main

September 15, 2021

1 Gene Ontology Semantic Similarity Analysis

```
[1]: suppressPackageStartupMessages({
         library(dplyr)
         library(ggplot2)
         library(GOSemSim)
         library(org.Hs.eg.db)
     })
```

1.1 Functions

```
[2]: get_semData <- function(ont){
         return(godata('org.Hs.eg.db', ont=ont))
     }

     save_img <- function(image, fn, w=7, h=7){
         for(ext in c(".pdf", ".png")){
             ggsave(file=paste0(fn, ext), plot=image, width=w, height=h)
         }
     }

     GO_semantic_similarity <- function(ont){
         hsGO = get_semData(ont)
         tissues = c("Caudate", "DLPFC", "Hippocampus")
         t1 = c(); t2 = c(); ss = c()
         for(tissue1 in c("Caudate", "DLPFC")){
             fn1 = paste0("../../_m/", tolower(tissue1), "_functional_enrichment.
      ↪txt")
             for(tissue2 in c("DLPFC", "Hippocampus")){
                 fn2 = paste0("../../_m/", tolower(tissue2), "_functional_enrichment.
      ↪txt")
                 if(tissue1 != tissue2 & file.exists(fn1) & file.exists(fn2)){
                     df1 = data.table::fread(fn1) %>% filter(source == paste0("GO:",␣
      ↪ont))
                     df2 = data.table::fread(fn2) %>% filter(source == paste0("GO:",␣
      ↪ont))
                     sim = mgoSim(df1$term_id, df2$term_id, semData=hsGO,
```

```
                      measure="Wang", combine="BMA")
            t1 = c(t1, tissue1); t2 = c(t2, tissue2);ss = c(ss, sim)
        }
      }
    }
    return(data.frame("Tissue_1"=t1, "Tissue_2"=t2, "Semantic_Similarity"=ss,␣
 ↪"Ont"=ont))
}
```

1.2   Semantic similarity analysis

```
[3]: datalist = list()
     for(ont in c("MF", "BP", "CC")){
         datalist[[ont]] = GO_semantic_similarity(ont)
     }
     dt = bind_rows(datalist) %>% mutate_if(is.character, as.factor)
     dt %>% data.table::fwrite("go_semantic_similarity.tsv", sep='\t')
     dt %>% head
```

preparing gene to GO mapping data…

preparing IC data…

preparing gene to GO mapping data…

preparing IC data…

preparing gene to GO mapping data…

preparing IC data…

		Tissue_1	Tissue_2	Semantic_Similarity	Ont
		<fct>	<fct>	<dbl>	<fct>
A data.frame: 6 × 4	1	Caudate	DLPFC	0.971	MF
	2	Caudate	Hippocampus	0.946	MF
	3	DLPFC	Hippocampus	0.976	MF
	4	Caudate	DLPFC	0.919	BP
	5	Caudate	Hippocampus	0.873	BP
	6	DLPFC	Hippocampus	0.957	BP

1.3   Plot semantic similarity
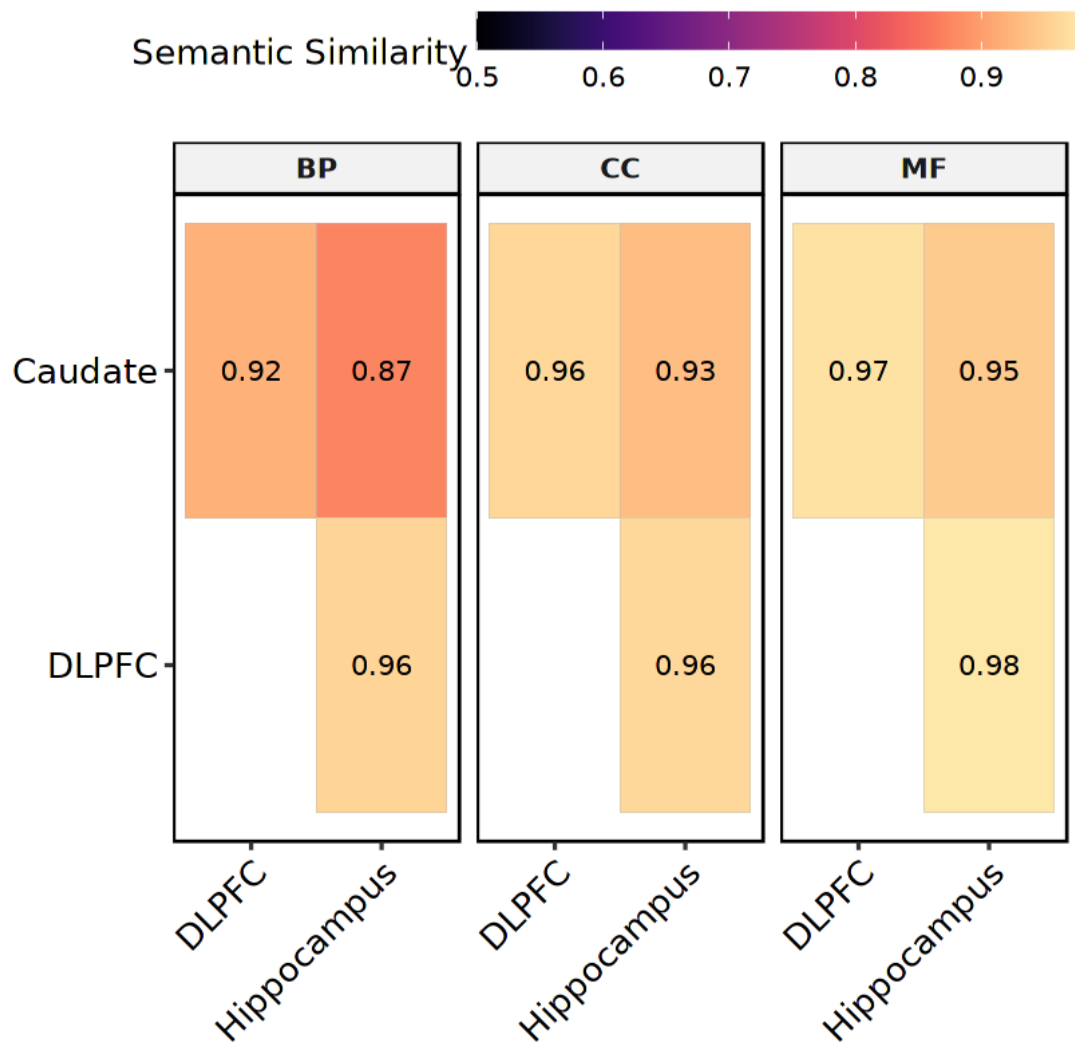
```
[4]: tile_plot <- dt %>% #tidyr::drop_na() %>%
         ggplot(aes(x=Tissue_2, y=Tissue_1, fill=Semantic_Similarity,
                     label=format(round(Semantic_Similarity, 2)))) +
         geom_tile(color="grey") + ggfittext::geom_fit_text(contrast=TRUE) +
         viridis::scale_color_viridis(option="magma") + facet_wrap("~Ont") +
```

```
        viridis::scale_fill_viridis(name="Semantic Similarity", limits=c(0.5,1),
                                     direction=1, option="magma") +
    labs(x="", y="") + ggpubr::theme_pubr(base_size=15, border=TRUE) +
    scale_y_discrete(limits = rev(levels(dt$Tissue_1))) +
    theme(axis.text.x=element_text(angle = 45, hjust=1),
          strip.text=element_text(face="bold"),
          legend.key.width=unit(2, 'cm'))
save_img(tile_plot, "GO_semantic_similarity", w=8, h=5)
tile_plot
```

1.4 Reproducibility Information

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

```
[1] "2021-09-15 09:18:10 EDT"

   user  system elapsed
 44.051   0.730  45.135

 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     2021-09-15

 Packages
 package       * version  date       lib source
 abind           1.4-5    2016-07-21 [1] CRAN (R 4.0.2)
 AnnotationDbi * 1.52.0   2020-10-27 [1] Bioconductor
 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)
 Biobase       * 2.50.0   2020-10-27 [1] Bioconductor
 BiocGenerics  * 0.36.1   2021-04-16 [1] Bioconductor
 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)
 blob            1.2.2    2021-07-23 [1] CRAN (R 4.0.3)
 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)
 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)
generics         0.1.0    2020-10-31 [1] CRAN (R 4.0.2)
ggfittext        0.9.1    2021-01-30 [1] CRAN (R 4.0.3)
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)
GO.db            3.12.1   2021-04-08 [1] Bioconductor
GOSemSim       * 2.16.1   2020-10-29 [1] Bioconductor
gridExtra        2.3      2017-09-09 [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)
IRanges        * 2.24.1   2020-12-12 [1] Bioconductor
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)
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)
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)
org.Hs.eg.db   * 3.12.0   2021-02-24 [1] Bioconductor
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)
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)
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)
RSQLite          2.2.8    2021-08-21 [1] CRAN (R 4.0.3)
rstatix          0.7.0    2021-02-13 [1] CRAN (R 4.0.3)
S4Vectors      * 0.28.1   2020-12-09 [1] Bioconductor
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)
shades           1.4.0    2019-08-02 [1] CRAN (R 4.0.3)
```

```
stringi         1.7.4    2021-08-25 [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)
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)
viridis         0.6.1    2021-05-11 [1] CRAN (R 4.0.3)
viridisLite     0.4.0    2021-04-13 [1] CRAN (R 4.0.3)
withr           2.4.2    2021-04-18 [1] CRAN (R 4.0.3)
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
```

[ ]: