## 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_excel(fn).sort_values('p_uncorrected').head(15)
         df['Log10'] = -np.log10(df['p_fdr_bh'])
         df['Tissue'] = tissue
         df['Bias'] = label
         return df
[3]: tissue = 'caudate'
     config = {
         'All': '../../_m/GO_analysis_allDEG.xlsx',
         'CTL': '../../_m/GO_analysis_downregulated.xlsx',
         'SZ': '../../_m/GO_analysis_upregulated.xlsx',
     }
     df = pd.DataFrame()
     for bias in ['CTL', 'SZ']:
         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 0.07553648068669527
[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("Blue", "Red")
         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() +
             scale_colour_manual(name="DEGs Enrichment", values=cbPalette,
                                  labels=c("Downregulated in SZ", "Upregulated in _{\mbox{\scriptsize LL}}
      →SZ")) +
             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]: tibble 3.1.2
                                                        1.0.7
                                                dplyr
                          stringr 1.4.0
     tidyr
             1.1.3
                          forcats 0.5.1
      readr
              1.4.0
      purrr
              0.3.4
    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, "GO_top15_stacked", 12, 8)
```

- voltage-gated potassium channel complex
  - voltage-gated potassium channel activity
    - synaptic vesicle membrane
      - synapse assembly
        - startle response-
      - small GTPase binding-
    - semaphorin-plexin signaling pathway
      - regulation of dopamine secretion
        - potassium ion transport-
    - potassium ion transmembrane transport
      - postsynaptic density-
- positive regulation of protein phosphorylation
  - phosphatase binding · ч
  - olfactory receptor activity
    - neuronal cell body-
  - nervous system development-
  - negative regulation of neuron differentiation
    - myelin sheath ◆ 0.
    - microtubule cytoskeleton- 0.
      - microtubule- 📥 0.
- ilic cell adhesion via plasma membrane adhesion molecules
  - guanyl-nucleotide exchange factor activity-
    - GTPase activator activity
      - glutamatergic synapse-
    - gene silencing by miRNA-
- of chemical stimulus involved in sensory perception of smell
  - cytoskeleton-
  - cytoplasmic vesicle-
  - chemical synaptic transmission
    - axon-
      - 3.6

## -Log10 (FDI

[]: