

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 <chr>	ensemblID <chr>	Symbol <chr>	logFC <dbl>	adj.P.Val <dbl>	logFC_Sex <dbl>
A data.table: 2 × 11	chr21:8402231-8402287(+)			-5.005042	0.006971000	3.218546
	chr21:8402231-8402287(-)			-6.033730	0.008380723	4.248603

1.2 Load phenotypes

```

[5]: pheno_file = '/ceph/projects/v3_phase3_paper/inputs/phenotypes/_m/
      ↪ caudate_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("BD", "CTL", "SZ")
      pheno %>% head(2)

```

		BrNum	RNum	Region	RIN	Age	Sex	Race	Dx
		<fct>	<fct>	<fct>	<dbl>	<dbl>	<fct>	<fct>	<fct>
A data.frame: 2 × 8	R12864	Br1303	R12864	Caudate	9.6	42.98	Female	AA	SZ
	R12865	Br1320	R12865	Caudate	9.5	53.12	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 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."

		RNum	chr21:8402231-8402287(+)	chr21:8402231-8402287(-)
		<chr>	<dbl>	<dbl>
A data.frame: 2 × 3	1	R12864	1.7741225	1.6403453
	2	R12865	-0.5589144	-0.5376838

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. 393 2. 10

		BrNum	RNum	Region	RIN	Age	Sex	Race	Dx	chr21:8402231-8402287(+)
		<fct>	<chr>	<fct>	<dbl>	<dbl>	<fct>	<fct>	<fct>	<dbl>
A data.frame: 2 × 10	1	Br1303	R12864	Caudate	9.6	42.98	Female	AA	SZ	1.7741225
	2	Br1320	R12865	Caudate	9.5	53.12	Male	AA	SZ	-0.5589144

```
[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]
```

	BrNum	RNum	Region	RIN	Age	Sex	Race	Dx	Feature
	<fct>	<chr>	<fct>	<dbl>	<dbl>	<fct>	<fct>	<fct>	<fct>
A tibble: 2 × 10	Br1303	R12864	Caudate	9.6	42.98	Female	AA	SZ	chr21:8402231-8402
	Br1303	R12864	Caudate	9.6	42.98	Female	AA	SZ	chr21:8402231-8402

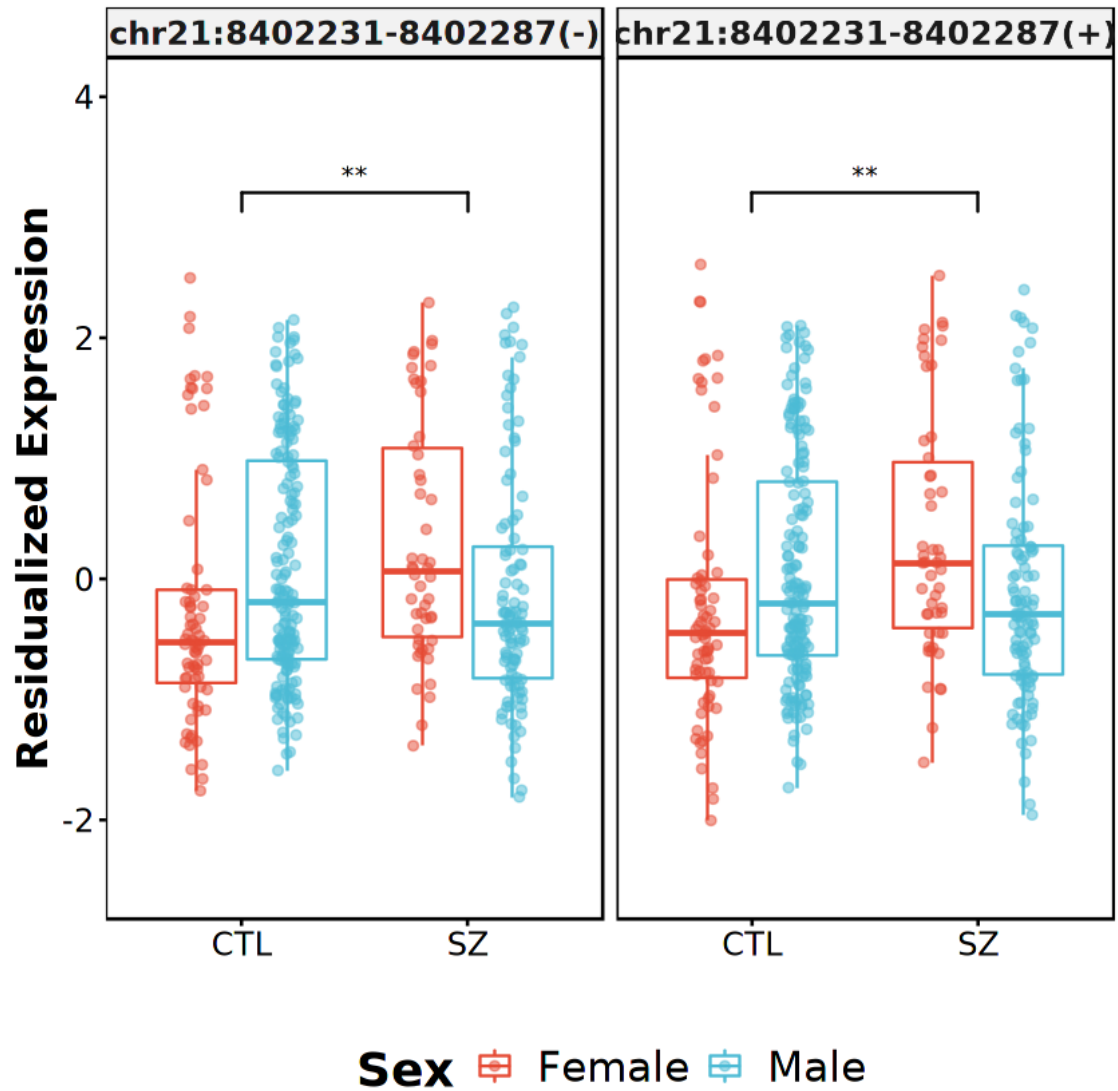
```
[9]: tmp = df %>% mutate(group1='CTL', group2='SZ', y_pos=3.2)

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"



```
[10]: ggplot_save(bxp, 'junc_interaction', 6, 4.2)
```

1.5 Session Info

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

```
[1] "2021-07-15 20:23:42 EDT"
```

```
user system elapsed
25.670 24.963 22.894
```

```

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)
fanshi       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