main_r

July 9, 2021

1 Generate a prettier plot with statistics on the plot

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

```
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]:
```
config <- list('caudate'= '../../../_m/genes/diffExpr_szVctl_full.txt',
               'dlpfc'= '/ceph/projects/v4_phase3_paper/inputs/public_data/_m/
↪phase2/dlpfc_diffExpr_szVctl_full.txt',
               'hippo'= '/ceph/projects/v4_phase3_paper/inputs/public_data/_m/
↪phase2/hippo_diffExpr_szVctl_full.txt')
```

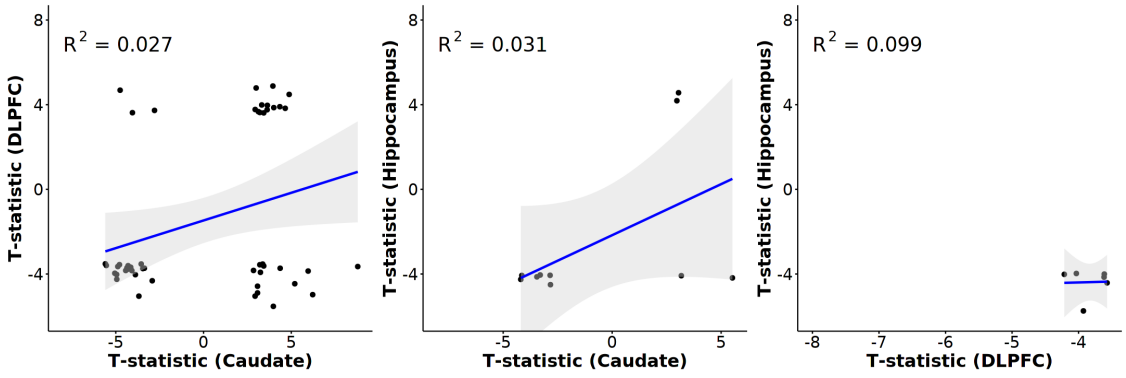
[3]:
```
get_deg <- function(fn){
    dft <- data.table::fread(fn)
    if('gene_id' %in% colnames(dft)){
        dft <- dft %>%
            mutate(Feature=gene_id, Dir=sign(t)) %>%
            rename(ensemblID=ensembl_gene_id) %>%
            select('Feature', 'ensemblID', 'adj.P.Val', 'logFC', 't', 'Dir')
    } else if ('gencodeID' %in% colnames(dft)){
        dft <- dft %>%
            mutate(Feature=gencodeID, Dir=sign(t)) %>%
```

```
            select("Feature", "ensemblID", "adj.P.Val", "logFC", "t", "Dir")
    } else {
        dft <- dft %>%
            mutate(Feature=V1, Dir=sign(dft$t)) %>%
            select('Feature', 'ensemblID', 'adj.P.Val', 'logFC', 't', 'Dir')
    }
    return(dft)
}

get_deg_sig <- function(fn, fdr){
    dft <- get_deg(fn)
    return(subset(dft, adj.P.Val < fdr))
}

merge_dataframe <- function(tissue1, tissue2){
    return(merge(get_deg(config[[tissue1]]), get_deg(config[[tissue2]]),
                 by='Feature', suffixes=c(paste0('_',tissue1),␣
 ↪paste0('_',tissue2))))
}

merge_dataframes_sig <- function(tissue1, tissue2){
    fdr1 = ifelse(tissue1 != 'dlpfc', 0.05, 0.05)
    fdr2 = ifelse(tissue2 != 'dlpfc', 0.05, 0.05)
    return(merge(get_deg_sig(config[[tissue1]], fdr1),␣
 ↪get_deg_sig(config[[tissue2]], fdr2),
                 by='Feature', suffixes=c(paste0('_',tissue1),␣
 ↪paste0('_',tissue2))))
}

tissue_annotation <- function(tissue){
    return(list('dlpfc'='DLPFC', 'hippo'='Hippocampus',␣
 ↪'caudate'='Caudate')[[tissue]])
}

get_scatter_plot <- function(tissue1, tissue2, merge_fnc, coords){
    dft <- merge_fnc(tissue1, tissue2)
    sp = ggscatter(dft, x=paste0('t_', tissue1), y=paste0('t_', tissue2),␣
 ↪add="reg.line",
                   xlab=paste0('T-statistic (',tissue_annotation(tissue1), ')'),
                   ylab=paste0('T-statistic (',tissue_annotation(tissue2), ')'),
                   add.params=list(color="blue", fill="lightgray"), conf.
 ↪int=TRUE,
                   cor.method="pearson", cor.coef=FALSE, cor.coef.size=7,
                   cor.coeff.args=list(label.sep="\n"), ylim=c(-6,8)) +
        stat_cor(aes(label=..rr.label..), label.sep='\n', size=8,
                 method="spearman",  label.x=-8, label.y=7) +
```

```
        font("xylab", size=20, face='bold') +
        font("xy.text", size=18)
    return(sp)
}


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

[4]:
```
options(repr.plot.width=18, repr.plot.height=6)
sp1_sig = get_scatter_plot('caudate', 'dlpfc', merge_dataframes_sig, c(-110,␣
↪85))
sp2_sig = get_scatter_plot('caudate', 'hippo', merge_dataframes_sig, c(-110,␣
↪85))
sp3_sig = get_scatter_plot('dlpfc', 'hippo', merge_dataframes_sig, c(-110, 85))
fig1 = ggarrange(sp1_sig, sp2_sig, sp3_sig, ncol=3, nrow=1, align='v')
print(fig1)
```

```
`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'
```

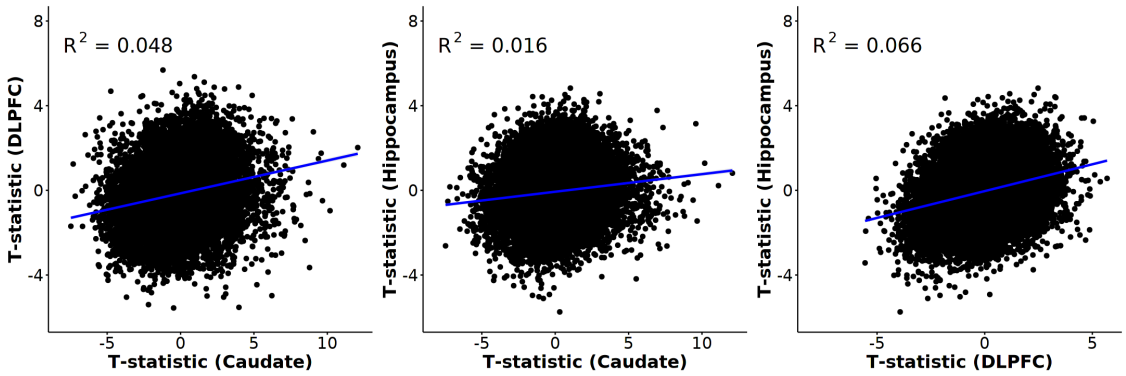


[5]:
```
save_ggplots("tstatistic_corr_sig", fig1, 18, 6)
```

[6]:
```
sp1 = get_scatter_plot('caudate', 'dlpfc', merge_dataframe, c(-110, 85))
sp2 = get_scatter_plot('caudate', 'hippo', merge_dataframe, c(-110, 85))
sp3 = get_scatter_plot('dlpfc', 'hippo', merge_dataframe, c(-110, 85))
fig2 = ggarrange(sp1, sp2, sp3, ncol=3, nrow=1, align='v')
```

```
print(fig2)
```

```
`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'
```



[7]:
```
save_ggplots("tstatistic_corr", fig2, 18, 6)
```

1.1 Reproducibility Information

[8]:
```
print("Reproducibility Information:")
Sys.time()
proc.time()
options(width=120)
sessioninfo::session_info()
```

```
[1] "Reproducibility Information:"

[1] "2021-07-09 15:09:52 EDT"

   user  system elapsed
 28.918   1.038  23.338

 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     2021-07-09

 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)
bit           4.0.4    2020-08-04 [1] CRAN (R 4.0.2)
bit64         4.0.5    2020-08-30 [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)
cowplot       1.1.1    2020-12-30 [1] CRAN (R 4.0.2)
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)
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)
lattice       0.20-41  2020-04-02 [2] CRAN (R 4.0.3)
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)
 Matrix        1.3-4    2021-06-01 [1] CRAN (R 4.0.3)
 mgcv          1.8-33   2020-08-27 [2] CRAN (R 4.0.3)
 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)
 nlme          3.1-152  2021-02-04 [1] CRAN (R 4.0.3)
 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
```