

main

July 9, 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_csv(fn, sep='\t').sort_values('p_value').head(10)
    df['Log10'] = -np.log10(df['p_value'])
    df['Tissue'] = tissue
    df['Bias'] = label
    return df
```

```
[3]: tissue = 'dlpfc'
config = {
    'All': '../_m/allDEGs_functional_enrichment.txt',
    'Female': '../_m/female_bias_DEGs_functional_enrichment.txt',
    'Male': '../_m/male_bias_DEGs_functional_enrichment.txt',
}

df = pd.DataFrame()
for bias in ['All', 'Female', 'Male']:
    df = pd.concat([df, get_top_GO(tissue, config[bias], bias)], axis=0)
```

```
[4]: df.to_csv("%s_functional_analysis.txt" % tissue, sep='\t', index=False)
```

```
[5]: df.shape
```

```
[5]: (28, 17)
```

### 1.1 Plot

```
[6]: %load_ext rpy2.ipython
```

```
[7]: %%R -i df
library(ggplot2)
library(tidyverse)

save_plot <- function(p, fn, w, h){
```

```

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

plot_G0 <- function(){
  cbPalette <- c("#000000", "Red", "Blue")
  gg1 = df %>%
    ggplot(aes(x=Log10, y=term_name, color=Bias)) +
    geom_point(shape=18, alpha=0.8, size=4) + labs(y='', x=''-Log10 (p
    ↪adjust)') +
    theme_bw() +
    scale_colour_manual(name="Sex Bias", values=cbPalette,
                        labels=c("All", "Female Bias", "Male Bias")) +
    geom_vline(xintercept = -log10(0.05), linetype = "dotted") +
    theme(axis.text=element_text(size=14),
          axis.title=element_text(size=18, face='bold'),
          strip.text=element_text(size=18, face='bold'))
  return(gg1)
}

```

```

R[write to console]:  Attaching packages
                      tidyverse 1.3.1

```

```

R[write to console]:  tidyr    1.1.3    stringr 1.4.0
                      readr    1.4.0    forcats 0.5.1
                      purrr    0.3.4

```

```

R[write to console]:  Conflicts
tidyverse_conflicts()
dplyr::filter() masks stats::filter()
dplyr::lag()    masks stats::lag()

```

```

[8]: %>%R
      gg1 = plot_G0()
      print(gg1)
      save_plot(gg1, "dlpfc_G0_top10_stacked", 12, 6)

```

X-linked inheritance-  
 Terpenoid backbone biosynthesis-  
 pole plasm-  
 P granule-  
 Mevalonate pathway-  
 Intraflagellar transport-  
 Gonosomal inheritance-  
 germ plasm-  
 focal adhesion-  
 Factor: ZF5; motif: NRNGNGCGCGCWN; match class: 1-  
 Factor: ZF5; motif: GSGCGCGR-  
 Factor: ZF5; motif: GGSGCGCGS; match class: 1-  
 Factor: MAZ; motif: GGGMGGGGSSGGGGGGGGGGGG-  
 Factor: ETF; motif: GVGGMGG-  
 Factor: E2F-4; motif: SNGGGCGGGAANN; match class: 1-  
 2F-3:HES-7; motif: NNNSGCGCSNNNNNCRCGYGNN; match class: 1-  
 Factor: E2F-2; motif: GCGCGCGCNCS; match class: 1-  
 Factor: E2F-2; motif: GCGCGCGCNCS-  
 Factor: E2F-2; motif: GCGCGCGCGYW-  
 2F-1:HES-7; motif: GGCRCGTGSYNNWNNGGCGCSM; match class: 1-  
 Cholesterol Biosynthesis Pathway-  
 Cholesterol biosynthesis-  
 cell-substrate junction-  
 anchoring junction-  
 -Log10 (p

[ ]: