main

August 21, 2021

1 Visualize GO analysis

```
[1]: import numpy as np
     import pandas as pd
[2]: def get_top_GO(tissue, fn, label):
         df = pd.read_excel(fn).sort_values('p_uncorrected').head(10)
         df['Log10'] = -np.log10(df['p_fdr_bh'])
         df['Tissue'] = tissue
         df['Bias'] = label
         return df
[3]: tissue = 'hippocampus'
     config = {
         'All': '../../_m/GO_analysis_allDEG.xlsx',
         'AA': '../../_m/GO_analysis_downregulated.xlsx',
         'EA': '../../_m/GO_analysis_upregulated.xlsx',
     }
     df = pd.DataFrame()
     for bias in ['All', 'AA', 'EA']:
         df = pd.concat([df, get_top_GO(tissue, config[bias], bias)], axis=0)
     fac = []
     for ii in range(df.shape[0]):
         xx, yy = df[['ratio_in_study']].iloc[ii, 0].split('/')
         fac.append((int(xx) / int(yy)) * 2)
     df['geneRatio'] = fac
     print(np.min(fac), np.max(fac))
    0.0 0.17540842648323302
[4]: df.to_csv("%s_GO_analysis.tsv" % tissue, sep='\t', index=False)
```

1.1 Plot

```
[5]: %load ext rpy2.ipython
[6]: \%\R -i df
     library(ggplot2)
     library(tidyverse)
     save_plot <- function(p, fn, w, h){</pre>
         for(ext in c('.svg', '.png', '.pdf')){
             ggsave(file=paste0(fn,ext), plot=p, width=w, height=h)
         }
     }
     plot_GO <- function(){</pre>
         cbPalette <- c("#000000", "Red", "Blue")
         gg1 = df \%
             ggplot(aes(x=Log10, y=name, color=Bias, size=geneRatio)) +
             geom_point(shape=18, alpha=0.8) + labs(y='', x='-Log10 (FDR)') +
             theme_bw(base_size=15) +
             scale_colour_manual(name="Ancestry", values=cbPalette,
                                 labels=c("All", "AA Bias", "EA Bias")) +
             geom_vline(xintercept = -log10(0.05), linetype = "dotted") +
             theme(axis.title=element_text(face='bold'),
                   strip.text=element_text(face='bold'))
         return(gg1)
     }
    R[write to console]:
                            Attaching packages
                          tidyverse 1.3.1
    R[write to console]: tibble 3.1.2
                                               dplyr 1.0.7
     tidyr
             1.1.3
                          stringr 1.4.0
                          forcats 0.5.1
     readr
              1.4.0
              0.3.4
     purrr
    R[write to console]:
                           Conflicts
    tidyverse_conflicts()
     dplyr::filter() masks stats::filter()
     dplyr::lag()
                      masks stats::lag()
[7]: \%\%R
     gg1 = plot_GO()
     print(gg1)
     save_plot(gg1, "ancestry_G0_top10_stacked", 10, 6)
```

- sphingolipid biosynthetic process
 - protein transport -
- roteasome-mediated ubiquitin-dependent protein catabolic process
 - positive regulation of superoxide anion generation -
 - positive regulation of NF-kappaB transcription factor activity
 - positive regulation of interleukin-6 production -
 - positive regulation of interleukin-1 beta production geneR
 - positive regulation of I-kappaB kinase/NF-kappaB signaling 0.0
 - phagocytosis, engulfment • 0.0
 - olfactory receptor activity • 0.1
 - mitochondrion • 0.1
 - mitochondrial inner membrane
 - lysosomal membrane -
 - late endosome membrane Ances
 - innate immune response • All
 - inflammatory response AA
 - gene silencing by miRNA • EA
 - external side of plasma membrane -
 - endoplasmic reticulum membrane
 - endoplasmic reticulum -
- ection of chemical stimulus involved in sensory perception of smell
 - cilium assembly
 - axoneme -

3.8

-Log10 (FDR)

[]: