main

July 11, 2021

[1]:
```
import functools
import numpy as np
import pandas as pd
from scipy.stats import fisher_exact, binom_test
```

[2]:
```
@functools.lru_cache()
def get_deg(tissue):
    dft = pd.read_csv(config[tissue], sep='\t', index_col=0)
    dft['Feature'] = dft.index
    dft['Dir'] = np.sign(dft['t'])
    if 'gene_id' in dft.columns:
        dft['ensemblID'] = dft.gene_id.str.replace('\\..*', '', regex=True)
    elif 'ensembl_gene_id' in dft.columns:
        dft.rename(columns={'ensembl_gene_id': 'ensemblID'}, inplace=True)
    return dft[['Feature', 'ensemblID', 'adj.P.Val', 'logFC', 't', 'Dir']]


@functools.lru_cache()
def get_deg_sig(tissue, fdr):
    dft = get_deg(tissue)
    return dft[(dft['adj.P.Val'] < fdr)]


@functools.lru_cache()
def get_gtex():
    gtex_file = '/ceph/users/jbenja13/projects/sex_sz_ria/input/'+\
                'public_results/extract_geneLists/_m/gp_gtex_sex_diff_genes.csv'
    gtex = pd.read_csv(gtex_file).rename(columns={'Ensembl': 'Gencode'}).
 ↪drop('Gene', axis=1)
    gtex['Ensembl'] = gtex.Gencode.str.replace("\\..*", "", regex=True)
    return gtex.set_index('Ensembl')
```

[3]:
```
def cal_fishers(bs_tissue, tissue_col):
    df = get_deg(bs_tissue)
    gtex = get_gtex()
    fdr = 0.05
    table = [[len(set(df[(df['adj.P.Val']<fdr)].ensemblID) &
                  set(gtex[(gtex[tissue_col] != 0)].index)),
```

```
                len(set(df[(df['adj.P.Val']<fdr)].ensemblID) &
                    set(gtex[(gtex[tissue_col] == 0)].index))],
               [len(set(df[(df['adj.P.Val']>=fdr)].ensemblID) &
                    set(gtex[(gtex[tissue_col] != 0)].index)),
                len(set(df[(df['adj.P.Val']>=fdr)].ensemblID) &
                    set(gtex[(gtex[tissue_col] == 0)].index))]]
    print(table)
    return fisher_exact(table)
```

[4]:
```
config = {
    'caudate': '../../../../caudate/_m/genes/diffExpr_maleVfemale_full.txt',
    'dlpfc': '../../../../dlpfc/_m/genes/diffExpr_maleVfemale_full.txt',
    'hippo': '../../../../hippocampus/_m/genes/diffExpr_maleVfemale_full.txt',
}
```

0.1 GTEx directionality

[5]:
```
brain_col = get_gtex().columns[get_gtex().columns.str.contains("Brain")]
brain_col
```

[5]:
```
Index(['Brain-Amygdala', 'Brain-Anterior_cingulate_cortex', 'Brain-Caudate',
       'Brain-Cerebellar_Hemisphere', 'Brain-Cerebellum', 'Brain-Cortex',
       'Brain-Frontal_Cortex', 'Brain-Hippocampus', 'Brain-Hypothalamus',
       'Brain-Nucleus_accumbens', 'Brain-Putamen', 'Brain-Spinal_cord',
       'Brain-Substantia_nigra'],
      dtype='object')
```

0.1.1 Fisher Exact Test, Enrichment of Overlap

[6]:
```
for bs_tissue in ['caudate', 'dlpfc', 'hippo']:
    for tissue_col in brain_col:
        print("Enrichment for %s from BrainSeq and %s from GTEx!" %
              (bs_tissue, tissue_col))
        print(cal_fishers('caudate', tissue_col))
    print("")
```

```
Enrichment for caudate from BrainSeq and Brain-Amygdala from GTEx!
[[10, 243], [2, 15019]]
(309.0329218106996, 8.348481397310317e-17)
Enrichment for caudate from BrainSeq and Brain-Anterior_cingulate_cortex from
GTEx!
[[14, 239], [124, 14897]]
(7.037319476312593, 6.93516674357797e-08)
Enrichment for caudate from BrainSeq and Brain-Caudate from GTEx!
[[13, 240], [0, 15021]]
(inf, 5.190924068649964e-24)
Enrichment for caudate from BrainSeq and Brain-Cerebellar_Hemisphere from GTEx!
[[10, 243], [2, 15019]]
```

```
(309.0329218106996, 8.348481397310317e-17)
Enrichment for caudate from BrainSeq and Brain-Cerebellum from GTEx!
[[15, 238], [26, 14995]]
(36.34857789269554, 5.529547706631774e-17)
Enrichment for caudate from BrainSeq and Brain-Cortex from GTEx!
[[10, 243], [0, 15021]]
(inf, 1.3023439738464406e-18)
Enrichment for caudate from BrainSeq and Brain-Frontal_Cortex from GTEx!
[[11, 242], [2, 15019]]
(341.34090909090907, 1.5705137252605465e-18)
Enrichment for caudate from BrainSeq and Brain-Hippocampus from GTEx!
[[11, 242], [0, 15021]]
(inf, 2.0733070337637218e-20)
Enrichment for caudate from BrainSeq and Brain-Hypothalamus from GTEx!
[[12, 241], [1, 15020]]
(747.8838174273859, 4.2111963833078614e-21)
Enrichment for caudate from BrainSeq and Brain-Nucleus_accumbens from GTEx!
[[12, 241], [0, 15021]]
(inf, 3.2872980552944607e-22)
Enrichment for caudate from BrainSeq and Brain-Putamen from GTEx!
[[12, 241], [0, 15021]]
(inf, 3.2872980552944607e-22)
Enrichment for caudate from BrainSeq and Brain-Spinal_cord from GTEx!
[[10, 243], [2, 15019]]
(309.0329218106996, 8.348481397310317e-17)
Enrichment for caudate from BrainSeq and Brain-Substantia_nigra from GTEx!
[[12, 241], [0, 15021]]
(inf, 3.2872980552944607e-22)

Enrichment for dlpfc from BrainSeq and Brain-Amygdala from GTEx!
[[10, 243], [2, 15019]]
(309.0329218106996, 8.348481397310317e-17)
Enrichment for dlpfc from BrainSeq and Brain-Anterior_cingulate_cortex from
GTEx!
[[14, 239], [124, 14897]]
(7.037319476312593, 6.93516674357797e-08)
Enrichment for dlpfc from BrainSeq and Brain-Caudate from GTEx!
[[13, 240], [0, 15021]]
(inf, 5.190924068649964e-24)
Enrichment for dlpfc from BrainSeq and Brain-Cerebellar_Hemisphere from GTEx!
[[10, 243], [2, 15019]]
(309.0329218106996, 8.348481397310317e-17)
Enrichment for dlpfc from BrainSeq and Brain-Cerebellum from GTEx!
[[15, 238], [26, 14995]]
(36.34857789269554, 5.529547706631774e-17)
Enrichment for dlpfc from BrainSeq and Brain-Cortex from GTEx!
[[10, 243], [0, 15021]]
(inf, 1.3023439738464406e-18)
```

```
Enrichment for dlpfc from BrainSeq and Brain-Frontal_Cortex from GTEx!
[[11, 242], [2, 15019]]
(341.34090909090907, 1.5705137252605465e-18)
Enrichment for dlpfc from BrainSeq and Brain-Hippocampus from GTEx!
[[11, 242], [0, 15021]]
(inf, 2.0733070337637218e-20)
Enrichment for dlpfc from BrainSeq and Brain-Hypothalamus from GTEx!
[[12, 241], [1, 15020]]
(747.8838174273859, 4.2111963833078614e-21)
Enrichment for dlpfc from BrainSeq and Brain-Nucleus_accumbens from GTEx!
[[12, 241], [0, 15021]]
(inf, 3.2872980552944607e-22)
Enrichment for dlpfc from BrainSeq and Brain-Putamen from GTEx!
[[12, 241], [0, 15021]]
(inf, 3.2872980552944607e-22)
Enrichment for dlpfc from BrainSeq and Brain-Spinal_cord from GTEx!
[[10, 243], [2, 15019]]
(309.0329218106996, 8.348481397310317e-17)
Enrichment for dlpfc from BrainSeq and Brain-Substantia_nigra from GTEx!
[[12, 241], [0, 15021]]
(inf, 3.2872980552944607e-22)

Enrichment for hippo from BrainSeq and Brain-Amygdala from GTEx!
[[10, 243], [2, 15019]]
(309.0329218106996, 8.348481397310317e-17)
Enrichment for hippo from BrainSeq and Brain-Anterior_cingulate_cortex from
GTEx!
[[14, 239], [124, 14897]]
(7.037319476312593, 6.93516674357797e-08)
Enrichment for hippo from BrainSeq and Brain-Caudate from GTEx!
[[13, 240], [0, 15021]]
(inf, 5.190924068649964e-24)
Enrichment for hippo from BrainSeq and Brain-Cerebellar_Hemisphere from GTEx!
[[10, 243], [2, 15019]]
(309.0329218106996, 8.348481397310317e-17)
Enrichment for hippo from BrainSeq and Brain-Cerebellum from GTEx!
[[15, 238], [26, 14995]]
(36.34857789269554, 5.529547706631774e-17)
Enrichment for hippo from BrainSeq and Brain-Cortex from GTEx!
[[10, 243], [0, 15021]]
(inf, 1.3023439738464406e-18)
Enrichment for hippo from BrainSeq and Brain-Frontal_Cortex from GTEx!
[[11, 242], [2, 15019]]
(341.34090909090907, 1.5705137252605465e-18)
Enrichment for hippo from BrainSeq and Brain-Hippocampus from GTEx!
[[11, 242], [0, 15021]]
(inf, 2.0733070337637218e-20)
Enrichment for hippo from BrainSeq and Brain-Hypothalamus from GTEx!
```

```
[[12, 241], [1, 15020]]
(747.8838174273859, 4.2111963833078614e-21)
Enrichment for hippo from BrainSeq and Brain-Nucleus_accumbens from GTEx!
[[12, 241], [0, 15021]]
(inf, 3.2872980552944607e-22)
Enrichment for hippo from BrainSeq and Brain-Putamen from GTEx!
[[12, 241], [0, 15021]]
(inf, 3.2872980552944607e-22)
Enrichment for hippo from BrainSeq and Brain-Spinal_cord from GTEx!
[[10, 243], [2, 15019]]
(309.0329218106996, 8.348481397310317e-17)
Enrichment for hippo from BrainSeq and Brain-Substantia_nigra from GTEx!
[[12, 241], [0, 15021]]
(inf, 3.2872980552944607e-22)
```

0.1.2 Binomial enrichment for directionality

[7]:
```
@functools.lru_cache()
def get_gtex_col(tissue_col):
    dx = get_gtex().loc[:, ["Symbol", tissue_col]].copy()
    dx['Dir'] = -1*np.sign(dx[tissue_col]) #fix correlations (male upregulated␣
 ↪is negative here)
    return dx


@functools.lru_cache()
def get_gtex_col_sig(tissue_col):
    dx = get_gtex_col(tissue_col)
    return dx[(dx[tissue_col] != 0)]


@functools.lru_cache()
def merge_dataframes_sig(tissue, tissue_col):
    fdr = 0.05
    return get_deg_sig(tissue, fdr).merge(get_gtex_col_sig(tissue_col),
                                          right_index=True, left_on='ensemblID',
                                          suffixes=['_%s' % tissue, '_%s' %␣
 ↪tissue_col])
```

[8]:
```
def enrichment_binom(tissue1, tissue2, merge_fnc):
    df = merge_fnc(tissue1, tissue2)
    df['agree'] = df['Dir_%s' % tissue1] * df['Dir_%s' % tissue2]
    dft = df.groupby('agree').size().reset_index()
    print(dft)
    return binom_test(dft[0].iloc[1], dft[0].sum()) if dft.shape[0] != 1 else␣
 ↪print("All directions agree!")
```

[9]:
```
for tissue in ['caudate', 'dlpfc', 'hippo']:
    for col in brain_col:
        print("Binomial enrichment for %s from BrainSeq and %s from GTEx!" %
              (tissue, col))
        print(enrichment_binom(tissue, col, merge_dataframes_sig))
    print("\n")
```

```
Binomial enrichment for caudate from BrainSeq and Brain-Amygdala from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-
Anterior_cingulate_cortex from GTEx!
   agree   0
0    1.0  14
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Caudate from GTEx!
   agree   0
0    1.0  13
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Cerebellar_Hemisphere
from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Cerebellum from GTEx!
   agree   0
0    1.0  15
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Cortex from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Frontal_Cortex from
GTEx!
   agree   0
0    1.0  11
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Hippocampus from GTEx!
   agree   0
0    1.0  11
```

```
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Hypothalamus from GTEx!
   agree   0
0    1.0  12
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Nucleus_accumbens from
GTEx!
   agree   0
0    1.0  12
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Putamen from GTEx!
   agree   0
0    1.0  12
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Spinal_cord from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for caudate from BrainSeq and Brain-Substantia_nigra from
GTEx!
   agree   0
0    1.0  12
All directions agree!
None


Binomial enrichment for dlpfc from BrainSeq and Brain-Amygdala from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Anterior_cingulate_cortex
from GTEx!
   agree   0
0    1.0  15
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Caudate from GTEx!
   agree   0
0    1.0  12
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Cerebellar_Hemisphere from
```

```
GTEx!
   agree   0
0    1.0  11
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Cerebellum from GTEx!
   agree   0
0    1.0  15
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Cortex from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Frontal_Cortex from GTEx!
   agree   0
0    1.0  12
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Hippocampus from GTEx!
   agree   0
0    1.0  11
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Hypothalamus from GTEx!
   agree   0
0    1.0  11
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Nucleus_accumbens from
GTEx!
   agree   0
0    1.0  11
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Putamen from GTEx!
   agree   0
0    1.0  11
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Spinal_cord from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for dlpfc from BrainSeq and Brain-Substantia_nigra from
GTEx!
```

```
   agree   0
0    1.0  12
All directions agree!
None


Binomial enrichment for hippo from BrainSeq and Brain-Amygdala from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Anterior_cingulate_cortex
from GTEx!
   agree   0
0    1.0  14
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Caudate from GTEx!
   agree   0
0    1.0  13
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Cerebellar_Hemisphere from
GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Cerebellum from GTEx!
   agree   0
0    1.0  14
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Cortex from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Frontal_Cortex from GTEx!
   agree   0
0    1.0  11
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Hippocampus from GTEx!
   agree   0
0    1.0  11
All directions agree!
None
```

```
Binomial enrichment for hippo from BrainSeq and Brain-Hypothalamus from GTEx!
   agree   0
0    1.0  12
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Nucleus_accumbens from
GTEx!
   agree   0
0    1.0  12
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Putamen from GTEx!
   agree   0
0    1.0  12
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Spinal_cord from GTEx!
   agree   0
0    1.0  10
All directions agree!
None
Binomial enrichment for hippo from BrainSeq and Brain-Substantia_nigra from
GTEx!
   agree   0
0    1.0  12
All directions agree!
None
```

[ ]: