main

September 6, 2021

1 Generate boxplots for DRD2 expression

```
[1]: library(tidyverse)
     library(ggpubr)
```

```
  Attaching packages                              tidyverse
1.3.1

  ggplot2 3.3.5       purrr   0.3.4
  tibble  3.1.4       dplyr   1.0.7
  tidyr   1.1.3       stringr 1.4.0
  readr   2.0.1       forcats 0.5.1

  Conflicts
tidyverse_conflicts()
  dplyr::filter() masks stats::filter()
  dplyr::lag()    masks stats::lag()
```

1.1 Functions

1.1.1 Cached functions

```
[2]: get_pheno <- function(){
         df = data.table::fread("/ceph/projects/v4_phase3_paper/inputs/phenotypes/_m/
      caudate_phenotypes.csv") %>%
             select(RNum, BrNum, Dx, Sex, Race, Age)
         return(df)
     }
     memPHENO <- memoise::memoise(get_pheno)

     get_drd2_features <- function(feature){
         filepath = paste0("../../_m/", feature, "/diffExpr_szVctl_full.txt")
         cmd = paste("cat <(head -1",filepath,") <(grep -iw drd2", filepath, ")")
         df = data.table::fread(cmd=cmd) %>% rename("Feature"="V1") %>%
             select("Feature", "logFC", "adj.P.Val")
         return(df)
     }
```

```r
memDRD2 <- memoise::memoise(get_drd2_features)

get_exon_id <- function(){
    exon_file = "/ceph/projects/v3_phase3_paper/analysis/dopamine/exons/_m/
 exons.csv"
    exon_annotation = data.table::fread(exon_file) %>%
        select(brainseq_exon_id, exon_id, exon_number, gene_id, gene_name) %>%
        filter(gene_name == "DRD2")
    return(exon_annotation)
}
memEXONS <- memoise::memoise(get_exon_id)

get_resdf <- function(feature){
    filename = paste0("../../_m/", feature, "/residualized_expression.tsv")
    df = data.table::fread(filename) %>% filter(V1 %in%
 memDRD2(feature)$Feature) %>%
        tibble::column_to_rownames("V1") %>% t %>% as.data.frame %>%
        tibble::rownames_to_column("RNum")
    return(df)
}
memRES <- memoise::memoise(get_resdf)

merge_data <- function(feature){
    df = inner_join(memPHENO(), memRES(feature), by="RNum") %>%
        select(-c("BrNum", "Sex", "Race", "Age")) %>%
        pivot_longer(!c("RNum", "Dx"), names_to="Feature",
 values_to="Residualized") %>%
        inner_join(memDRD2(feature), by="Feature")
    return(df)
}
memDF <- memoise::memoise(merge_data)
```

1.1.2 Simple functions

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

annotate_drd2_exons <- function(){
    bs_exon_id = c('e667152', 'e667153', 'e667154', 'e667155', 'e667156',
                   'e667157', 'e667158', 'e667159', 'e667164')
    annot = c('DRD2 exon 1S', 'DRD2 exon 2', 'DRD2 exon 3', 'DRD2 exon 4',
              'DRD2 exon 5', 'DRD2 exon 7', 'DRD2 exon 8', 'DRD2 exon 1L',
              'DRD2 exon 6')
    df = memEXONS() %>% select(brainseq_exon_id, exon_id) %>%
```

```
        filter(brainseq_exon_id %in% bs_exon_id) %>% distinct() %>%
        group_by(brainseq_exon_id) %>% slice(1) %>% as.data.frame %>%
        mutate("DRD2_ID"=annot)
    return(df)
}


annotate_drd2_junctions <- function(){
    feature_id = c("chr11:113424683-113475075(-)", "chr11:
 ↪113418137-113424366(-)",
                   "chr11:113417000-113418026(-)", "chr11:
 ↪113415612-113416862(-)",
                   "chr11:113414462-113415420(-)", "chr11:
 ↪113412884-113415420(-)",
                   "chr11:113412884-113414374(-)", "chr11:
 ↪113410921-113412555(-)")
    drd2_id = c("DRD2 junction 1-2", "DRD2 junction 2-3", "DRD2 junction 3-4",
                "DRD2 junction 4-5", "DRD2 junction 5-6", "DRD2 junction 5-7",
                "DRD2 junction 6-7", "DRD2 junction 7-8")
    return(data.frame("Feature"=feature_id, "DRD2_ID"=drd2_id))
}
```

1.2 Generate annotated boxplots
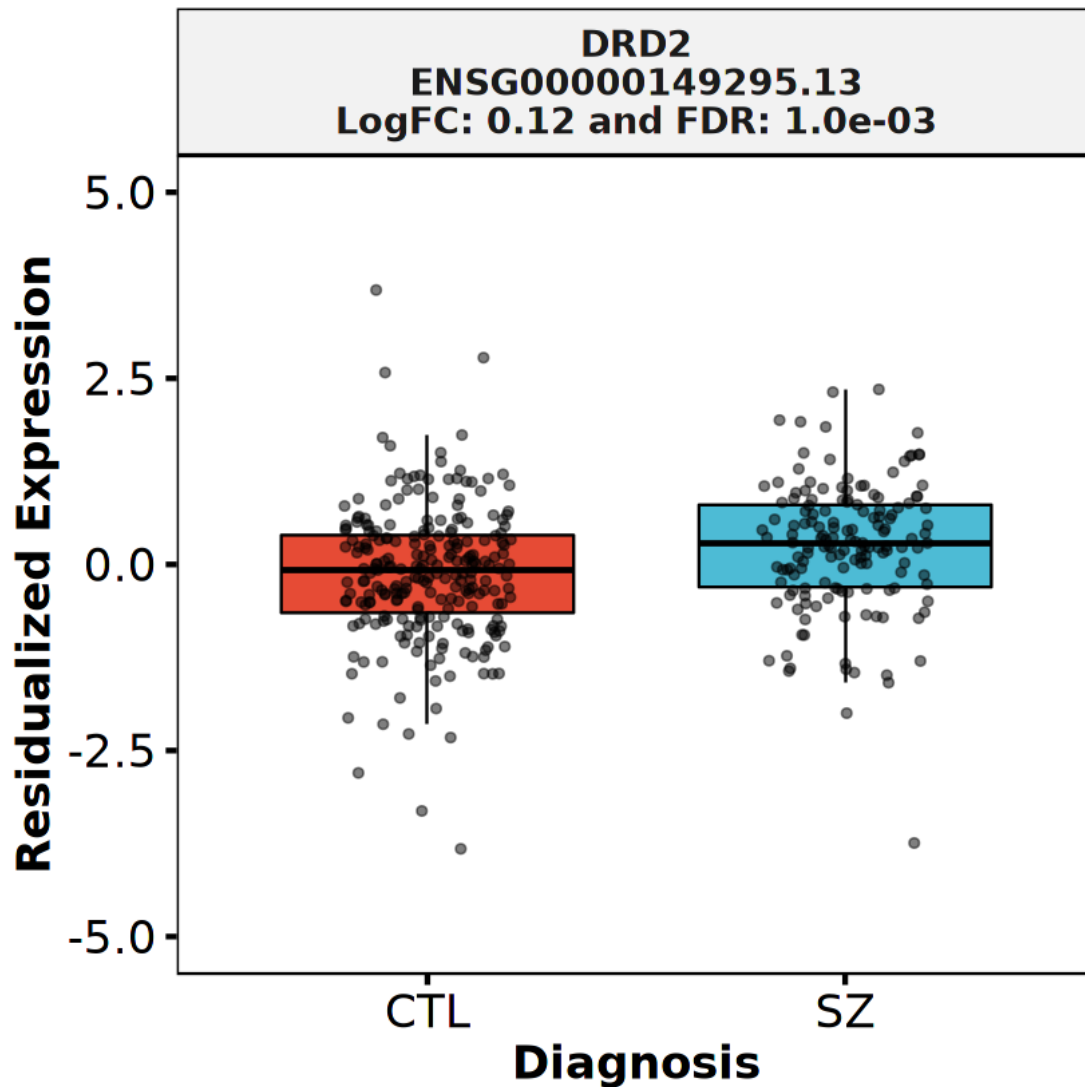
1.2.1 Genes

```
[4]: feature = "genes"

bxp = memDF(feature) %>%
    mutate(ID=paste0("DRD2\n", Feature, "\nLogFC: ", round(logFC, 2), " and FDR:
 ↪ ",
                     formatC(`adj.P.Val`, format = "e", digits = 1))) %>%
    mutate_if(is.character, as.factor) %>%
    ggboxplot(x="Dx", y="Residualized", fill="Dx", facet.by="ID", palette="npg",
              add="jitter", ggtheme=theme_pubr(base_size=20, border=TRUE),␣
 ↪xlab="Diagnosis",
              panel.labs.font=list(face='bold'), ylab="Residualized Expression",
              add.params=list(alpha=0.5), outlier.shape=NA, ylim=c(-5, 5),
              legend="None") +
    font("xy.title", face="bold")
save_ggplots(bxp, "drd2_gene_boxplot", 6, 6)
bxp
```

Warning message in data.table::fread(filename):
"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."

1.3 Transcripts

```
[5]: feature = "transcripts"

bxp = memDF(feature) %>% filter(Feature %in% c("ENST00000362072.7",␣
 ↪"ENST00000346454.7")) %>%
    mutate(DRD2_ID=ifelse(Feature == "ENST00000362072.7", "DRD2 Long", "DRD2␣
 ↪Short")) %>%
    mutate(ID=paste0(DRD2_ID, "\n", Feature, "\nLogFC: ", round(logFC, 2), "␣
 ↪and FDR: ",
                    formatC(`adj.P.Val`, format = "e", digits = 1))) %>%
    mutate_if(is.character, as.factor) %>%
```
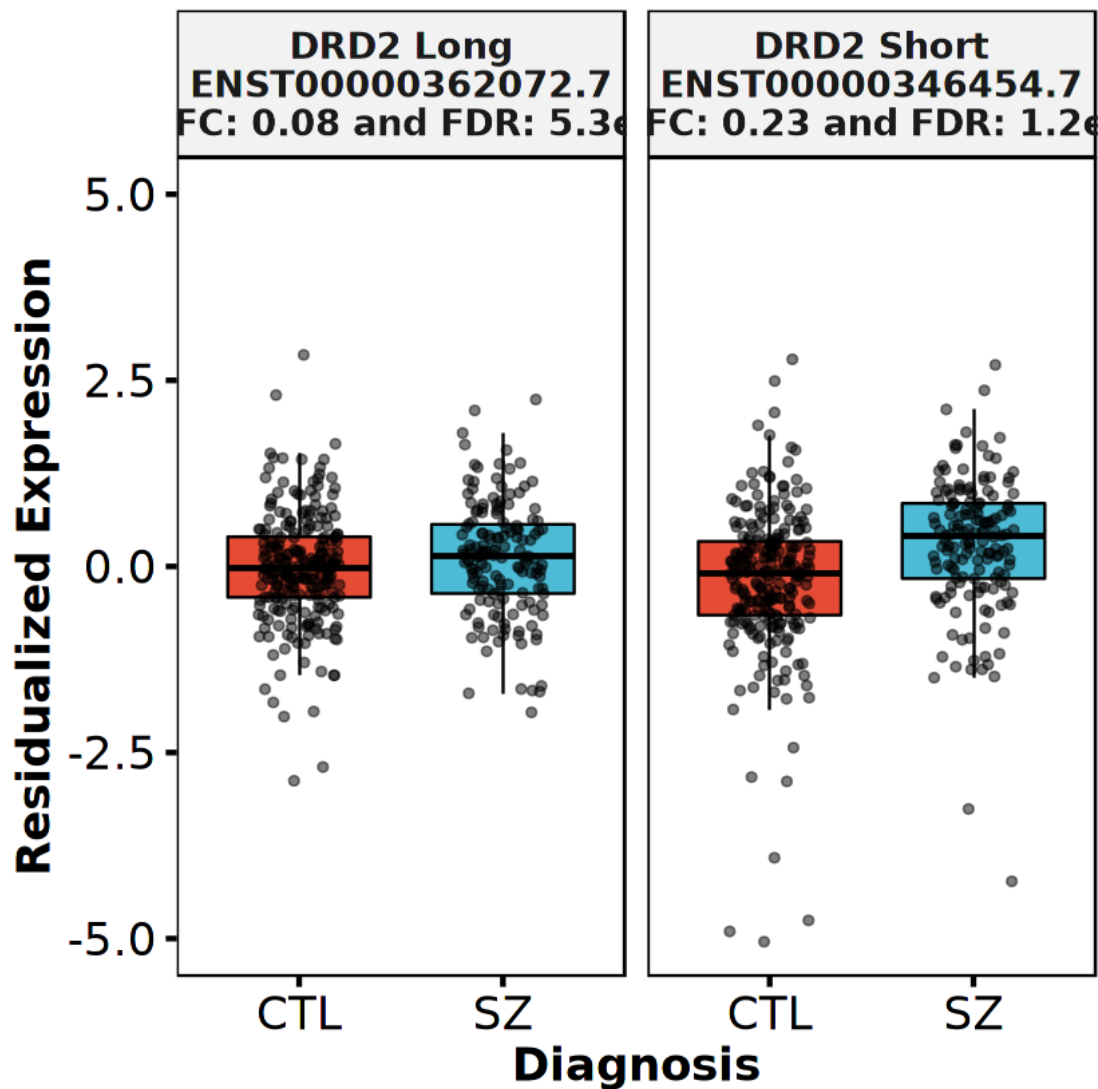
```
    ggboxplot(x="Dx", y="Residualized", fill="Dx", facet.by="ID", palette="npg",
              add="jitter", ggtheme=theme_pubr(base_size=20, border=TRUE),
              xlab="Diagnosis", legend="None", ylim=c(-5, 5), outlier.shape=NA,
              panel.labs.font=list(face='bold'), ylab="Residualized Expression",
              add.params=list(alpha=0.5)) +
    font("xy.title", face="bold")
save_ggplots(bxp, "drd2_transcripts_boxplot", 9, 6)
bxp
```

Warning message in data.table::fread(filename):
"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."

```
[6]: dir.create(feature)
     for(feature_id in c("ENST00000362072.7", "ENST00000346454.7")){
         drd2_id = ifelse(feature_id == "ENST00000362072.7", "DRD2 Long", "DRD2␣
     ↪Short")
         bxp = memDF(feature) %>% filter(Feature == feature_id) %>%␣
     ↪mutate(DRD2_ID=drd2_id) %>%
             mutate(ID=paste0(DRD2_ID, "\n", Feature, "\nLogFC: ", round(logFC, 2),␣
     ↪" and FDR: ",
                             formatC(`adj.P.Val`, format = "e", digits = 1))) %>%
             mutate_if(is.character, as.factor) %>%
             ggboxplot(x="Dx", y="Residualized", fill="Dx", facet.by="ID",␣
     ↪palette="npg",
                     add="jitter", ggtheme=theme_pubr(base_size=20, border=TRUE),
                     xlab="Diagnosis", outlier.shape=NA, ylim=c(-5, 5),␣
     ↪legend="None",
                     panel.labs.font=list(face='bold'), ylab="Residualized␣
     ↪Expression",
                     add.params=list(alpha=0.5)) +
             font("xy.title", face="bold")
         save_ggplots(bxp, paste0(feature,"/drd2_transcripts_boxplot_", gsub(" ",␣
     ↪"_", drd2_id)), 6, 6)
     }
```
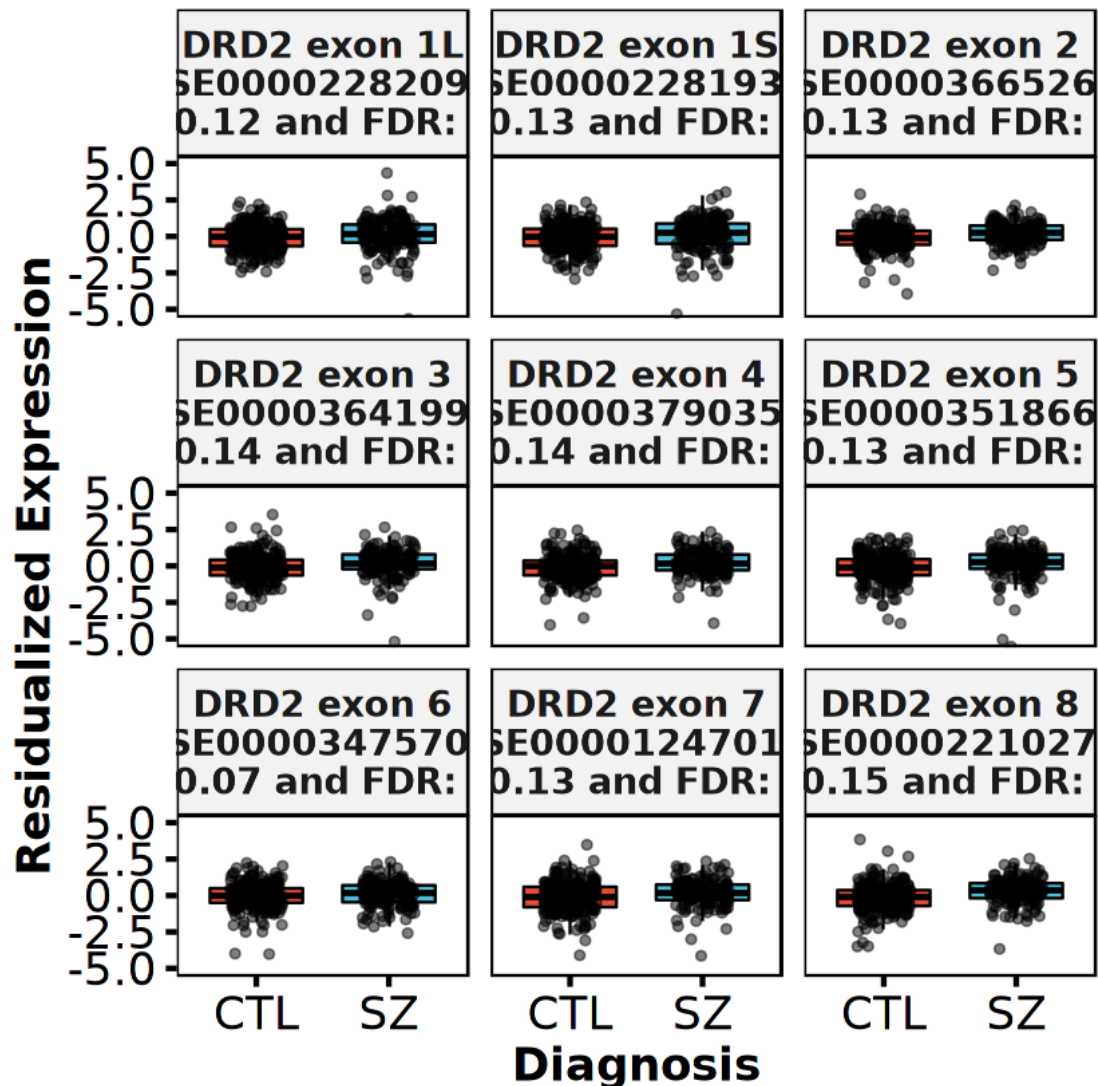
1.4   Exons

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

     bxp = memDF(feature) %>% inner_join(annotate_drd2_exons(),␣
      ↪by=c("Feature"="brainseq_exon_id")) %>%
         mutate(ID=paste0(DRD2_ID, "\n", exon_id, "\nLogFC: ", round(logFC, 2), "␣
      ↪and FDR: ",
                         formatC(`adj.P.Val`, format = "e", digits = 1))) %>%
         mutate_if(is.character, as.factor) %>%
         ggboxplot(x="Dx", y="Residualized", fill="Dx", facet.by="ID", palette="npg",
                 add="jitter", ggtheme=theme_pubr(base_size=20, border=TRUE),
                 xlab="Diagnosis", outlier.shape=NA, ylim=c(-5, 5), legend="None",
                 panel.labs.font=list(face='bold'), ylab="Residualized Expression",
                 add.params=list(alpha=0.5)) +
         font("xy.title", face="bold")
     save_ggplots(bxp, "drd2_exons_boxplot", 12, 12)
     bxp
```

```
Warning message in data.table::fread(filename):
"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."


```
[8]: dir.create(feature)
     for(feature_id in annotate_drd2_exons()$brainseq_exon_id){
         drd2_id = filter(annotate_drd2_exons(), brainseq_exon_id ==␣
     ↪feature_id)$DRD2_ID
        bxp = memDF(feature) %>% filter(Feature == feature_id) %>%
            inner_join(annotate_drd2_exons(), by=c("Feature"="brainseq_exon_id"))␣
     ↪%>%
            mutate(ID=paste0(DRD2_ID, "\n", exon_id, "\nLogFC: ", round(logFC, 2),␣
     ↪" and FDR: ",
                        formatC(`adj.P.Val`, format = "e", digits = 1))) %>%
```

```
        mutate_if(is.character, as.factor) %>%
        ggboxplot(x="Dx", y="Residualized", fill="Dx", facet.by="ID",␣
↪palette="npg",
                  add="jitter", ggtheme=theme_pubr(base_size=20, border=TRUE),
                  xlab="Diagnosis", outlier.shape=NA, ylim=c(-5, 5),␣
↪legend="None",
                  panel.labs.font=list(face='bold'), ylab="Residualized␣
↪Expression",
                  add.params=list(alpha=0.5)) +
        font("xy.title", face="bold")
    save_ggplots(bxp, paste0(feature,"/drd2_exons_boxplot_", gsub(" ", "_",␣
↪drd2_id)), 6, 6)
}
```
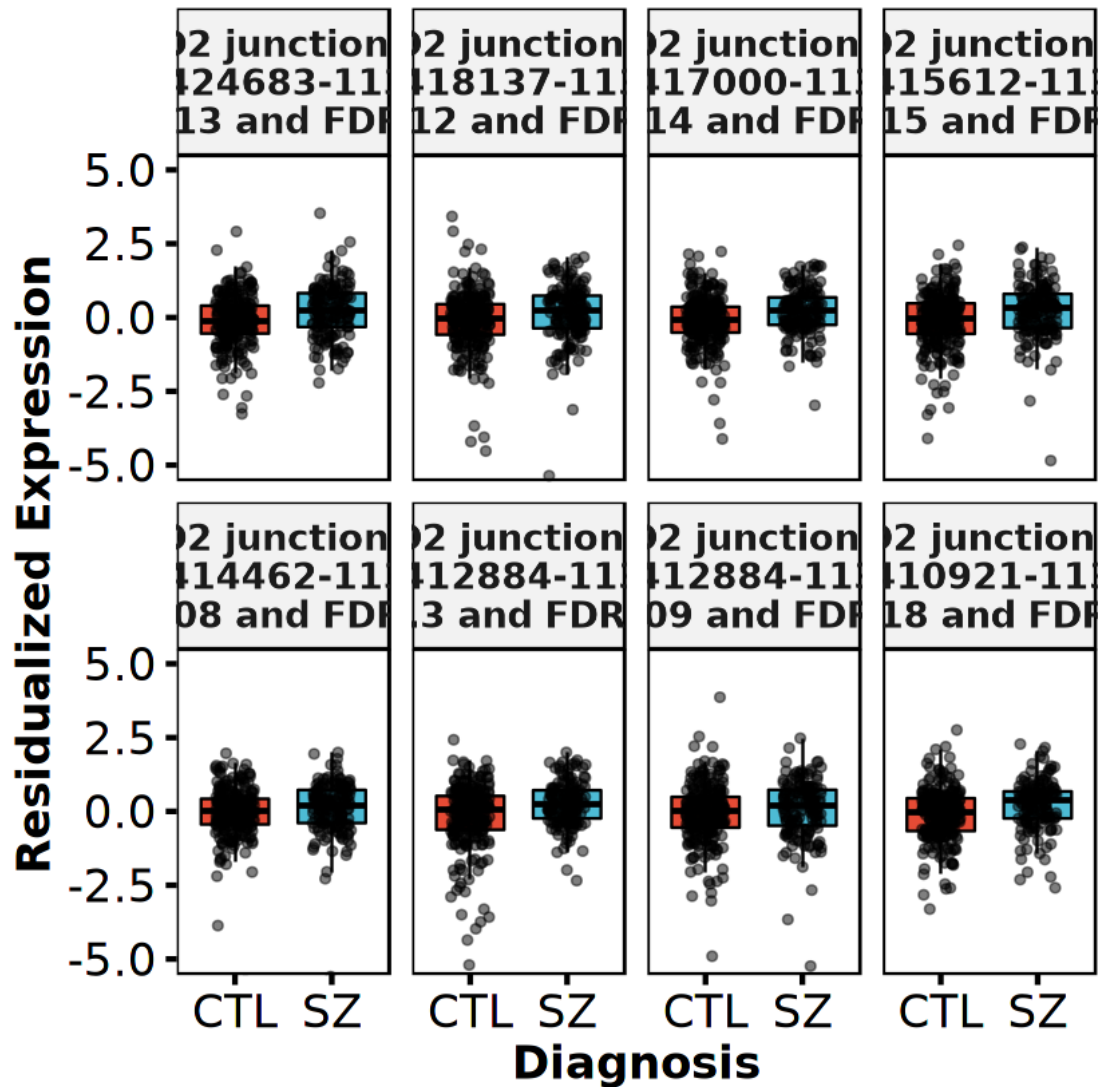
1.5   Junctions

```
[9]: feature = "junctions"

bxp = memDF(feature) %>% inner_join(annotate_drd2_junctions(), by=c("Feature"))␣
↪%>%
    mutate(ID=paste0(DRD2_ID, "\n", Feature, "\nLogFC: ", round(logFC, 2), "␣
↪and FDR: ",
                     formatC(`adj.P.Val`, format = "e", digits = 1))) %>%
    mutate_if(is.character, as.factor) %>%
    ggboxplot(x="Dx", y="Residualized", fill="Dx", facet.by="ID", palette="npg",
              add="jitter", ggtheme=theme_pubr(base_size=20, border=TRUE),
              xlab="Diagnosis", outlier.shape=NA, ylim=c(-5, 5), legend="None",
              panel.labs.font=list(face='bold'), ylab="Residualized Expression",
              add.params=list(alpha=0.5), ncol=4) +
    font("xy.title", face="bold")
save_ggplots(bxp, "drd2_junctions_boxplot", 16, 10)
bxp
```

```
Warning message in data.table::fread(filename):
"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."
```

```
[10]: dir.create(feature)
      for(feature_id in annotate_drd2_junctions()$Feature){
          drd2_id = filter(annotate_drd2_junctions(), Feature == feature_id)$DRD2_ID
          bxp = memDF(feature) %>% filter(Feature == feature_id) %>%
              inner_join(annotate_drd2_junctions(), by=c("Feature")) %>%
              mutate(ID=paste0(DRD2_ID, "\n", Feature, "\nLogFC: ", round(logFC, 2),␣
      ↪" and FDR: ",
                                  formatC(`adj.P.Val`, format = "e", digits = 1))) %>%
              mutate_if(is.character, as.factor) %>%
              ggboxplot(x="Dx", y="Residualized", fill="Dx", facet.by="ID",␣
      ↪palette="npg",
                      add="jitter", ggtheme=theme_pubr(base_size=20, border=TRUE),
```

```
                xlab="Diagnosis", outlier.shape=NA, ylim=c(-5, 5),␣
→legend="None",
                panel.labs.font=list(face='bold'), ylab="Residualized␣
→Expression",
                add.params=list(alpha=0.5)) +
        font("xy.title", face="bold")
    save_ggplots(bxp, paste0(feature, "/drd2_junctions_boxplot_",
                            gsub("-", "_", gsub(" ", "_", drd2_id)))), 6, 6)
}
```

1.6 Session Info

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

```
[1] "2021-09-06 08:47:21 EDT"

   user  system elapsed
 75.067  41.450  90.154

 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     2021-09-06

 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)
ggsci         2.9       2018-05-14 [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)
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
```