main

July 14, 2021

1 Extract differential expression analsys from Phase 2 Rdata

```
[1]: library(dplyr)
```

```
Attaching package: 'dplyr'


The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

1.1 Load data

```
[2]: lname = load('/ceph/users/jbenja13/phase3_paper/phase2/_m/deres.Rdata')
     lname
```

'deres'

```
[3]: names(deres)
```

1. 'development' 2. 'region' 3. 'sczd'

1.2 Extract features and select common columns

```
[4]: tx = deres$sczd$tx %>% as.data.frame %>% tibble::rownames_to_column("Feature")␣
     ↪%>%
         mutate(Symbol=gene_name, gencodeID=gene_id, gencodeTx=transcript_id)
     t_col = colnames(tx)

     jxn = deres$sczd$jxn %>% tibble::rownames_to_column("Feature") %>%
         mutate(gencodeID=newGeneID, Symbol=newGeneSymbol)
```

```r
j_col = colnames(jxn)

gene = deres$sczd$gene %>% as.data.frame %>% tibble::
  ↪rownames_to_column("Feature")
g_col = colnames(gene)

exon = deres$sczd$exon %>% as.data.frame %>% tibble::
  ↪rownames_to_column("Feature")
e_col = colnames(exon)

cols = intersect(intersect(intersect(j_col, e_col), g_col), t_col)
cols
```

Loading required package: S4Vectors

Loading required package: stats4

Loading required package: BiocGenerics

Loading required package: parallel

Attaching package: 'BiocGenerics'

The following objects are masked from 'package:parallel':

    clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
    clusterExport, clusterMap, parApply, parCapply, parLapply,
    parLapplyLB, parRapply, parSapply, parSapplyLB

The following objects are masked from 'package:dplyr':

    combine, intersect, setdiff, union

The following objects are masked from 'package:stats':

    IQR, mad, sd, var, xtabs

The following objects are masked from 'package:base':

    anyDuplicated, append, as.data.frame, basename, cbind, colnames,
    dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
    grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,

```
    order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
    rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
    union, unique, unsplit, which.max, which.min


Attaching package: 'S4Vectors'


The following objects are masked from 'package:dplyr':

    first, rename


The following object is masked from 'package:base':

    expand.grid
```

1. 'Feature' 2. 'ensemblID' 3. 'Symbol' 4. 'gencodeTx' 5. 'meanExprs' 6. 'passExprsCut' 7. 'logFC'
8. 'AveExpr' 9. 't' 10. 'P.Value' 11. 'adj.P.Val' 12. 'B' 13. 'region' 14. 'type' 15. 'gencodeID'

1.3   Merge data

```
[5]: df = bind_rows(gene %>% select(cols, -passExprsCut, -gencodeTx),
                  tx %>% select(cols, -passExprsCut, -gencodeTx),
                  exon %>% select(cols, -passExprsCut, -gencodeTx),
                  jxn %>% select(cols, -passExprsCut, -gencodeTx))
     df %>% dim
```

```
Note: Using an external vector in selections is ambiguous.
  Use `all_of(cols)` instead of `cols` to silence this message.
  See <https://tidyselect.r-lib.org/reference/faq-external-
vector.html>.
This message is displayed once per session.
```

1. 1622296 2. 13

1.4   Write output

```
[6]: df %>% filter(region == "DLPFC") %>%
         data.table::fwrite("dlpfc_diffExpr_szVctl_full.tsv", sep='\t')

     df %>% filter(region == "HIPPO") %>%
         data.table::fwrite("hippo_diffExpr_szVctl_full.tsv", sep='\t')
```

1.5 Reproducibility Information

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

```
[1] "2021-07-14 13:57:24 EDT"

   user  system elapsed
 40.930   1.476  38.171

 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     2021-07-14

 Packages
 package      * version date       lib source
 assertthat     0.2.1   2019-03-21 [1] CRAN (R 4.0.2)
 base64enc      0.1-3   2015-07-28 [1] CRAN (R 4.0.2)
 BiocGenerics * 0.36.1  2021-04-16 [1] Bioconductor
 cli            3.0.0   2021-06-30 [1] CRAN (R 4.0.3)
 crayon         1.4.1   2021-02-08 [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)
 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)
 generics       0.1.0   2020-10-31 [1] CRAN (R 4.0.2)
 glue           1.4.2   2020-08-27 [1] CRAN (R 4.0.2)
 htmltools      0.5.1.1 2021-01-22 [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)
 lifecycle      1.0.0   2021-02-15 [1] CRAN (R 4.0.3)
 magrittr       2.0.1   2020-11-17 [1] CRAN (R 4.0.2)
 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)
repr           1.1.3    2021-01-21 [1] CRAN (R 4.0.2)
rlang          0.4.11   2021-04-30 [1] CRAN (R 4.0.3)
S4Vectors    * 0.28.1   2020-12-09 [1] Bioconductor
sessioninfo    1.1.1    2018-11-05 [1] CRAN (R 4.0.2)
tibble         3.1.2    2021-05-16 [1] CRAN (R 4.0.3)
tidyselect     1.1.1    2021-04-30 [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)

[1] /home/jbenja13/R/x86_64-pc-linux-gnu-library/4.0
[2] /usr/lib/R/library
```