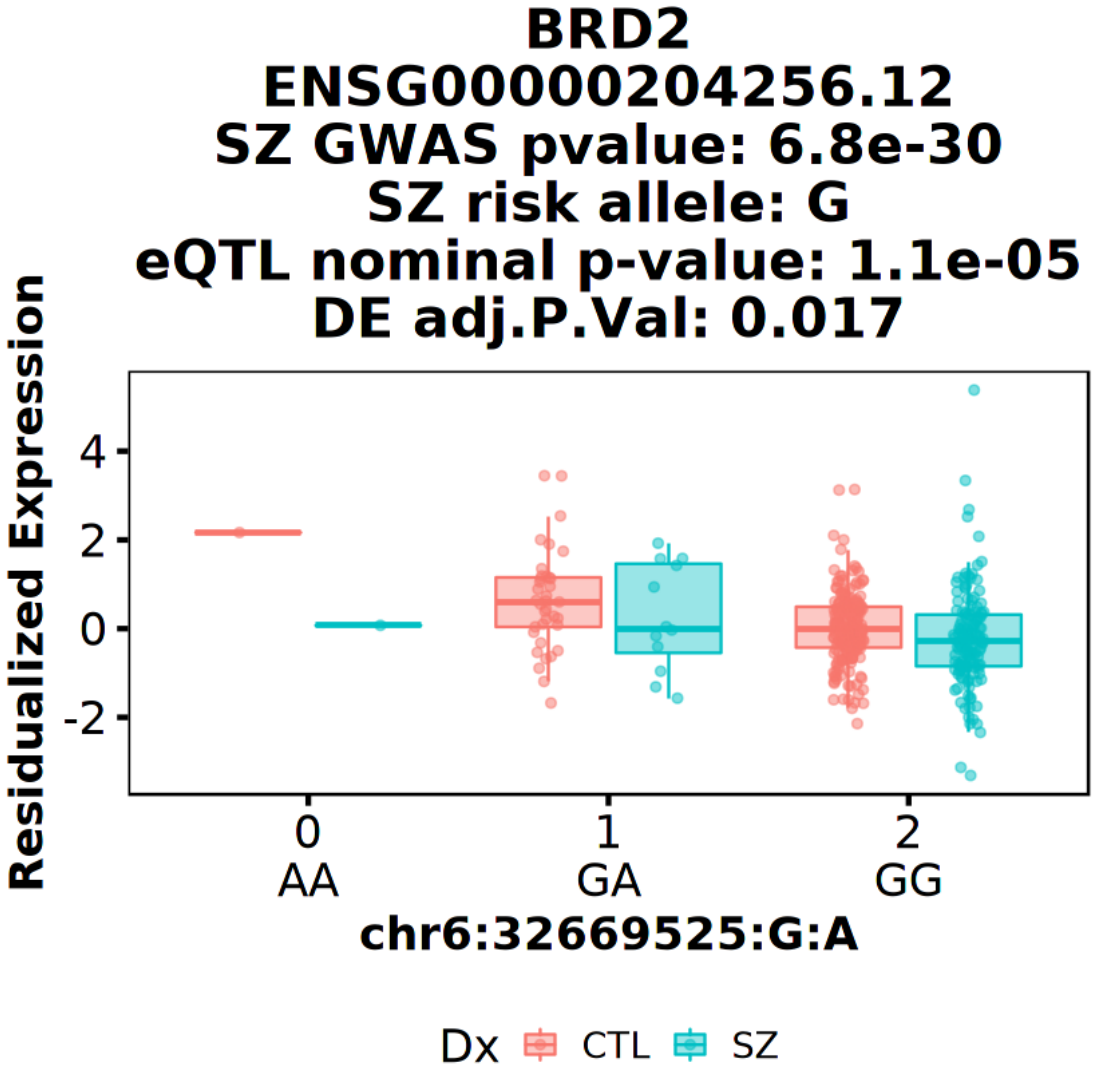


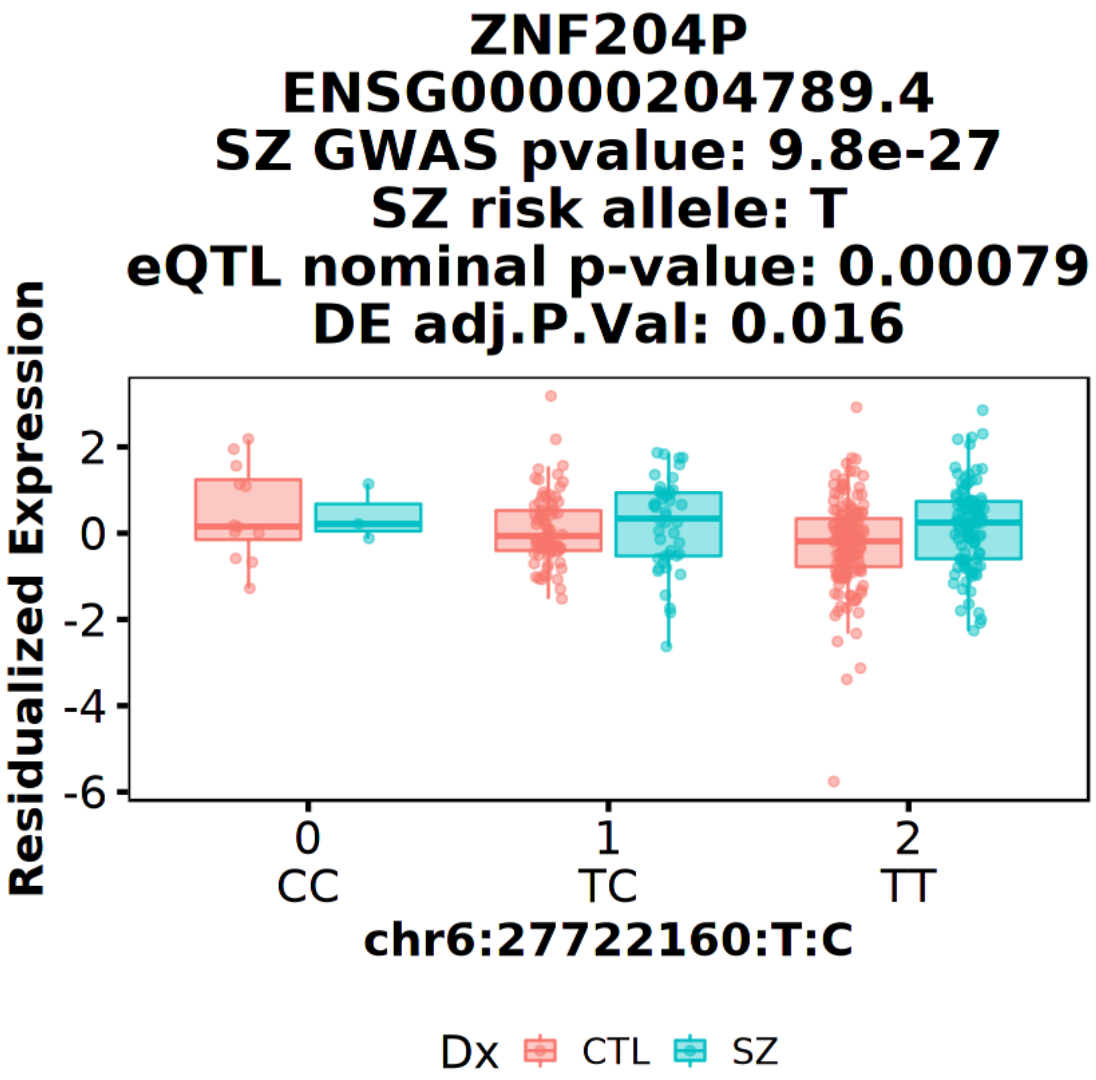

ZNF391
ENSG00000124613.8
SZ GWAS pvalue: 6.2e-41
SZ risk allele: C
eQTL nominal p-value: 0.00044
DE adj.P.Val: 0.021
Residualized Expression
2
0
-2
-4
0
TT
1
CT
2
CC
chr6:27424023:C:T
Dx
CTL
SZ

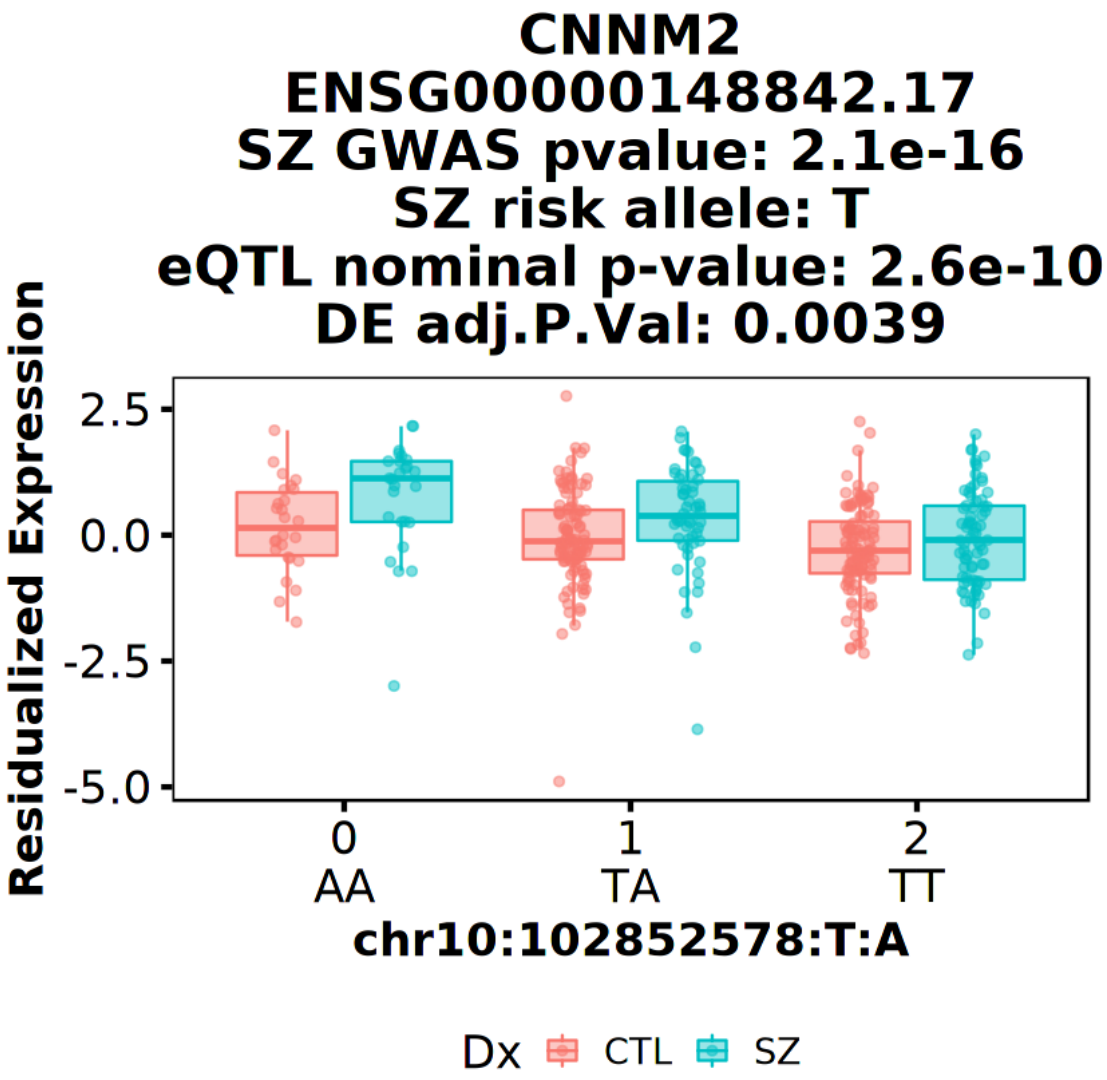

HCG4
ENSG00000176998.4
SZ GWAS pvalue: 1.3e-39
SZ risk allele: C
eQTL nominal p-value: 0.00077
DE adj.P.Val: 0.011
Residualized Expression
2
0
-2
-4
0
TT
1
CT
2
CC
chr6:29315895:C:T
Dx
CTL
SZ

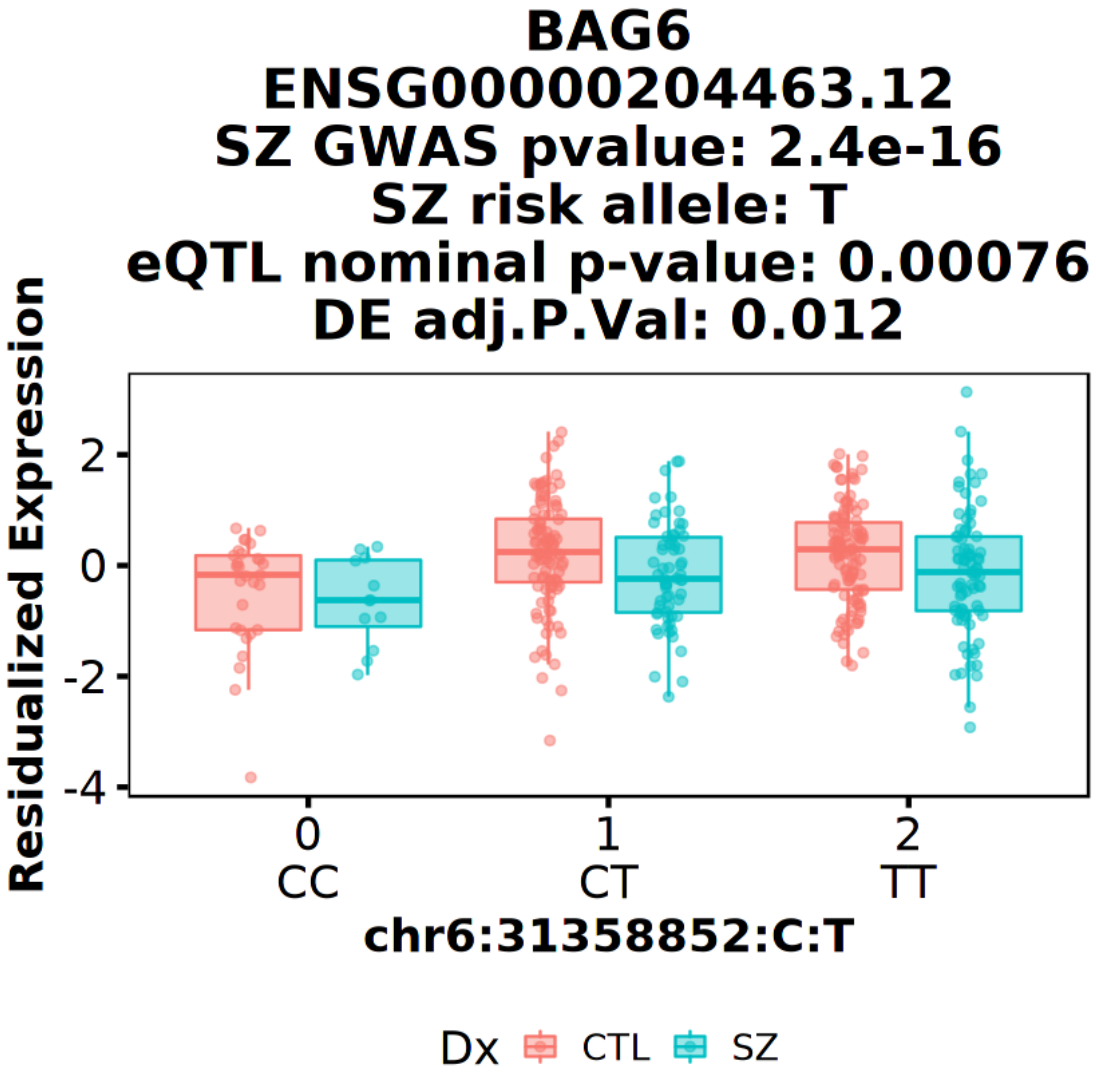

C4A
ENSG00000244731.7
SZ GWAS pvalue: 1.2e-31
SZ risk allele: G
eQTL nominal p-value: 1.7e-33
DE adj.P.Val: 0.0079
Residualized Expression
2
0
-2
-4
0
AA
1
GA
2
GG
chr6:31793436:G:A
Dx
CTL
SZ

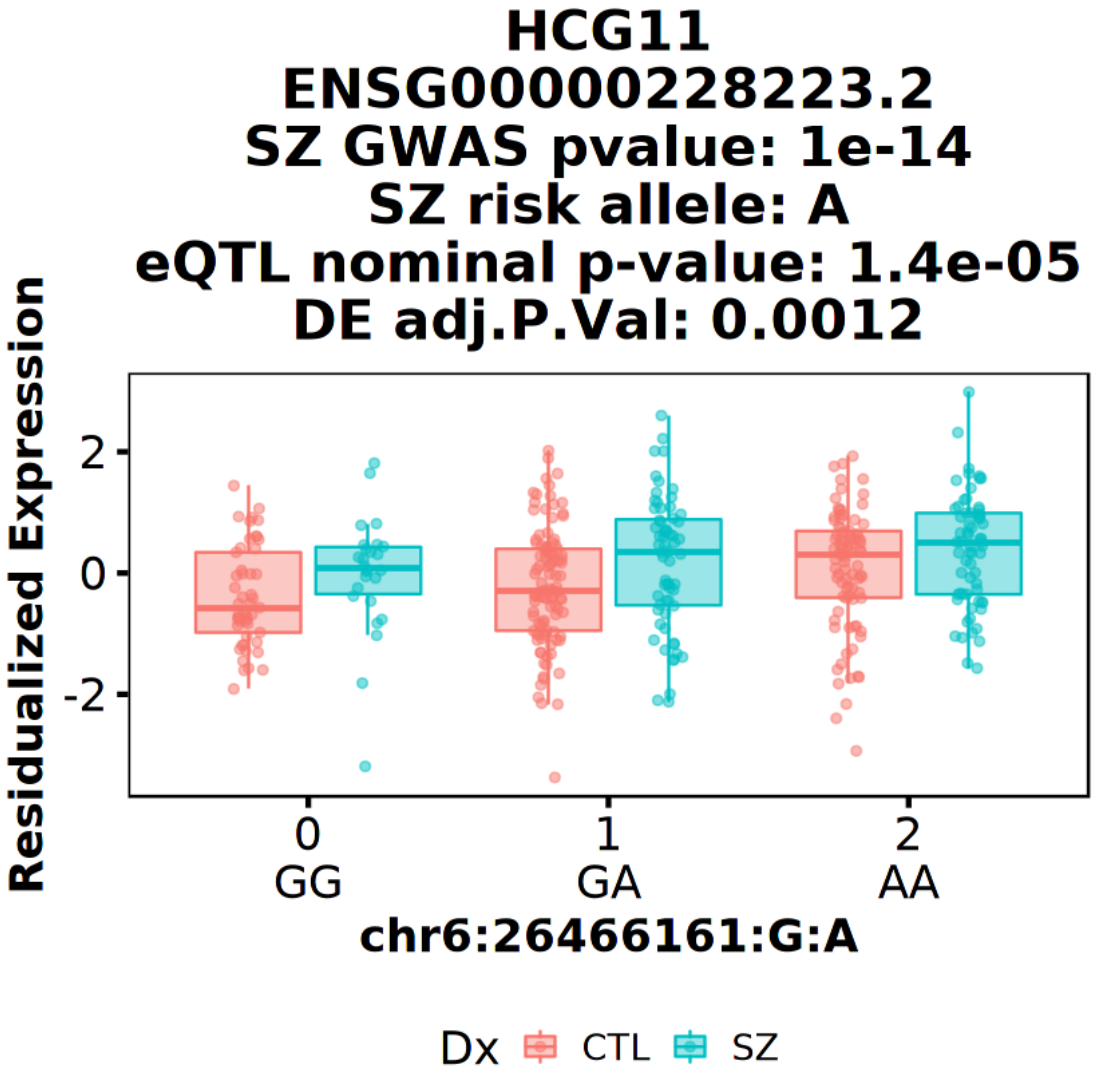

FLOT1
ENSG00000137312.14
SZ GWAS pvalue: 2.9e-30
SZ risk allele: C
eQTL nominal p-value: 0.00077
DE adj.P.Val: 0.013
Residualized Expression
2
0
-2
-4
0
TT
1
CT
2
CC
chr6:30765081:C:T
Dx
CTL
SZ

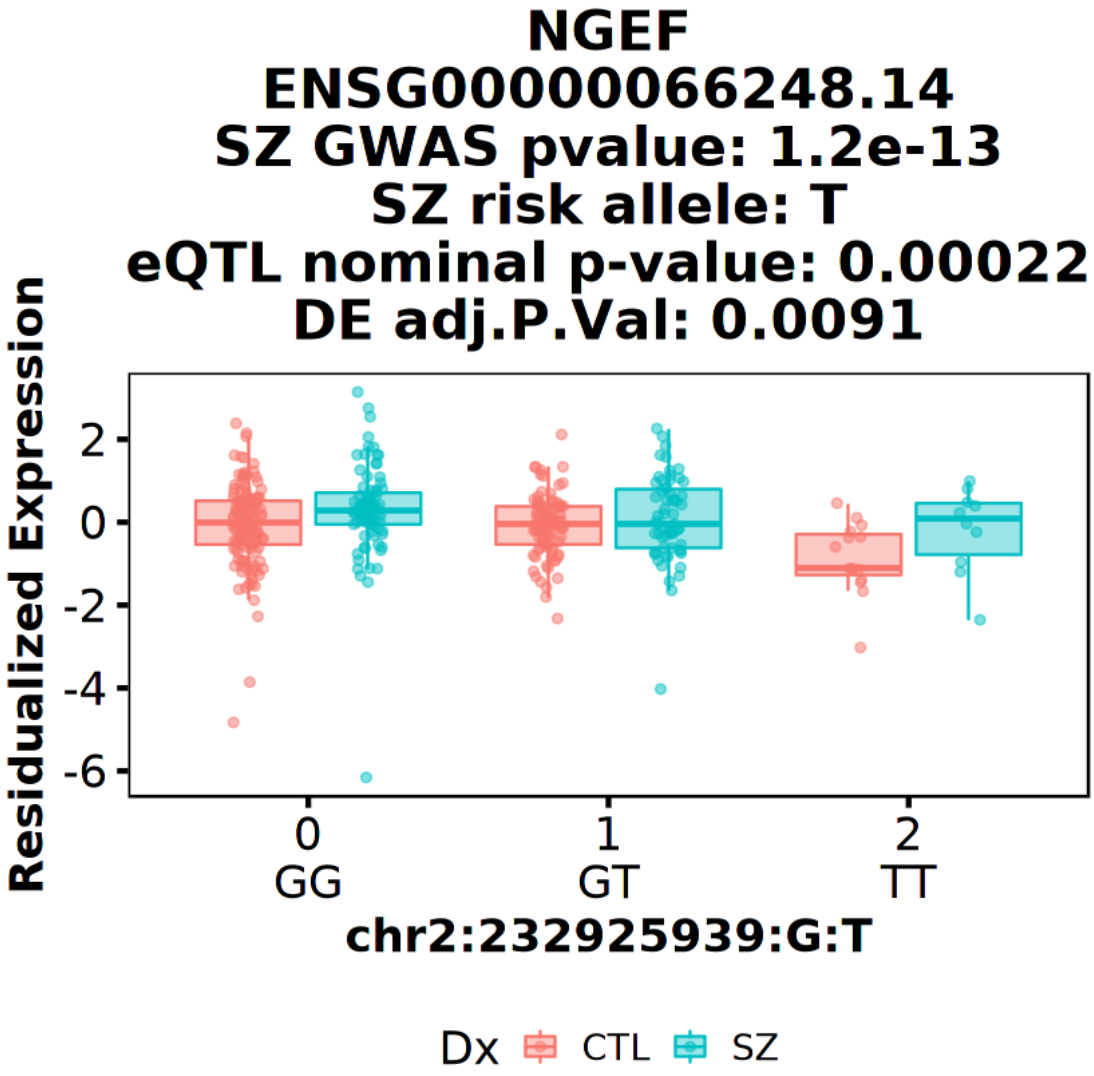

BRD2
ENSG00000204256.12
SZ GWAS pvalue: 6.8e-30
SZ risk allele: G
eQTL nominal p-value: 1.1e-05
DE adj.P.Val: 0.017
Residualized Expression
4
2
0
-2
0
AA
1
GA
2
GG
chr6:32669525:G:A
Dx
CTL
SZ

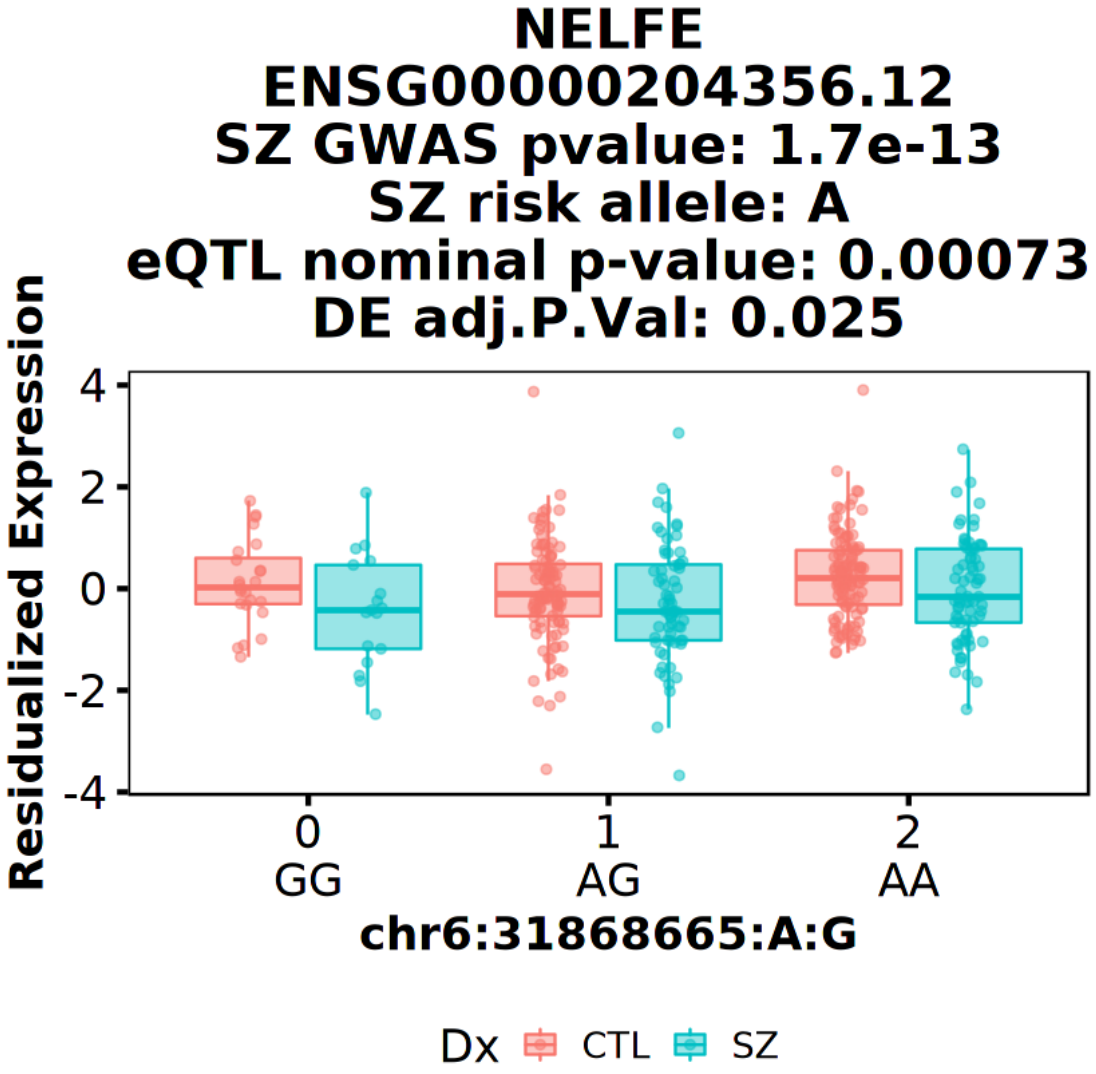

ZNF204P
ENSG00000204789.4
SZ GWAS pvalue: 9.8e-27
SZ risk allele: T
eQTL nominal p-value: 0.00079
DE adj.P.Val: 0.016
Residualized Expression
2
0
-2
-4
-6
0
CC
1
TC
2
TT
chr6:27722160:T:C
Dx
CTL
SZ

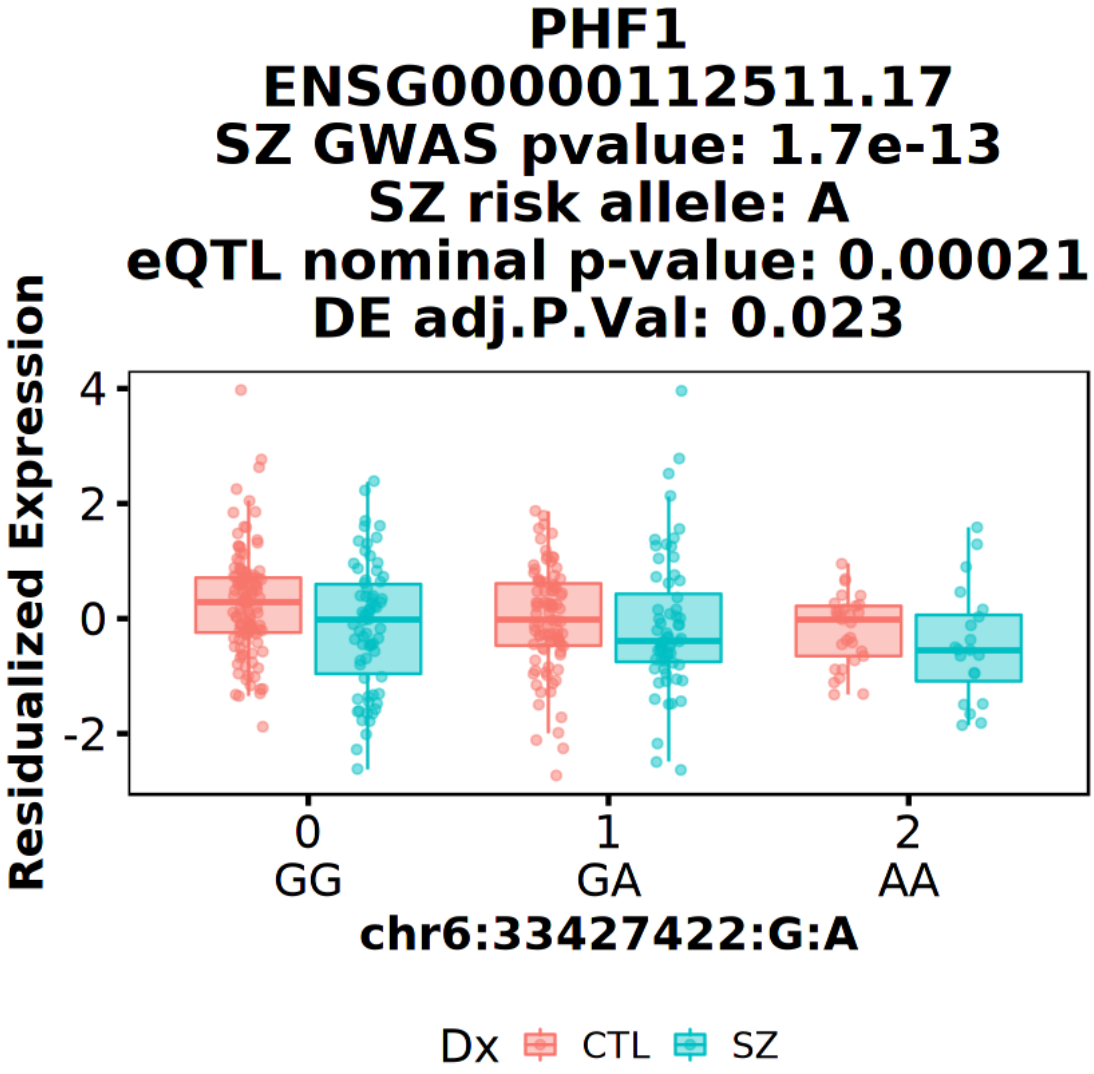

CNNM2
ENSG00000148842.17
SZ GWAS pvalue: 2.1e-16
SZ risk allele: T
eQTL nominal p-value: 2.6e-10
DE adj.P.Val: 0.0039
Residualized Expression
2.5
0.0
-2.5
-5.0
0
AA
1
TA
2
TT
chr10:102852578:T:A
Dx
CTL
SZ

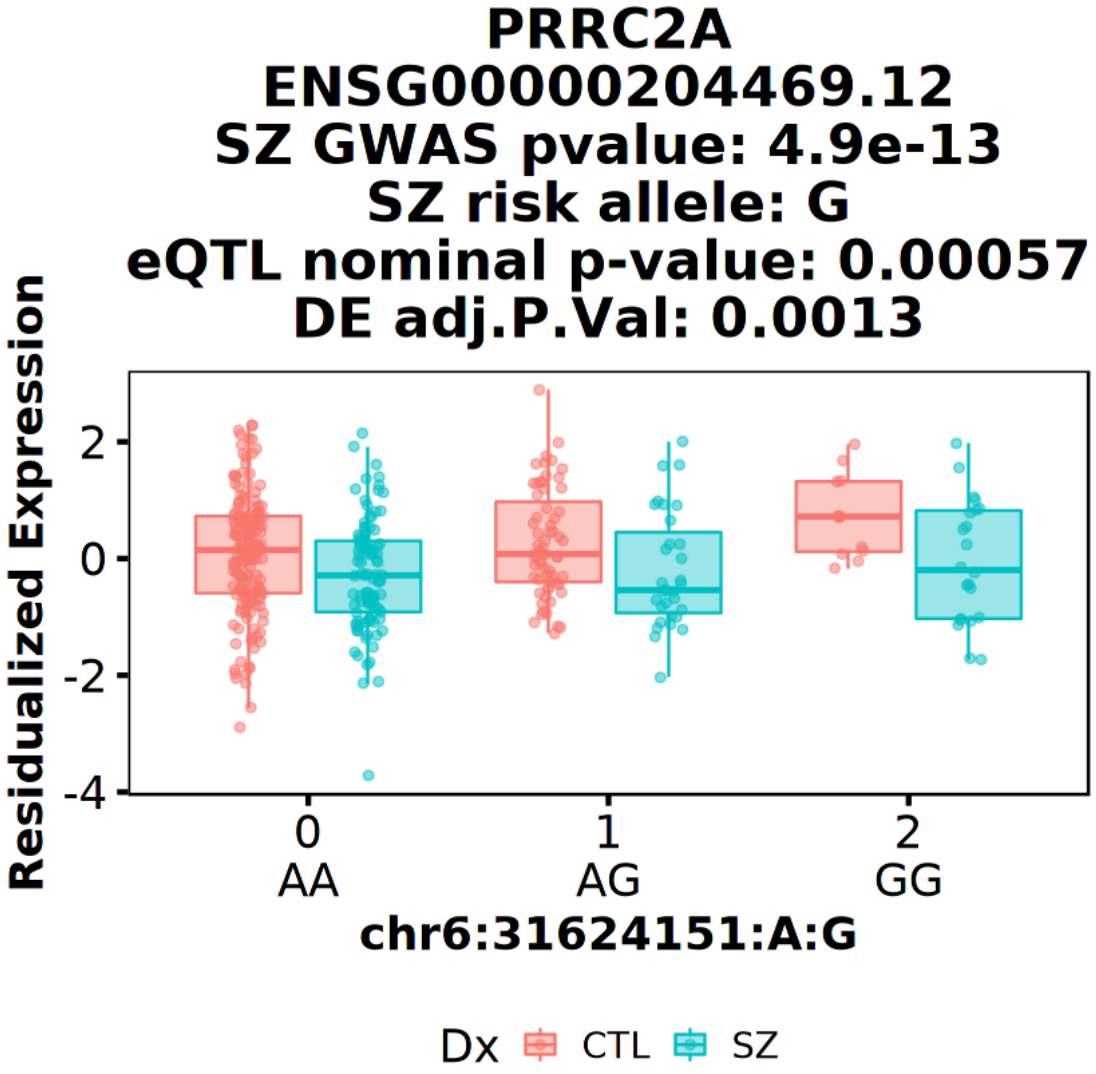

BAG6
ENSG00000204463.12
SZ GWAS pvalue: 2.4e-16
SZ risk allele: T
eQTL nominal p-value: 0.00076
DE adj.P.Val: 0.012
Residualized Expression
2
0
-2
-4
0
CC
1
CT
2
TT
chr6:31358852:C:T
Dx
CTL
SZ

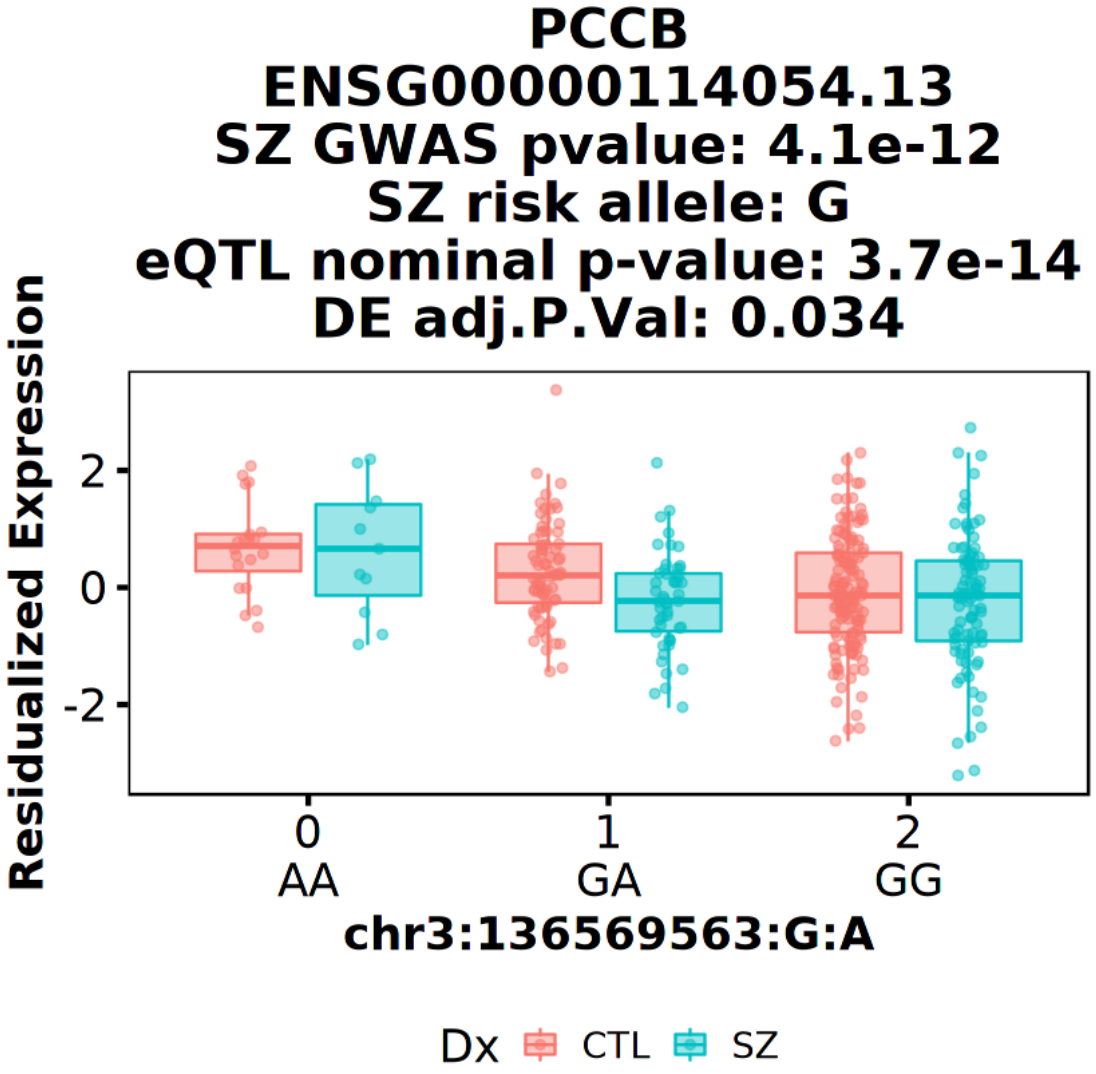

HCG11
ENSG00000228223.2
SZ GWAS pvalue: 1e-14
SZ risk allele: A
eQTL nominal p-value: 1.4e-05
DE adj.P.Val: 0.0012
Residualized Expression
2
0
-2
0
GG
1
GA
2
AA
chr6:26466161:G:A
Dx
CTL
SZ

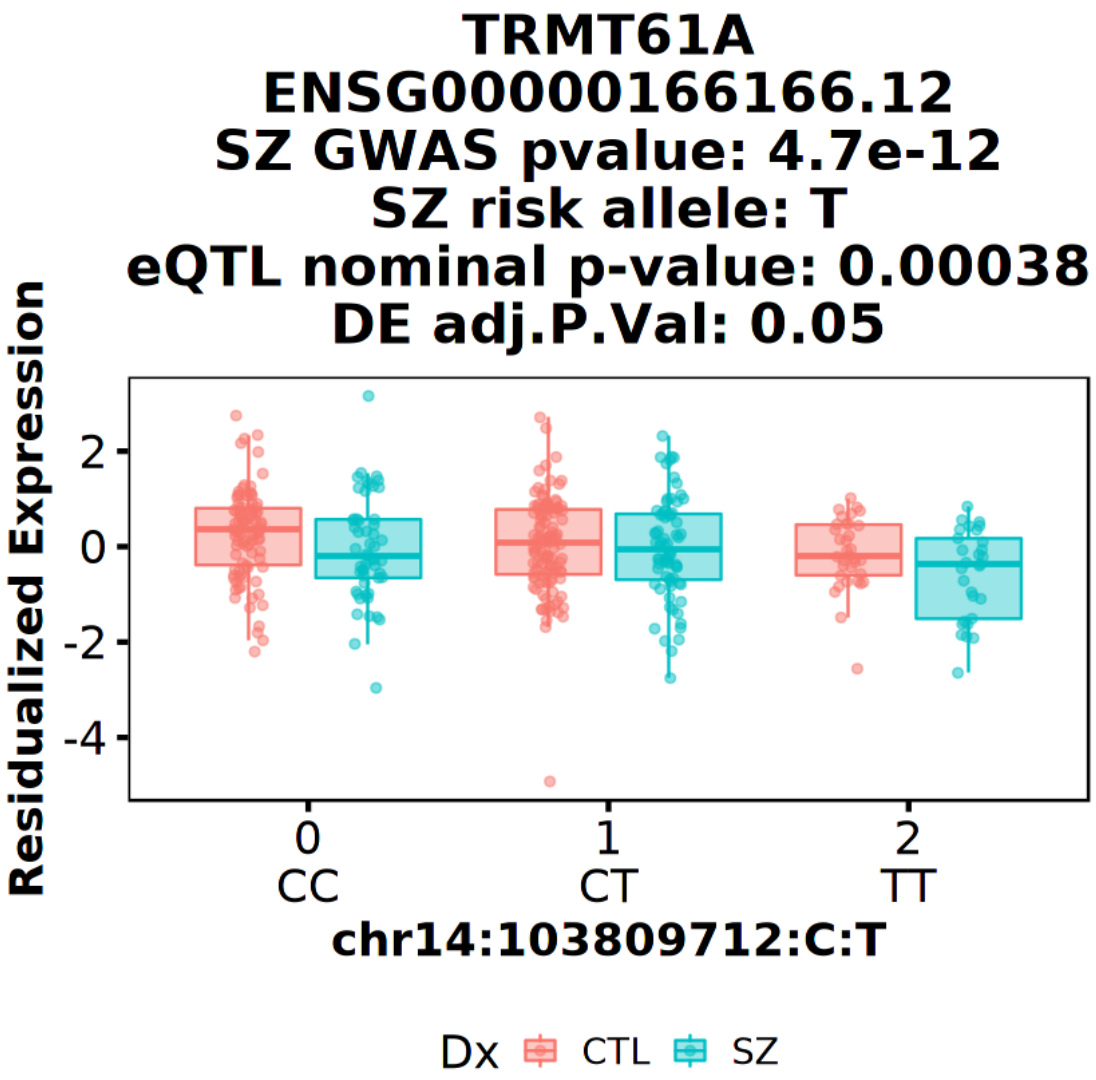

NGEF
ENSG00000066248.14
SZ GWAS pvalue: 1.2e-13
SZ risk allele: T
eQTL nominal p-value: 0.00022
DE adj.P.Val: 0.0091
Residualized Expression
2
0
-2
-4
-6
0
GG
1
GT
2
TT
chr2:232925939:G:T
Dx
CTL
SZ

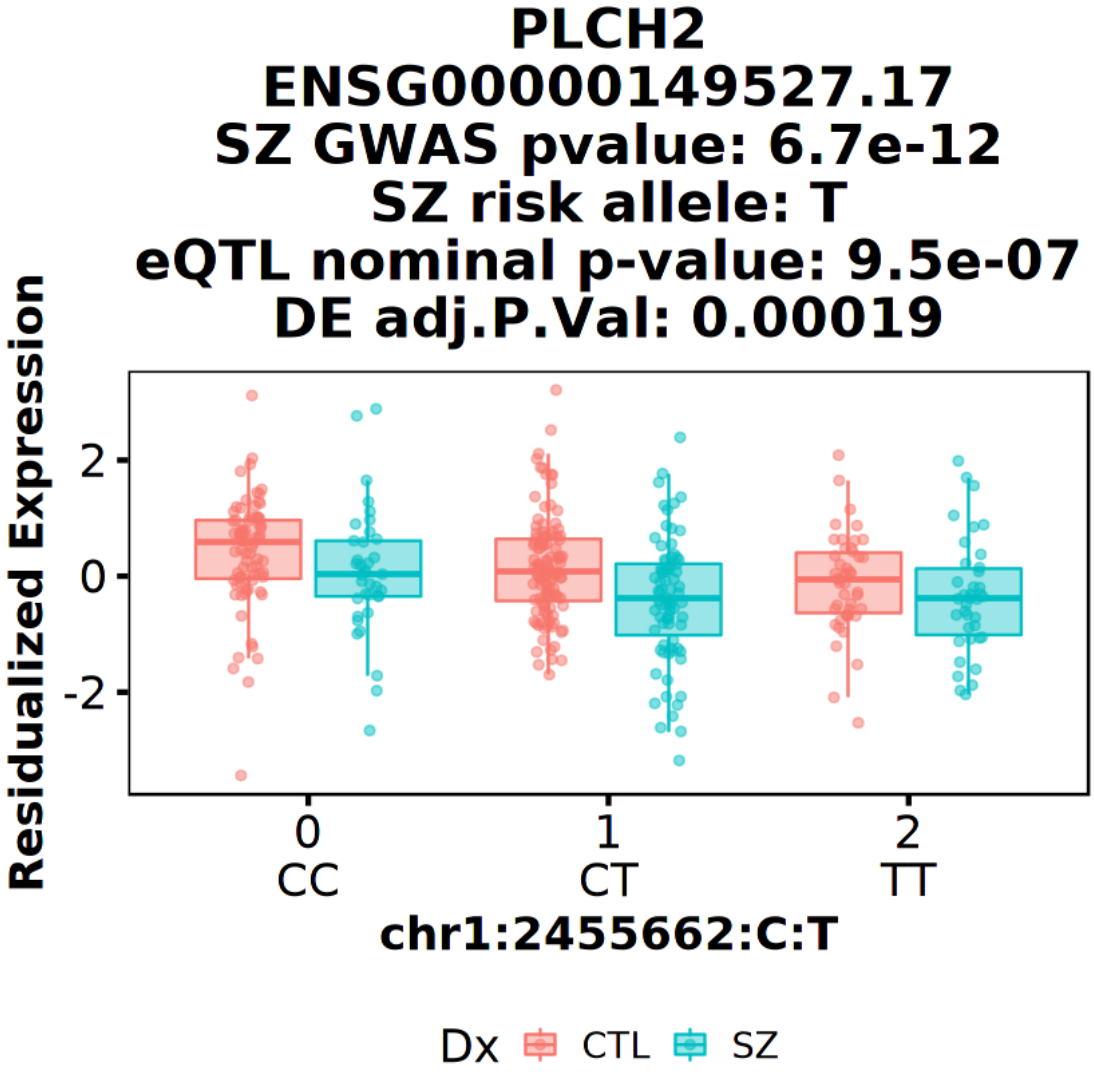

NELFE
ENSG00000204356.12
SZ GWAS pvalue: 1.7e-13
SZ risk allele: A
eQTL nominal p-value: 0.00073
DE adj.P.Val: 0.025
Residualized Expression
4
2
0
-2
-4
0
GG
1
AG
2
AA
chr6:31868665:A:G
Dx
CTL
SZ

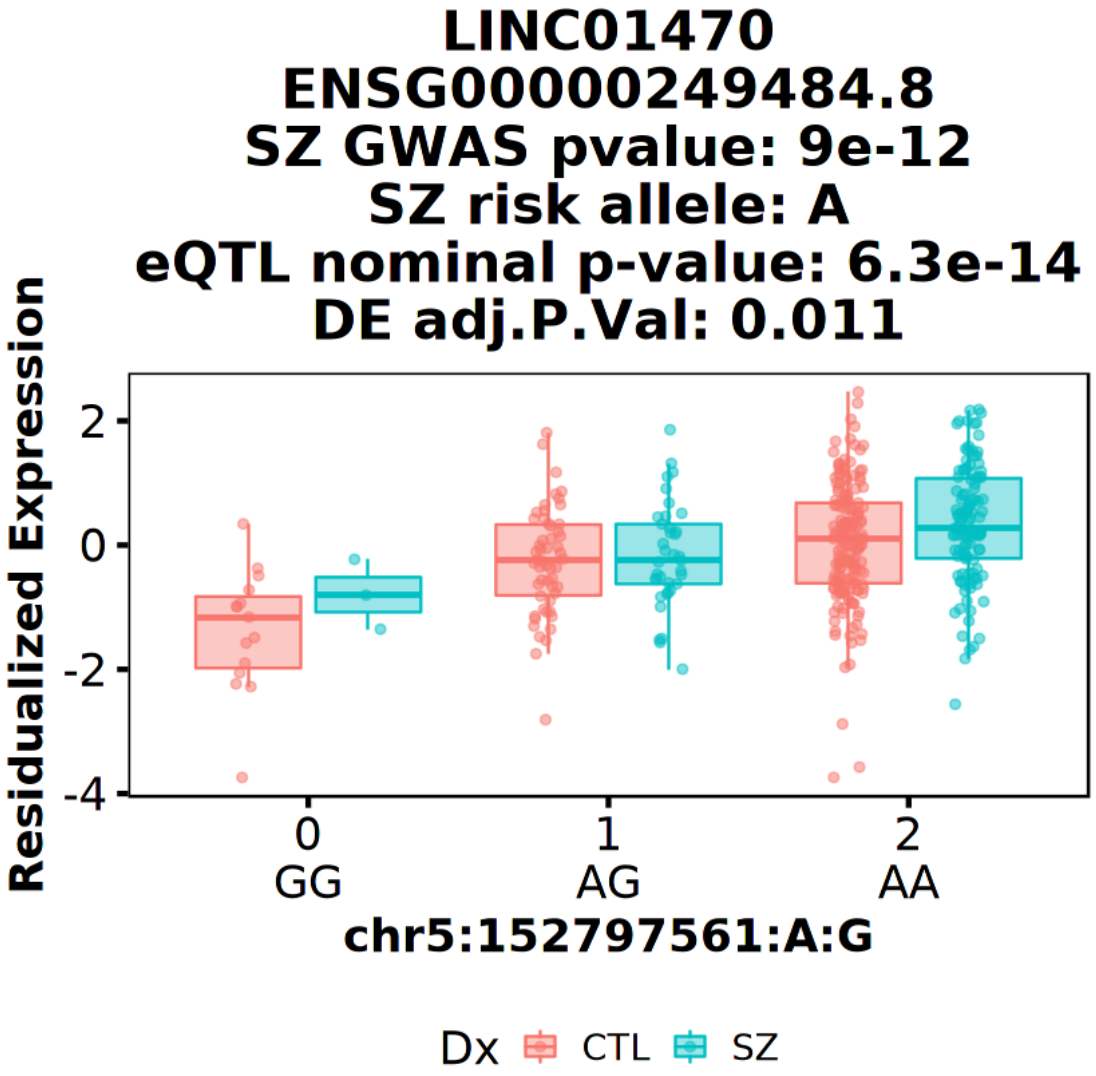

PHF1
ENSG00000112511.17
SZ GWAS pvalue: 1.7e-13
SZ risk allele: A
eQTL nominal p-value: 0.00021
DE adj.P.Val: 0.023
Residualized Expression
4
2
0
-2
0
GG
1
GA
2
AA
chr6:33427422:G:A
Dx
CTL
SZ

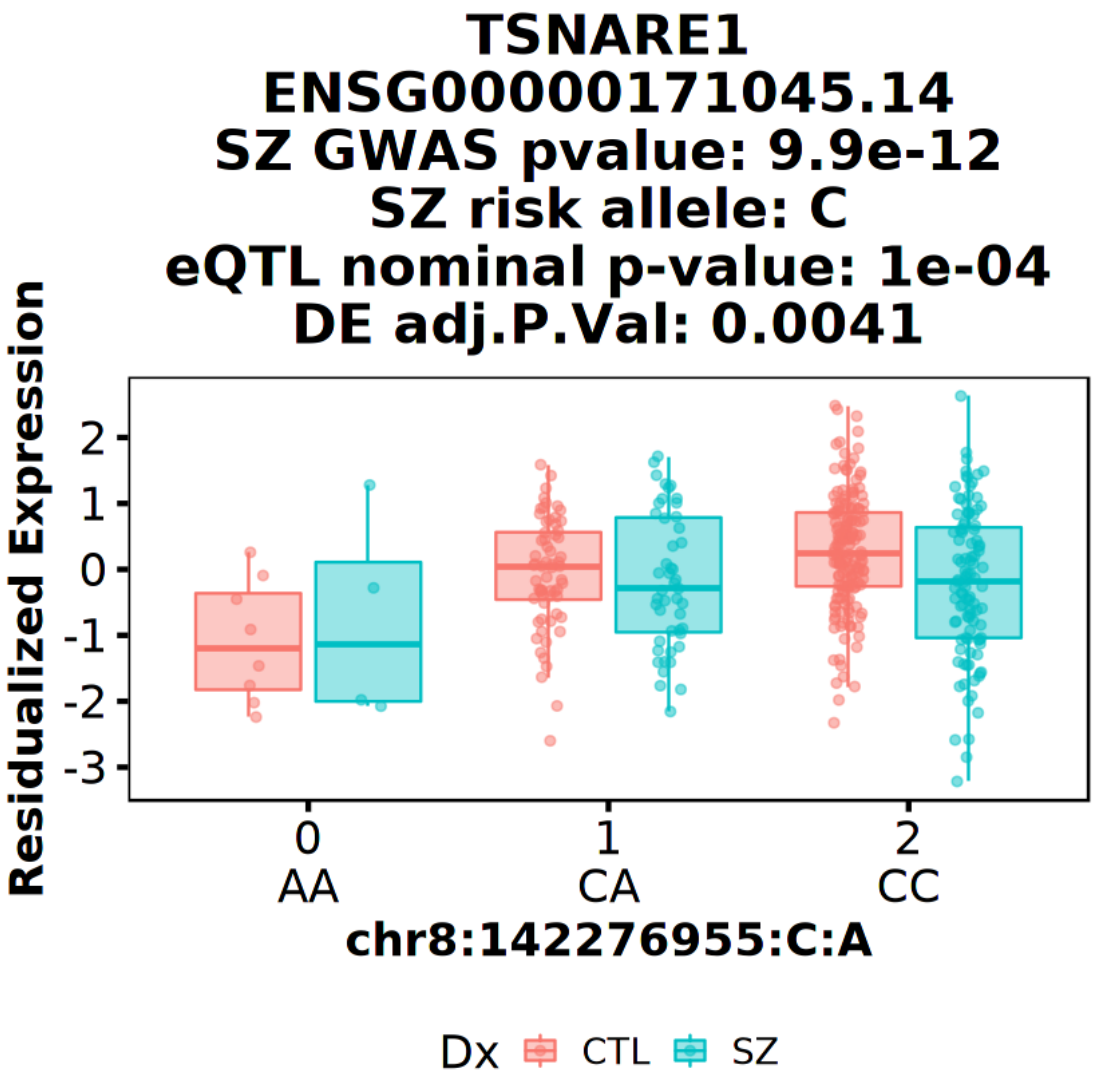

PRRC2A
ENSG00000204469.12
SZ GWAS pvalue: 4.9e-13
SZ risk allele: G
eQTL nominal p-value: 0.00057
DE adj.P.Val: 0.0013
Residualized Expression
2
0
-2
-4
0
AA
1
AG
2
GG
chr6:31624151:A:G
Dx
CTL
SZ

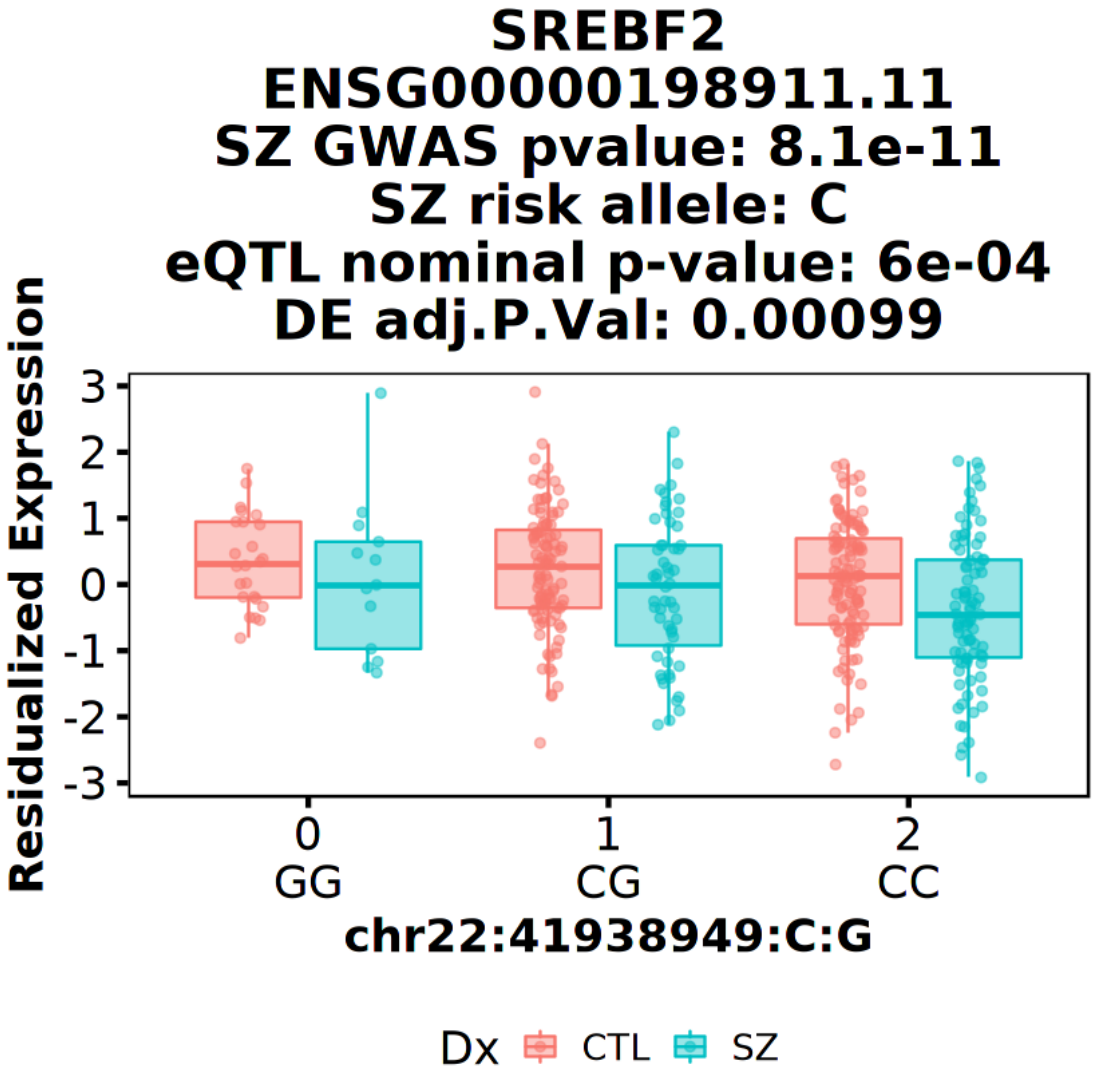

PCCB
ENSG00000114054.13
SZ GWAS pvalue: 4.1e-12
SZ risk allele: G
eQTL nominal p-value: 3.7e-14
DE adj.P.Val: 0.034
Residualized Expression
2
0
-2
0
AA
1
GA
2
GG
chr3:136569563:G:A
Dx
CTL
SZ

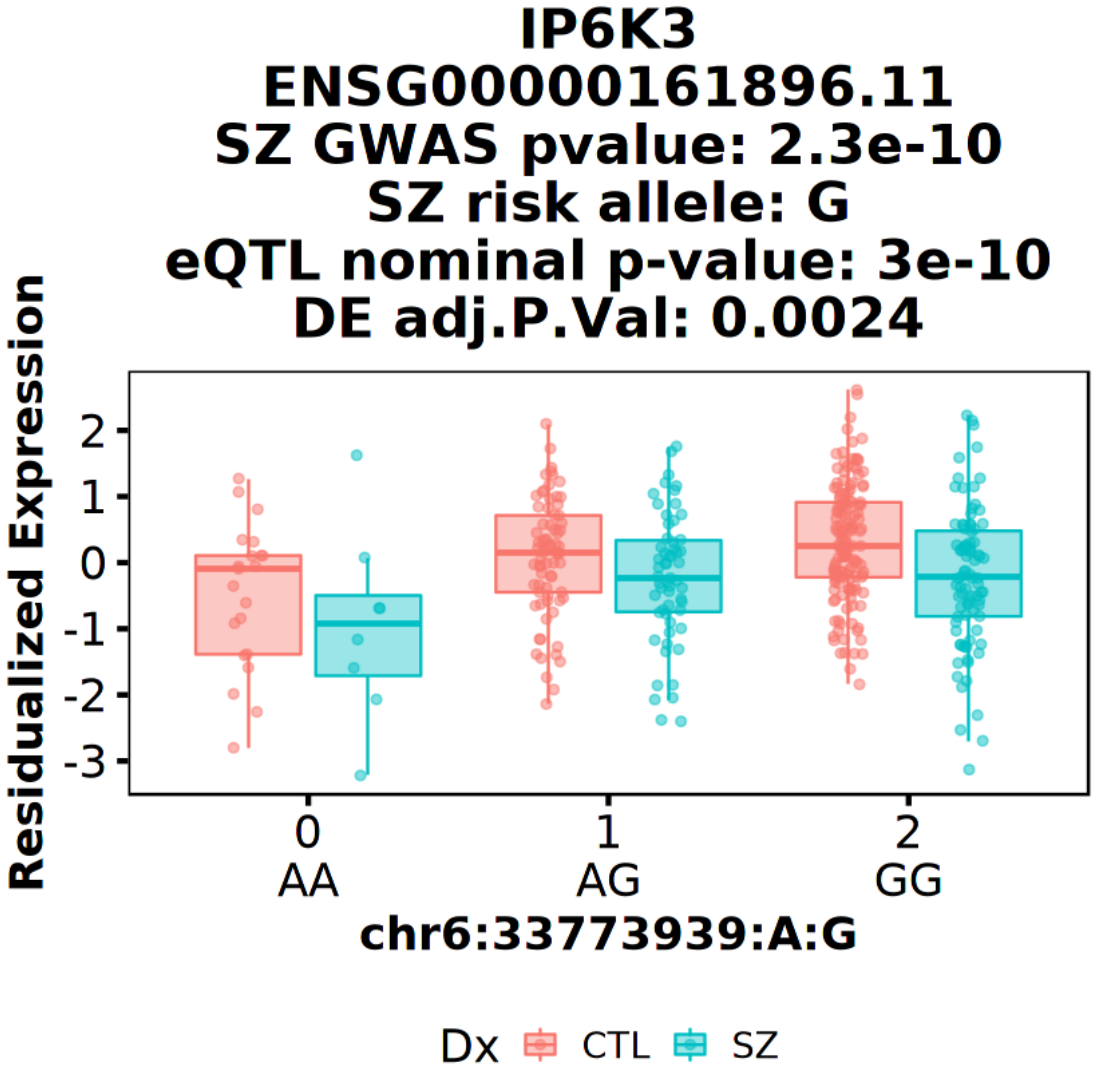

TRMT61A
ENSG00000166166.12
SZ GWAS pvalue: 4.7e-12
SZ risk allele: T
eQTL nominal p-value: 0.00038
DE adj.P.Val: 0.05
Residualized Expression
2
0
-2
-4
0
CC
1
CT
2
TT
chr14:103809712:C:T
Dx
CTL
SZ

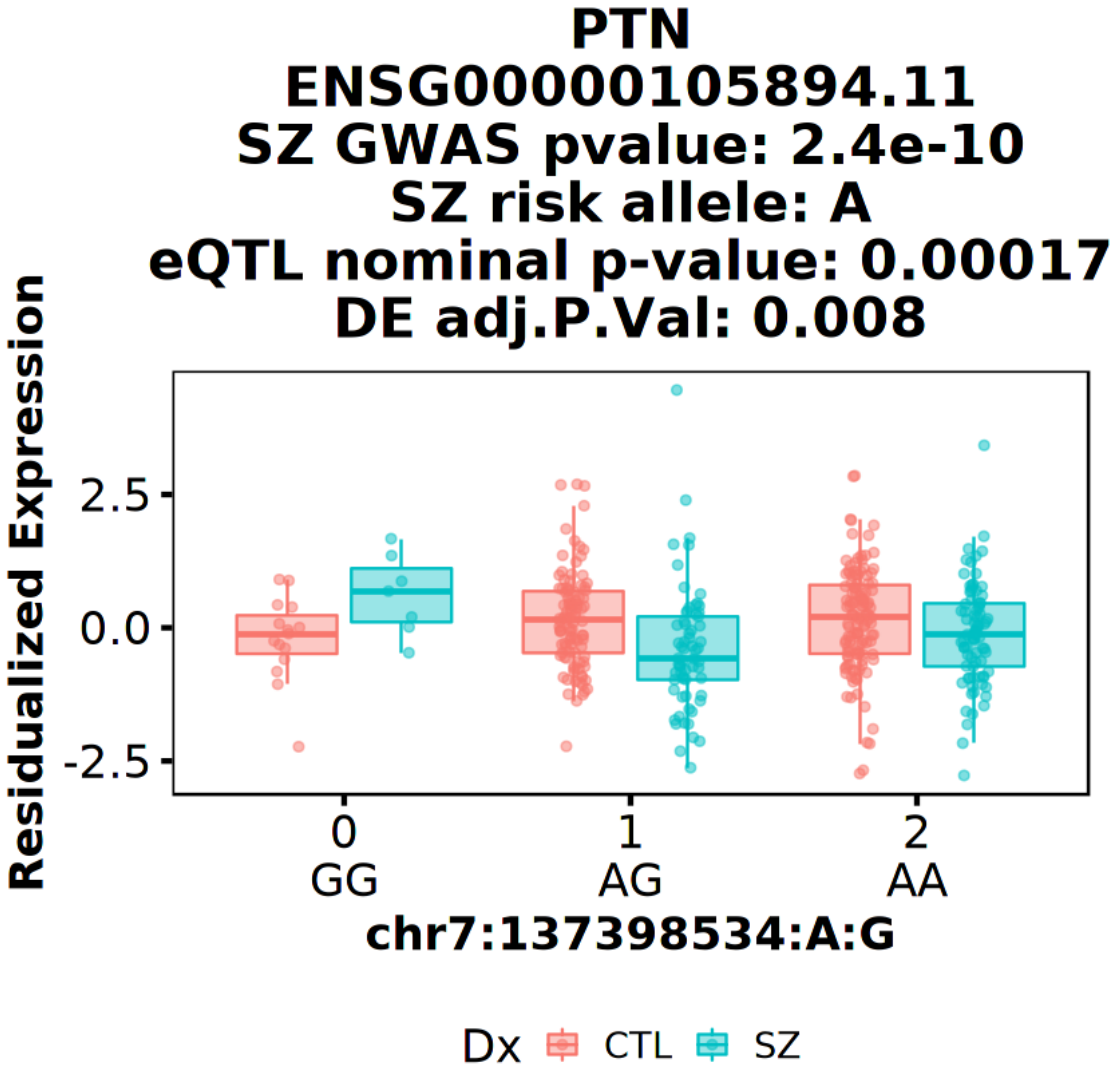

PLCH2
ENSG00000149527.17
SZ GWAS pvalue: 6.7e-12
SZ risk allele: T
eQTL nominal p-value: 9.5e-07
DE adj.P.Val: 0.00019
Residualized Expression
2
0
-2
0
CC
1
CT
2
TT
chr1:2455662:C:T
Dx
CTL
SZ

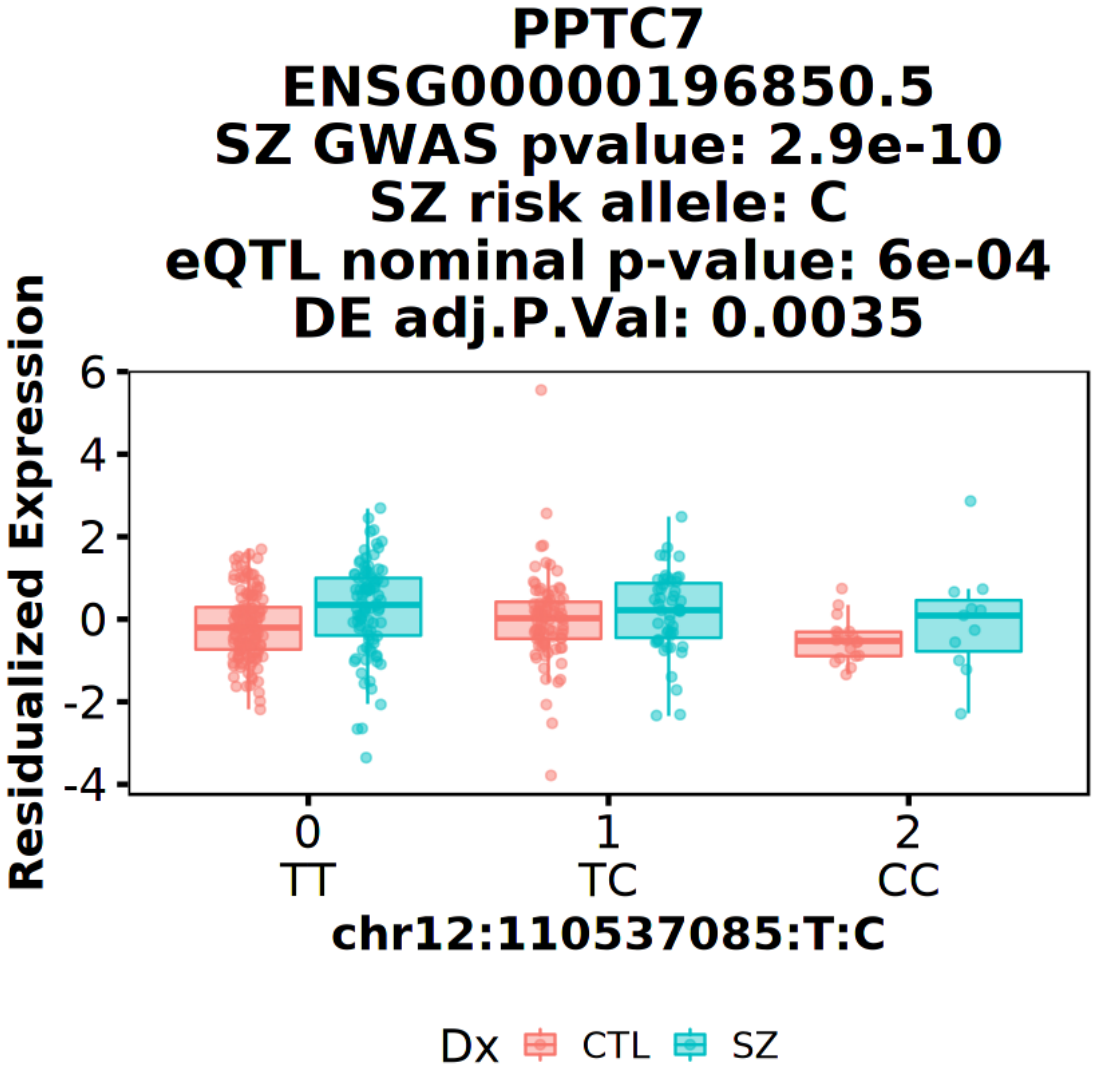

LINC01470
ENSG00000249484.8
SZ GWAS pvalue: 9e-12
SZ risk allele: A
eQTL nominal p-value: 6.3e-14
DE adj.P.Val: 0.011
Residualized Expression
2
0
-2
-4
0
GG
1
AG
2
AA
chr5:152797561:A:G
Dx
CTL
SZ

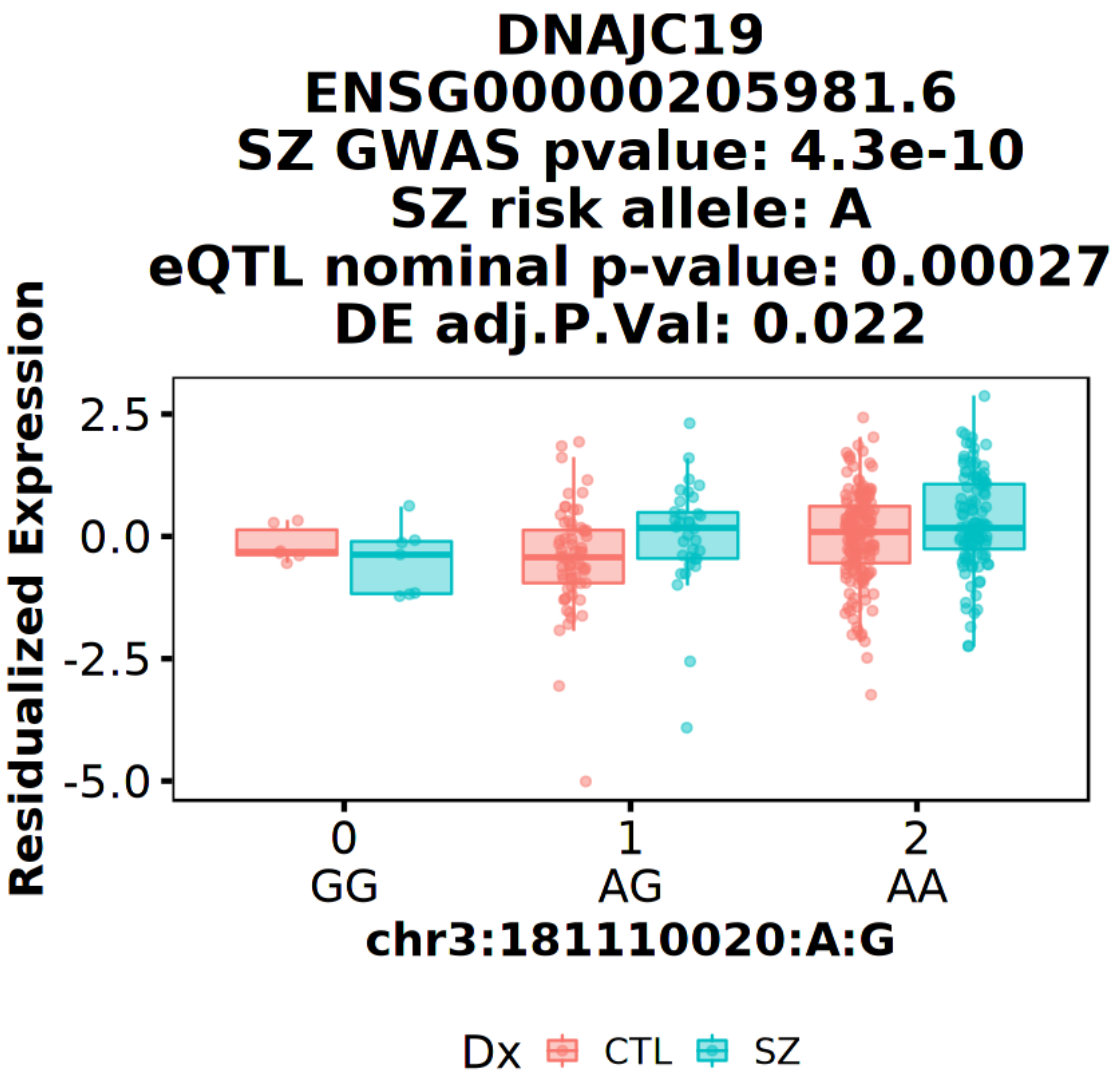

TSNARE1
ENSG00000171045.14
SZ GWAS pvalue: 9.9e-12
SZ risk allele: C
eQTL nominal p-value: 1e-04
DE adj.P.Val: 0.0041
Residualized Expression
2
1
0
-1
-2
-3
0
AA
1
CA
2
CC
chr8:142276955:C:A
Dx
CTL
SZ

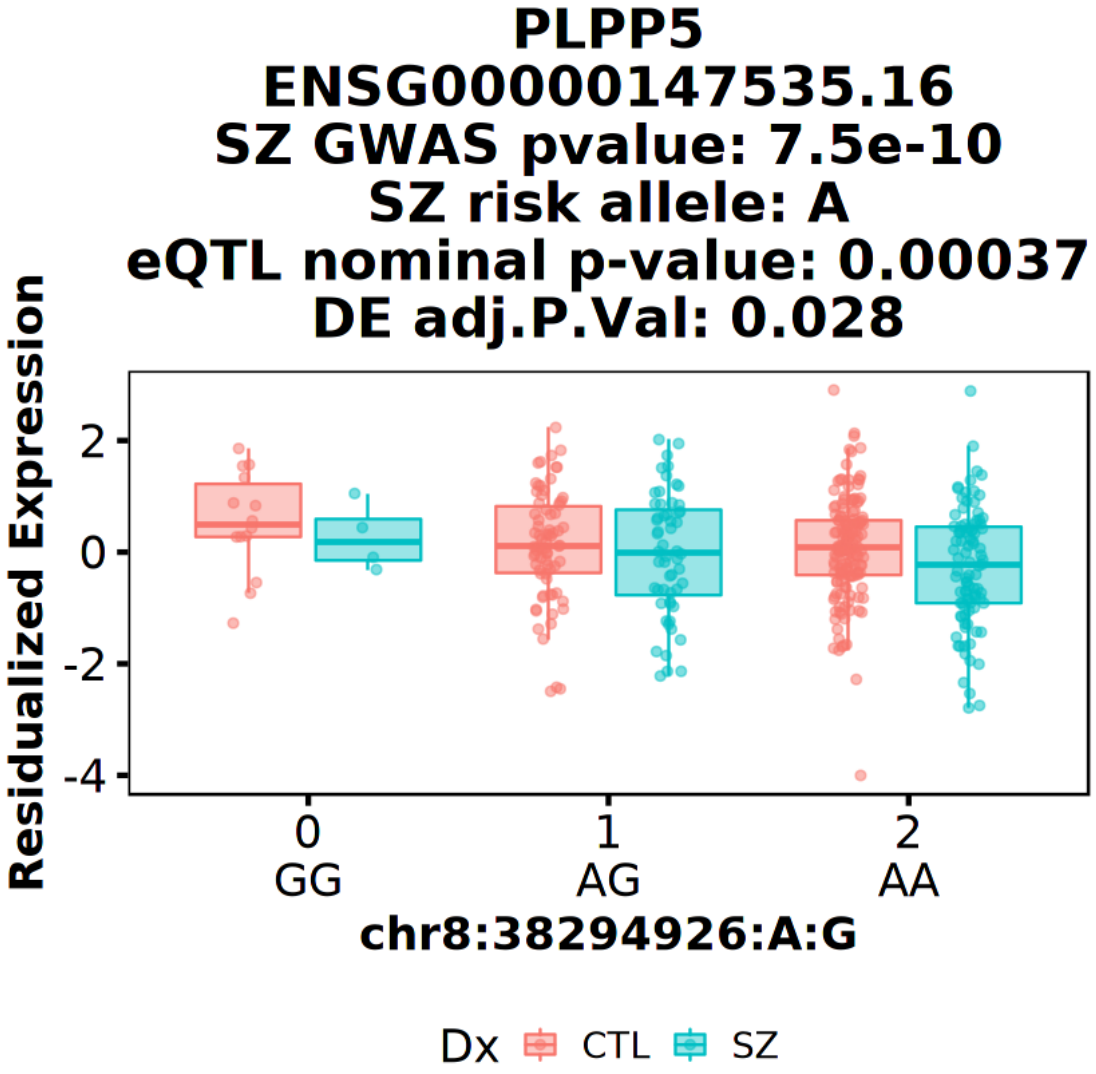

SREBF2
ENSG00000198911.11
SZ GWAS pvalue: 8.1e-11
SZ risk allele: C
eQTL nominal p-value: 6e-04
DE adj.P.Val: 0.00099
Residualized Expression
3
2
1
0
-1
-2
-3
0
GG
1
CG
2
CC
chr22:41938949:C:G
Dx
CTL
SZ

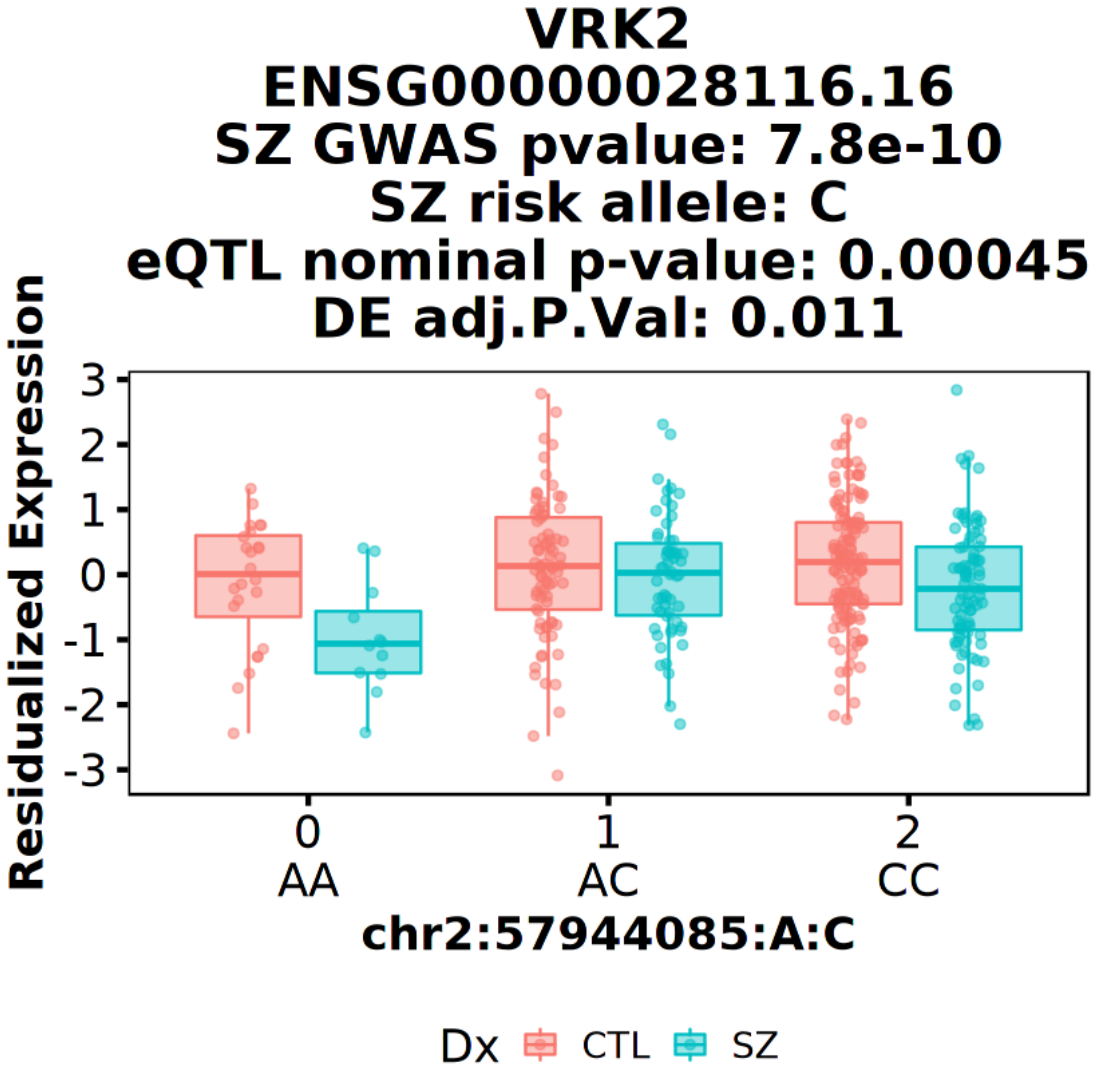

IP6K3
ENSG00000161896.11
SZ GWAS pvalue: 2.3e-10
SZ risk allele: G
eQTL nominal p-value: 3e-10
DE adj.P.Val: 0.0024
Residualized Expression
2
1
0
-1
-2
-3
0
AA
1
AG
2
GG
chr6:33773939:A:G
Dx
CTL
SZ

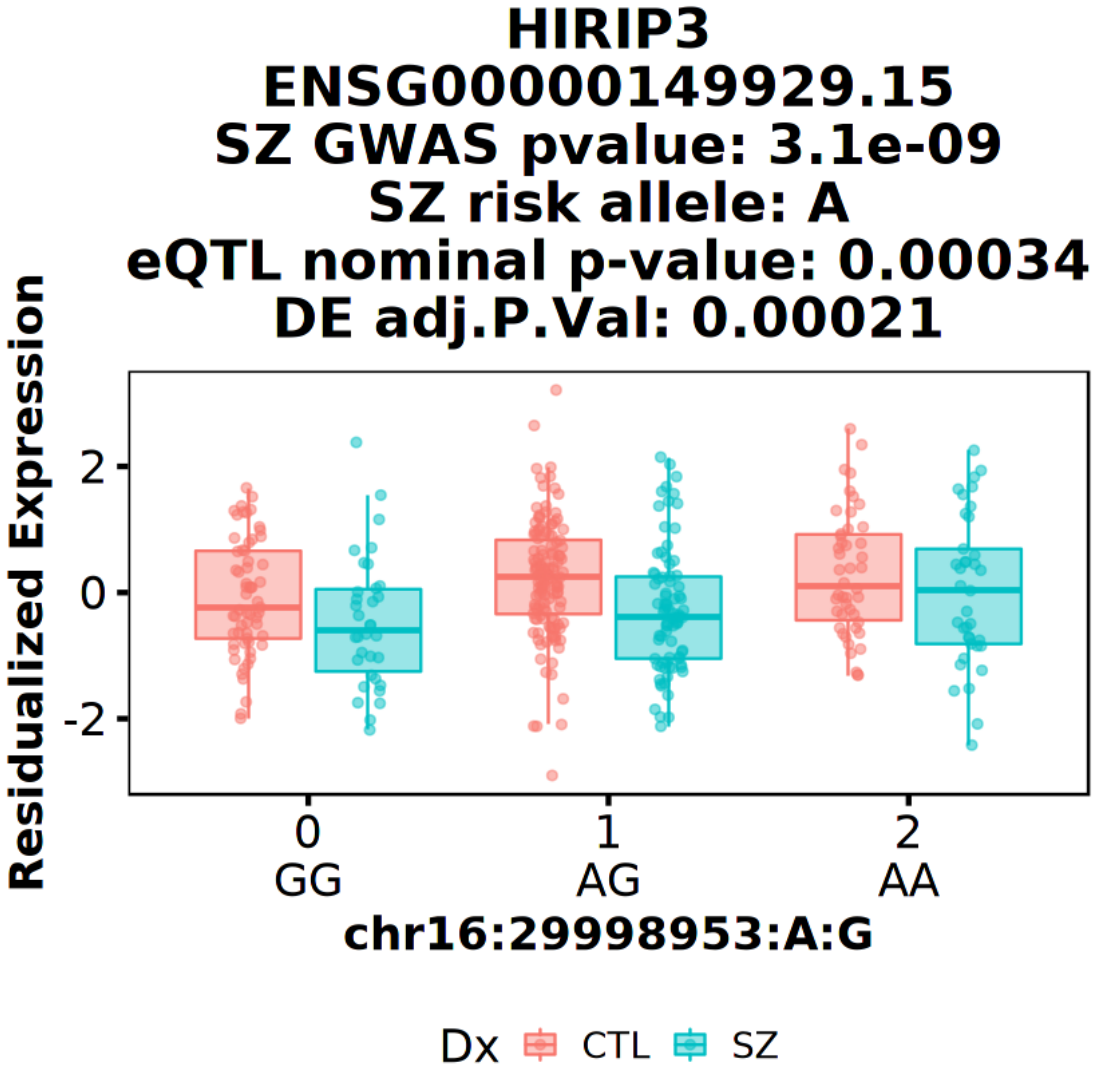

PTN
ENSG00000105894.11
SZ GWAS pvalue: 2.4e-10
SZ risk allele: A
eQTL nominal p-value: 0.00017
DE adj.P.Val: 0.008
Residualized Expression
2.5
0.0
-2.5
0
GG
1
AG
2
AA
chr7:137398534:A:G
Dx
CTL
SZ

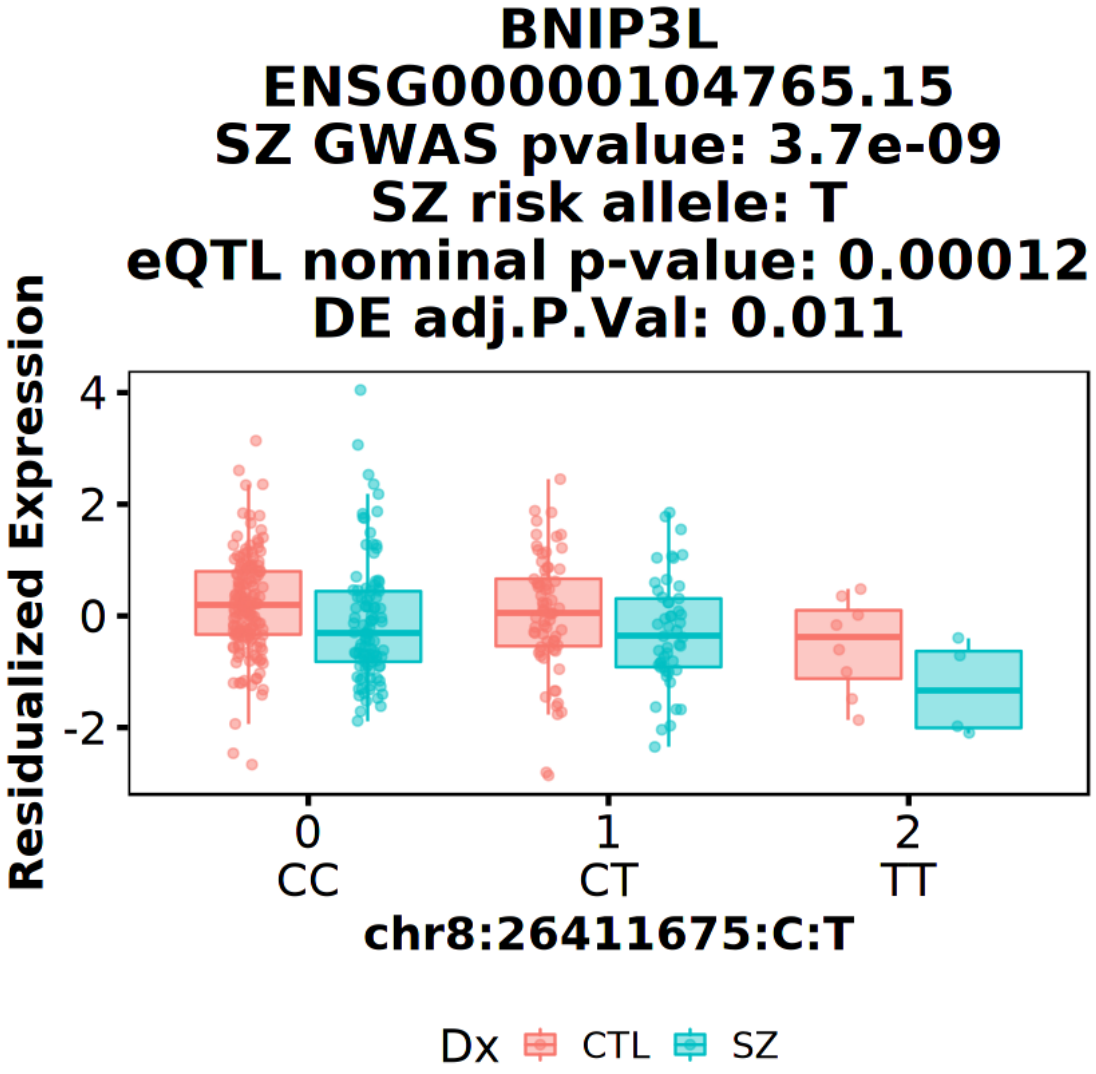

PPTC7
ENSG00000196850.5
SZ GWAS pvalue: 2.9e-10
SZ risk allele: C
eQTL nominal p-value: 6e-04
DE adj.P.Val: 0.0035
Residualized Expression
6
4
2
0
-2
-4
0
TT
1
TC
2
CC
chr12:110537085:T:C
Dx
CTL
SZ

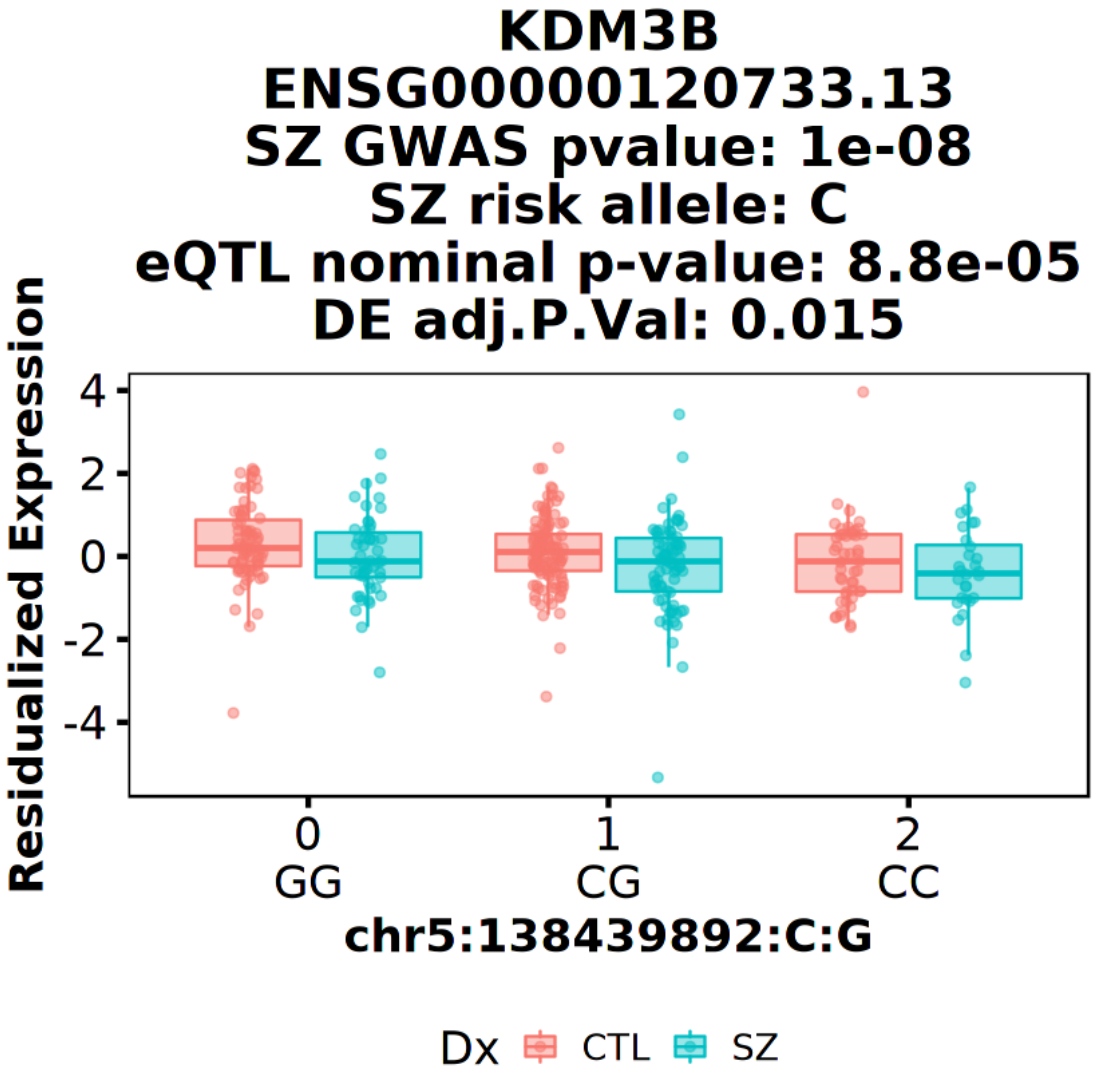

DNAJC19
ENSG00000205981.6
SZ GWAS pvalue: 4.3e-10
SZ risk allele: A
eQTL nominal p-value: 0.00027
DE adj.P.Val: 0.022
Residualized Expression
2.5
0.0
-2.5
-5.0
0
GG
1
AG
2
AA
chr3:181110020:A:G
Dx
CTL
SZ

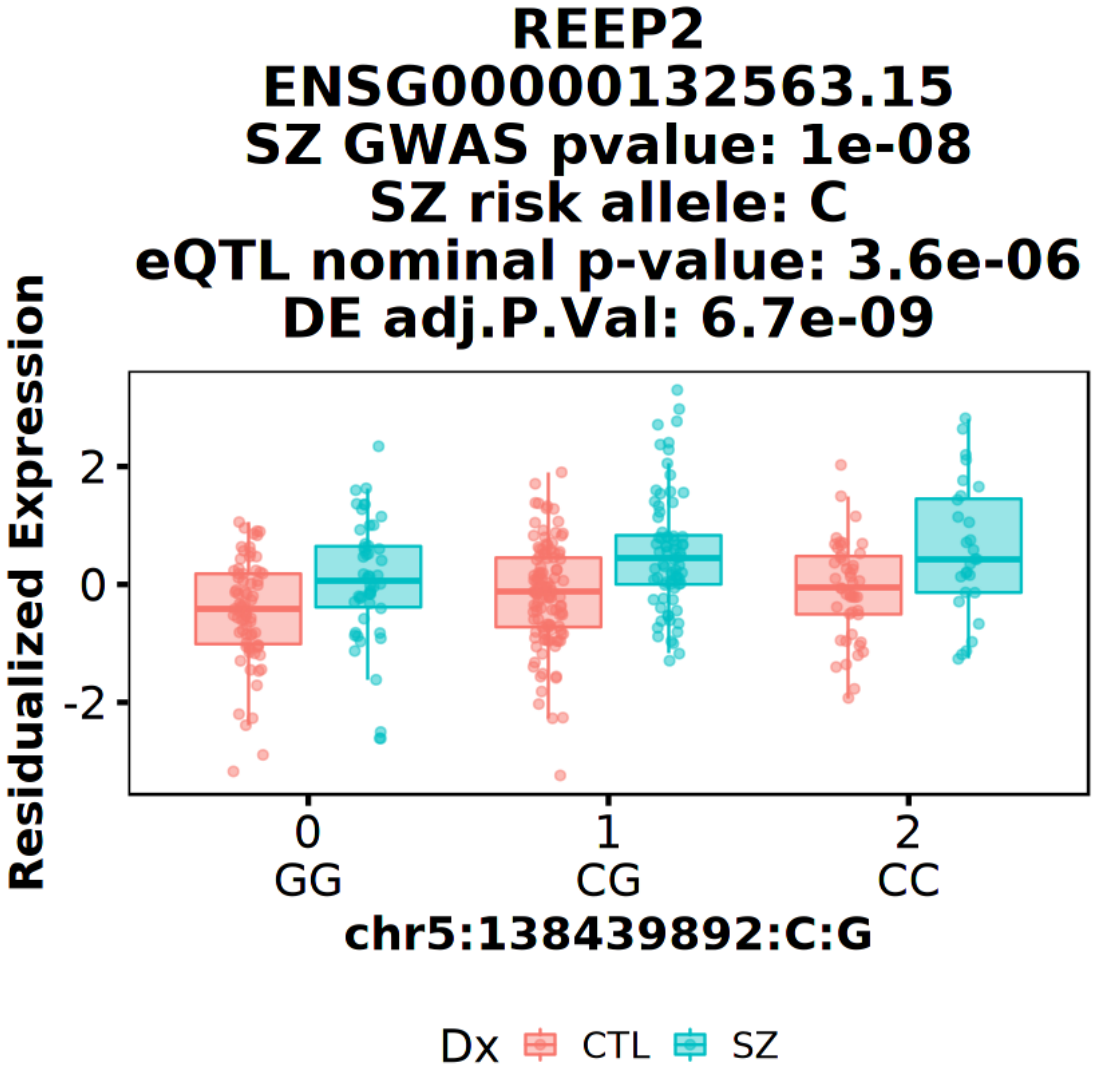

PLPP5
ENSG00000147535.16
SZ GWAS pvalue: 7.5e-10
SZ risk allele: A
eQTL nominal p-value: 0.00037
DE adj.P.Val: 0.028
Residualized Expression
2
0
-2
-4
0
GG
1
AG
2
AA
chr8:38294926:A:G
Dx
CTL
SZ

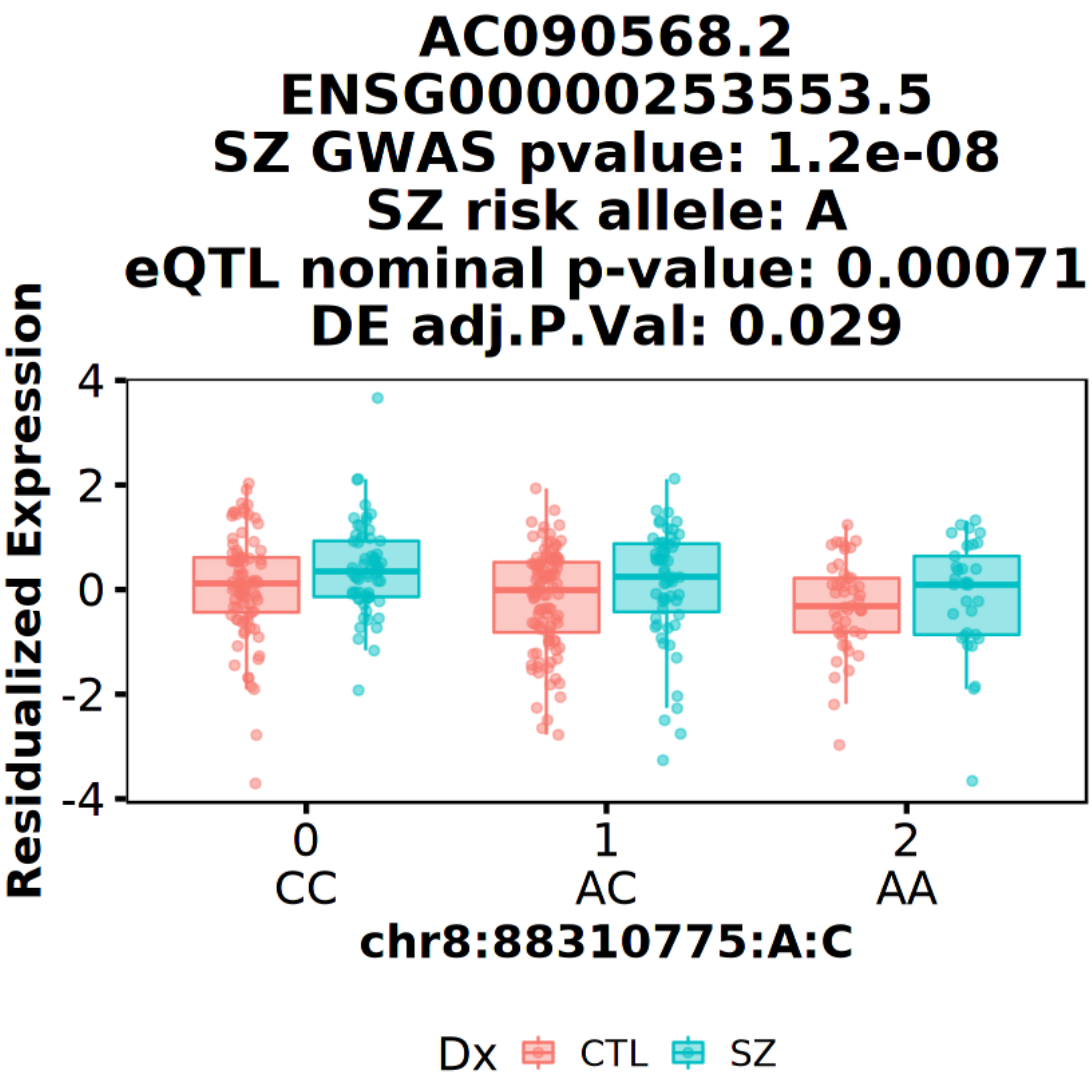

VRK2
ENSG00000028116.16
SZ GWAS pvalue: 7.8e-10
SZ risk allele: C
eQTL nominal p-value: 0.00045
DE adj.P.Val: 0.011
Residualized Expression
3
2
1
0
-1
-2
-3
0
AA
1
AC
2
CC
chr2:57944085:A:C
Dx
CTL
SZ

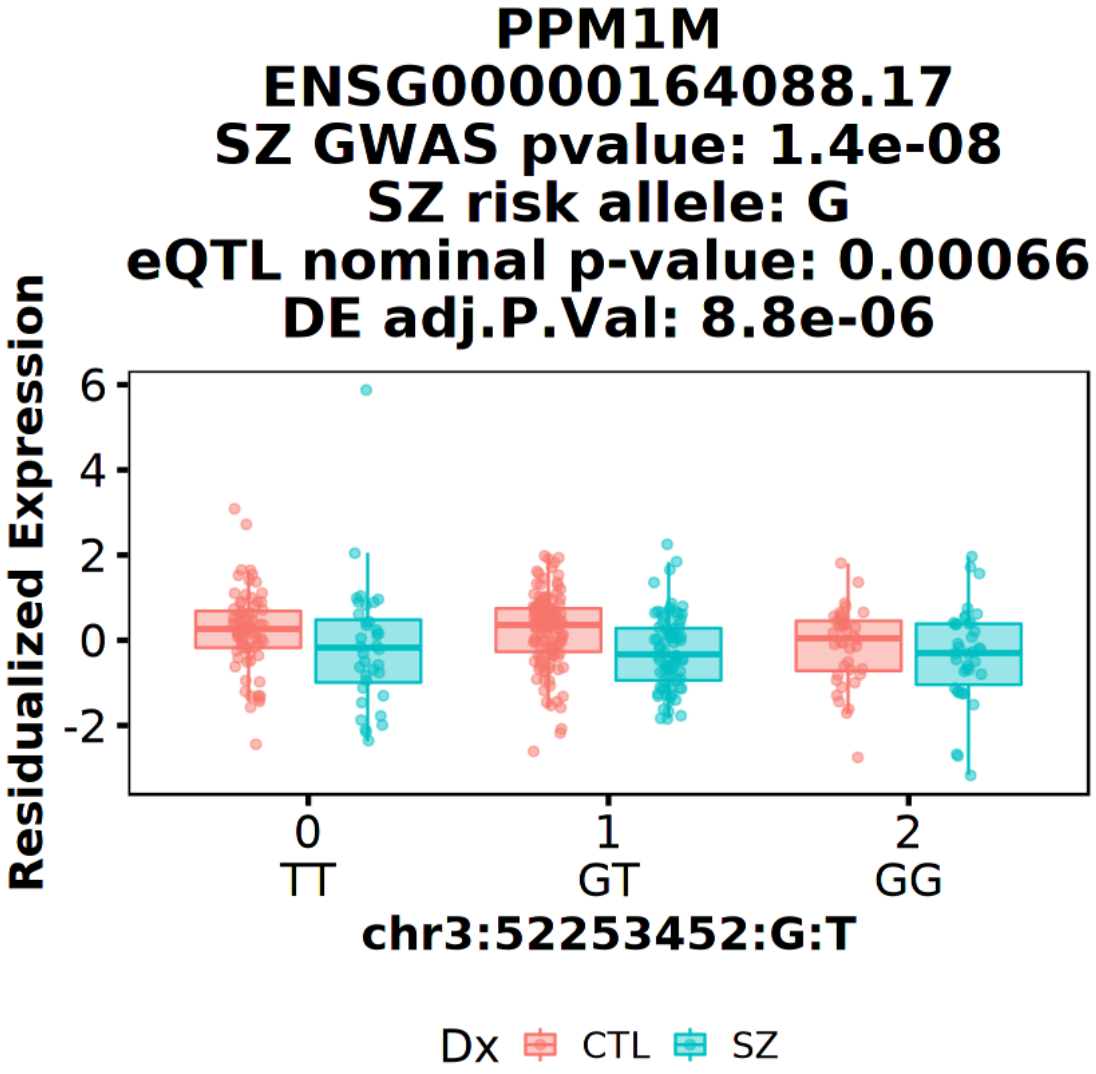

HIRIP3
ENSG00000149929.15
SZ GWAS pvalue: 3.1e-09
SZ risk allele: A
eQTL nominal p-value: 0.00034
DE adj.P.Val: 0.00021
Residualized Expression
2
0
-2
0
GG
1
AG
2
AA
chr16:29998953:A:G
Dx
CTL
SZ

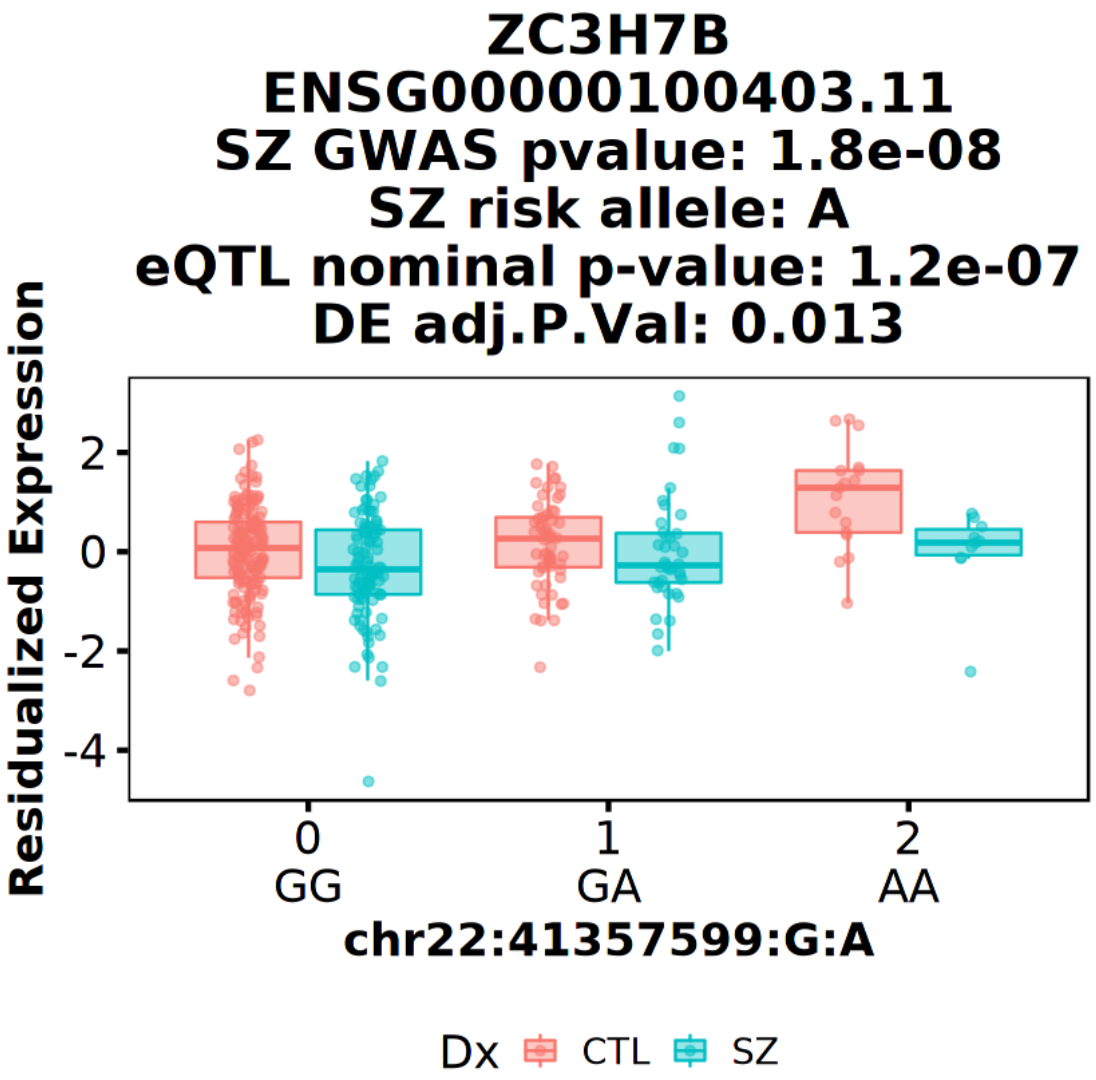

BNIP3L
ENSG00000104765.15
SZ GWAS pvalue: 3.7e-09
SZ risk allele: T
eQTL nominal p-value: 0.00012
DE adj.P.Val: 0.011
Residualized Expression
4
2
0
-2
0
CC
1
CT
2
TT
chr8:26411675:C:T
Dx
CTL
SZ

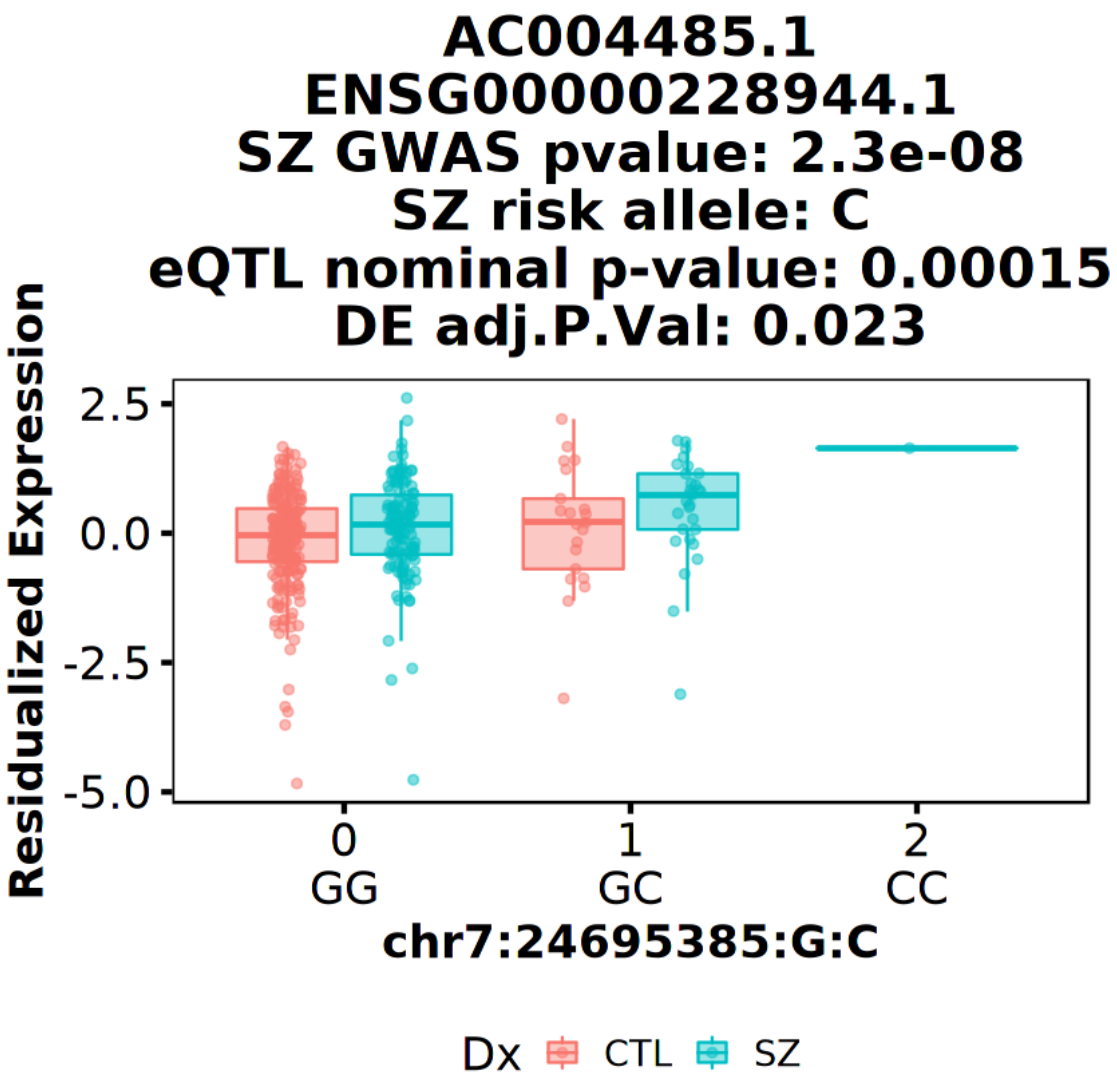

KDM3B
ENSG00000120733.13
SZ GWAS pvalue: 1e-08
SZ risk allele: C
eQTL nominal p-value: 8.8e-05
DE adj.P.Val: 0.015
Residualized Expression
4
2
0
-2
-4
0
GG
1
CG
2
CC
chr5:138439892:C:G
Dx
CTL
SZ

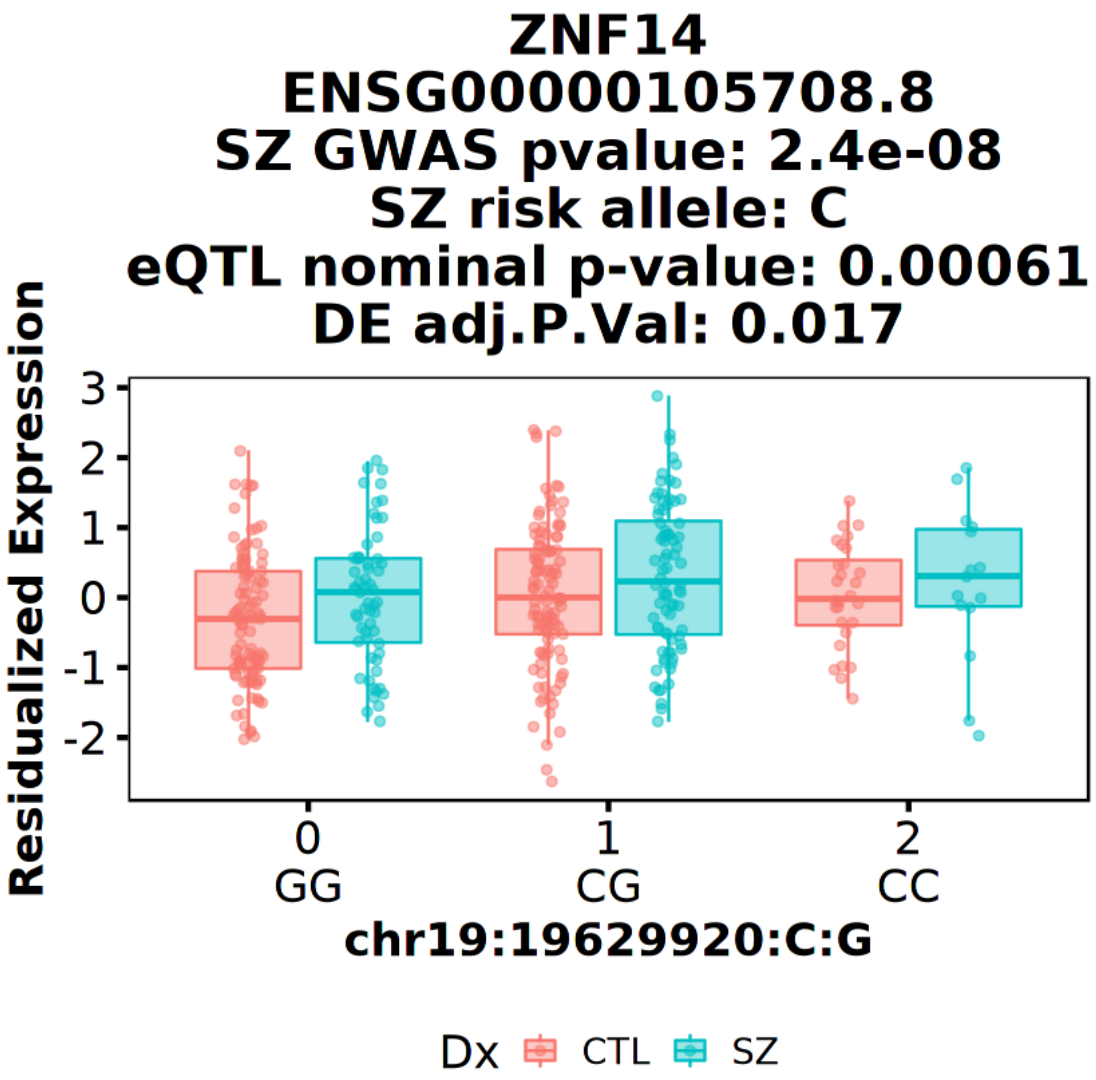

REEP2
ENSG00000132563.15
SZ GWAS pvalue: 1e-08
SZ risk allele: C
eQTL nominal p-value: 3.6e-06
DE adj.P.Val: 6.7e-09
Residualized Expression
2
0
-2
0
GG
1
CG
2
CC
chr5:138439892:C:G
Dx
CTL
SZ


AC090568.2
ENSG00000253553.5
SZ GWAS pvalue: 1.2e-08
SZ risk allele: A
eQTL nominal p-value: 0.00071
DE adj.P.Val: 0.029
Residualized Expression
4
2
0
-2
-4
0
CC
1
AC
2
AA
chr8:88310775:A:C
Dx
CTL
SZ


PPM1M
ENSG00000164088.17
SZ GWAS pvalue: 1.4e-08
SZ risk allele: G
eQTL nominal p-value: 0.00066
DE adj.P.Val: 8.8e-06
Residualized Expression
6
4
2
0
-2
0
TT
1
GT
2
GG
chr3:52253452:G:T
Dx
CTL
SZ


ZC3H7B
ENSG00000100403.11
SZ GWAS pvalue: 1.8e-08
SZ risk allele: A
eQTL nominal p-value: 1.2e-07
DE adj.P.Val: 0.013
Residualized Expression
2
0
-2
-4
0
GG
1
GA
2
AA
chr22:41357599:G:A
Dx
CTL
SZ


AC004485.1
ENSG00000228944.1
SZ GWAS pvalue: 2.3e-08
SZ risk allele: C
eQTL nominal p-value: 0.00015
DE adj.P.Val: 0.023
Residualized Expression
2.5
0.0
-2.5
-5.0
0
GG
1
GC
2
CC
chr7:24695385:G:C
Dx
CTL
SZ



1.3 Session Info

[14]:
```
Sys.time()
proc.time()
options(width = 120)
sessioninfo::session_info()
```

```
[1] "2021-09-08 11:39:47 EDT"
```

```
     user    system   elapsed
14588.436  5711.659  2791.350
```

```
 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
```