

main

July 13, 2021

1 Cell type enrichment analysis

```
[1]: import functools
import numpy as np
import pandas as pd
from scipy.stats import fisher_exact
from statsmodels.stats.multitest import multipletests
```

```
[2]: config = {
    'shared_file': '../../../../../differential_expression/tissue_comparison/
↳upset_plots/_m/BrainSeq_shared_degs_annotation.txt',
    'xci_file': '../_h/xci_status_hg19.txt',
    'background': "background.txt"
}
```

```
[3]: @functools.lru_cache()
def get_deg():
    return pd.read_csv(config['shared_file'], sep='\t')

@functools.lru_cache()
def get_xci():
    return pd.read_csv(config['xci_file'], sep='\t')

@functools.lru_cache()
def get_background():
    return pd.read_csv(config['background'], sep='\t')
```

```
[4]: def cal_fishers(status):
    xci = get_xci()
    xci['ensemblID'] = xci['Gene ID'].str.replace("\\\\.*", "", regex=True)
    df = get_background().drop_duplicates()\
        .merge(get_deg(), on="ensemblID", how="left")\
        .drop(['gene_id', 'gene_name'], axis=1)\
        .merge(xci, on='ensemblID', how="left")
    df['Dir'] = df.Dir.fillna(0)
```

```

table = [[np.sum((df['Dir']!=0) & (df['Combined XCI status'] == status)),
          np.sum((df['Dir']!=0) & (df['Combined XCI status'] != status))],
          [np.sum((df['Dir']==0) & (df['Combined XCI status'] == status)),
          np.sum((df['Dir']==0) & (df['Combined XCI status'] != status))]]
print(table)
return fisher_exact(table)

def cal_fishers_direction(status, direction):
    deg = get_deg()
    if direction == 'Up':
        deg = deg[(deg['Dir'] > 0)].copy()
    else:
        deg = deg[(deg['Dir'] < 0)].copy()
    xci = get_xci()
    xci['ensemblID'] = xci['Gene ID'].str.replace("\\\\..*", "", regex=True)
    df = get_background().drop_duplicates()\
        .merge(deg, on="ensemblID", how="left")\
        .drop(['gene_id', 'gene_name'], axis=1)\
        .merge(xci, on='ensemblID', how="left")
    df['Dir'] = df.Dir.fillna(0)
    table = [[np.sum((df['Dir']!=0) & (df['Combined XCI status'] == status)),
              np.sum((df['Dir']!=0) & (df['Combined XCI status'] != status))],
              [np.sum((df['Dir']==0) & (df['Combined XCI status'] == status)),
              np.sum((df['Dir']==0) & (df['Combined XCI status'] != status))]]
    print(table)
    return fisher_exact(table)

def cal_fisher_by_xci_status():
    xci_status = xci = get_xci().loc[:, 'Combined XCI status'].unique()
    xci_lt = []; pval_lt = []; oddratio_lt = []; dir_lt = [];
    for status in xci_status:
        odd_ratio, pval = cal_fishers(status)
        xci_lt.append(status); pval_lt.append(pval);
        oddratio_lt.append(odd_ratio); dir_lt.append('All')
        if pval < 0.05:
            print("There is a significant enrichment (p-value < %.1e) of %s!" %
                  (pval, status))

    for status in xci_status:
        odd_ratio, pval = cal_fishers_direction(status, "Up")
        xci_lt.append(status); pval_lt.append(pval);
        oddratio_lt.append(odd_ratio); dir_lt.append('Male Bias')
        if pval < 0.05:
            print("There is a significant enrichment of male bias genes_
→(p-value < %.1e) of %s!" %

```

```

        (pval, status))

    for status in xci_status:
        odd_ratio, pval = cal_fishers_direction(status, "Down")
        xci_lt.append(status); pval_lt.append(pval);
        oddratio_lt.append(odd_ratio); dir_lt.append('Female Bias')
        if pval < 0.05:
            print("There is a significant enrichment of female bias genes_
→(p-value < %.1e) of %s!" %
                (pval, status))
    return pd.DataFrame({'XCI status': xci_lt, 'OR': oddratio_lt,
                        'PValue': pval_lt, 'Direction': dir_lt})

```

1.1 Calculated fisher's exact for each XCI status

```

[5]: df = cal_fisher_by_xci_status()
_, fdr, _, _ = multipletests(df.PValue, method='fdr_bh')
df['FDR'] = fdr
df[(df.FDR <= 0.05)]

```

```
[[21, 52], [52, 24762]]
```

There is a significant enrichment (p-value < 2.6e-37) of escape!

```
[[1, 72], [73, 24741]]
```

```
[[0, 73], [391, 24423]]
```

```
[[2, 37], [71, 24777]]
```

There is a significant enrichment of male bias genes (p-value < 5.9e-03) of escape!

```
[[0, 39], [74, 24774]]
```

```
[[0, 39], [391, 24457]]
```

```
[[19, 15], [54, 24799]]
```

There is a significant enrichment of female bias genes (p-value < 1.1e-40) of escape!

```
[[1, 33], [73, 24780]]
```

```
[[0, 34], [391, 24462]]
```

```

[5]:   XCI status      OR      PValue  Direction      FDR
0    escape  192.308432  2.642862e-37        All  1.189288e-36
3    escape   18.863342  5.862299e-03   Male Bias  1.758690e-02
6    escape  581.704938  1.051047e-40  Female Bias  9.459419e-40

```

```
[6]: df
```

```

[6]:   XCI status      OR      PValue  Direction      FDR
0    escape  192.308432  2.642862e-37        All  1.189288e-36
1  variable    4.707192  1.956333e-01        All  3.521399e-01
2  inactive    0.000000  6.328014e-01        All  9.492021e-01
3    escape   18.863342  5.862299e-03   Male Bias  1.758690e-02
4  variable    0.000000  1.000000e+00   Male Bias  1.000000e+00

```

5	inactive	0.000000	1.000000e+00	Male Bias	1.000000e+00
6	escape	581.704938	1.051047e-40	Female Bias	9.459419e-40
7	variable	10.286426	9.635152e-02	Female Bias	2.167909e-01
8	inactive	0.000000	1.000000e+00	Female Bias	1.000000e+00

```
[7]: df.to_csv('xci_enrichment_analysis.txt', sep='\t', index=False)
```