main

July 15, 2021

1 Boxplot of DE junctions

```
[1]: library(ggpubr)
     library(ggplot2)
     library(tidyverse)
     suppressMessages(library(data.table))
```

```
Loading required package: ggplot2

  Attaching packages                                    tidyverse
1.3.1

  tibble  3.1.2        dplyr   1.0.7
  tidyr   1.1.3        stringr 1.4.0
  readr   1.4.0        forcats 0.5.1
  purrr   0.3.4

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

```
[2]: ggplot_save <- function(p, fn, w=7, h=7){
         for(ext in c('.svg', '.png', '.pdf')){
             ggsave(p, filename=paste0(fn, ext), width=w, height=h)
         }
     }


     subset_deg <- function(fn, thres){
         jt = fread(fn) %>% rename(Feature='V1') %>%
             filter(adj.P.Val < thres) %>%
             select(Feature, ensemblID, Symbol, logFC, adj.P.Val)
         return(jt)
     }
```

```
add_symnum <- function(res){
    symnum.args <- list(cutpoints = c(0, 0.0001, 0.001, 0.01, 0.05,  1),
                        symbols = c("****", "***", "**", "*",  "ns"))
    symnum.args$x <- res$adj.P.Val
    pvalue.signif <- do.call(stats::symnum, symnum.args) %>%
        as.character()
    pvalue.format <- format.pval(res$adj.P.Val, digits = 2)
    res <- res %>%
        dplyr::ungroup() %>%
        mutate(FDR = pvalue.format, p.signif = pvalue.signif)
    return(res)
}
```

1.1 Load DEG annotation

```
[3]: fn1 = '../../_m/junctions/diffExpr_interaction_full.txt'
     fn2 = '../../_m/junctions/diffExpr_maleVfemale_full.txt'
     fn3 = '../../_m/junctions/diffExpr_szVctl_full.txt'
     juncs = subset_deg(fn1, 0.05)
     sex = subset_deg(fn2, 1)
     sz = subset_deg(fn3, 1)
```

```
[4]: df = juncs %>%
         inner_join(sex, by=c("Feature", "ensemblID", "Symbol"), suffix=c('',␣
     ↪"_Sex")) %>%
         inner_join(sz, by=c("Feature", "ensemblID", "Symbol"), suffix=c('',␣
     ↪"_Diagnosis"))
     df = add_symnum(df)
     df
```

	Feature	ensemblID	Symbol	logFC	adj.P.Val	logFC_S
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
	chr21:8401487-8445715(-)			6.462429	1.977227e-06	-0.690033
	chr21:8401417-8445654(-)			3.818954	2.615920e-05	-0.241479
A data.table: 7 × 11	chr14:49566087-49586737(+)			-4.428507	1.286513e-04	1.9490009
	chr21:8401395-8445553(-)			5.809398	3.625764e-04	0.5161015
	chr21:8401970-8446050(+)			6.047697	6.728559e-04	-3.459245
	chrM:2246-2357(+)			3.699453	6.038357e-03	-2.324742
	chr2:132254419-132255143(+)			-3.425367	3.939449e-02	1.7834593

1.2 Load phenotypes

```
[5]: pheno_file = '/ceph/projects/v3_phase3_paper/inputs/phenotypes/_m/
     ↪dlpfc_phenotypes.csv'
     pheno = fread(pheno_file) %>% column_to_rownames(var='V1') %>%
         mutate_if(is.character, as.factor)
     levels(pheno$Sex) <- c("Female", "Male")
```

```
levels(pheno$Dx) <- c("CTL", "SZ")
pheno %>% head(2)
```

A data.frame: 2 × 8

	BrNum	RNum	Region	RIN	Age	Sex	Race	Dx
	\<fct\>	\<fct\>	\<fct\>	\<dbl\>	\<dbl\>	\<fct\>	\<fct\>	\<fct\>
R10424	Br5168	R10424	DLPFC	6.7	64.08	Male	CAUC	CTL
R12195	Br5073	R12195	DLPFC	8.4	62.61	Male	AA	SZ

1.3 Load residualized expression

```
[6]: res_file = '../../_m/junctions/residualized_expression.tsv'
resdf = fread(res_file) %>%
    column_to_rownames(var="V1") %>% t
resdf = as.data.frame(resdf) %>%
    rownames_to_column() %>%
    rename(RNum=rowname) %>%
    select(RNum, df$Feature)
resdf[1:2, 1:3]
```

```
Warning message in fread(res_file):
"Detected 359 column names but the data has 360 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."
```

A data.frame: 2 × 3

	RNum	chr21:8401487-8445715(-)	chr21:8401417-8445654(-)
	\<chr\>	\<dbl\>	\<dbl\>
1	R10424	0.1357274	-0.02991936
2	R12195	-0.8260069	-1.15152280

1.4 Merge data

```
[7]: jj = pheno %>%
    inner_join(resdf, by='RNum')
jj$Dx = fct_drop(jj$Dx)
jj %>% dim
jj[1:2, 1:10]
```

1. 359 2. 15

A data.frame: 2 × 10

	BrNum	RNum	Region	RIN	Age	Sex	Race	Dx	chr21:84014...
	\<fct\>	\<chr\>	\<fct\>	\<dbl\>	\<dbl\>	\<fct\>	\<fct\>	\<fct\>	\<dbl\>
1	Br5168	R10424	DLPFC	6.7	64.08	Male	CAUC	CTL	0.1357274
2	Br5073	R12195	DLPFC	8.4	62.61	Male	AA	SZ	-0.8260069

```
[8]: dft = jj %>% pivot_longer(-c(BrNum, RNum, Region, RIN, Age, Sex, Race, Dx),
                    names_to="Feature", values_to="Res") %>%
    inner_join(df, by="Feature") %>% mutate_at("Feature", as.factor)
dft[1:2, 1:10]
```

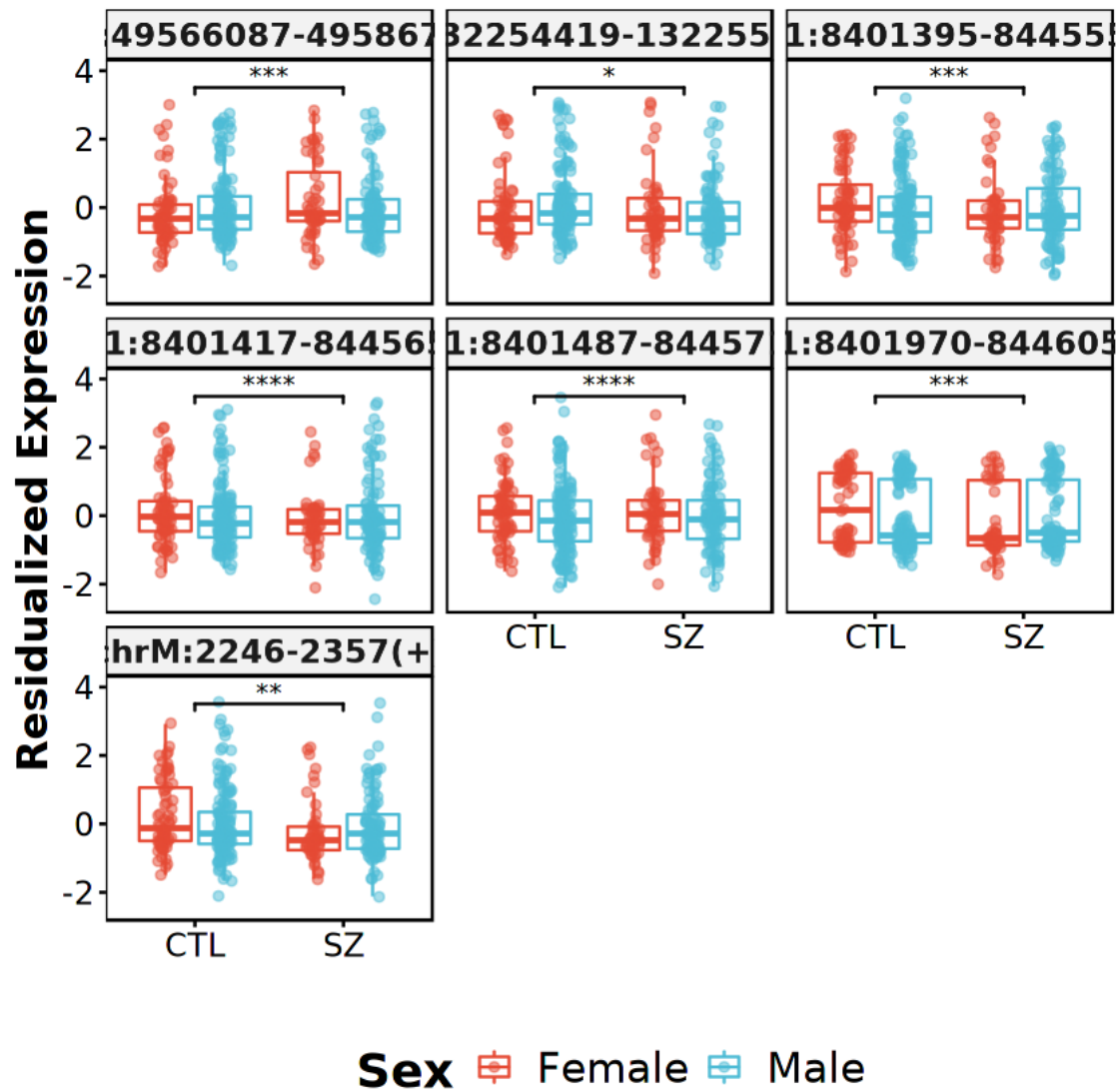| A tibble: 2 × 10	BrNum <fct>	RNum <chr>	Region <fct>	RIN <dbl>	Age <dbl>	Sex <fct>	Race <fct>	Dx <fct>	Feature <fct>
	Br5168	R10424	DLPFC	6.7	64.08	Male	CAUC	CTL	chr21:8401487-84457
	Br5168	R10424	DLPFC	6.7	64.08	Male	CAUC	CTL	chr21:8401417-84456

```
[9]: tmp = df %>% mutate(group1='CTL', group2='SZ', y_pos=3.5)

     bxp <- ggboxplot(dft, x="Dx", y="Res", color="Sex", add='jitter',
                      panel.labs.font=list(face='bold', size = 14),
                      facet.by=c("Feature"), palette="npg", xlab='',
                      ylim=c(-2.5, 4), ylab='Residualized Expression',
                      add.params=list(alpha=0.5), legend="bottom",
                      outlier.shape=NA) +
             geom_signif(data=tmp,
                     aes(xmin=group1, xmax=group2,
                         annotations=p.signif, y_position=y_pos),
                     manual=TRUE) +
             font("xy.title", size=18, face="bold") + font("xy.text", size=14) +
             font("legend.title", size=21, face="bold") + font("legend.text",␣
      ↪size=18)

     bxp
```

Warning message:
"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"

Sex ⊟ Female ⊟ Male

```
[10]: ggplot_save(bxp, 'junc_interaction', 10, 9)
```

1.5 Session Info

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

```
[1] "2021-07-15 20:38:12 EDT"

   user  system elapsed
 24.413  18.587  21.605
```

```
 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      2021-07-15

 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           0.7.8    2021-06-24 [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             3.0.0    2021-06-30 [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)
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           2.4.1    2021-04-23 [1] CRAN (R 4.0.3)
hms             1.1.0    2021-05-17 [1] CRAN (R 4.0.3)
```

```
htmltools      0.5.1.1   2021-01-22 [1] CRAN (R 4.0.2)
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)
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         1.6.1     2021-05-16 [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             2.5.0     2020-10-28 [1] CRAN (R 4.0.2)
Rcpp           1.0.7     2021-07-07 [1] CRAN (R 4.0.3)
readr        * 1.4.0     2020-10-05 [1] CRAN (R 4.0.2)
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         2.0.0     2021-04-02 [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          1.0.0     2021-03-09 [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        1.6.2     2021-05-17 [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       * 3.1.2     2021-05-16 [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)
utf8           1.2.1     2021-03-12 [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
```