

main

July 12, 2021

1 Examine genes that are disconcordant across brain regions

```
[1]: import functools
import numpy as np
import pandas as pd

[2]: config = {
    'caudate': '../.../caudate/_m/genes/diffExpr_EAvsAA_full.txt',
    'dlpfc': '../.../dlpfc/_m/genes/diffExpr_EAvsAA_full.txt',
    'hippo': '../.../hippocampus/_m/genes/diffExpr_EAvsAA_full.txt',
    'gyrus': '../.../dentateGyrus/_m/genes/diffExpr_EAvsAA_full.txt',
}

[3]: @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', 'Symbol', '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 merge_dataframes(tissue1, tissue2):
    return get_deg(tissue1).merge(get_deg(tissue2), on='Feature',
                                   suffixes=['_%s' % tissue1, '_%s' % tissue2])
```

```

@functools.lru_cache()
def merge_dataframes_sig(tissue1, tissue2):
    fdr1 = 0.05 if tissue1 != 'dlpfc' else 0.05
    fdr2 = 0.05 if tissue2 != 'dlpfc' else 0.05
    return get_deg_sig(tissue1, fdr1).merge(get_deg_sig(tissue2, fdr2),
on='Feature',
                                                    suffixes=['_s' % tissue1, '_s' %
tissue2])

```

```

[4]: def extract_disconcordant(tissue1, tissue2):
    df = merge_dataframes_sig(tissue1, tissue2)
    df = df[((df['Dir_%s' % tissue1] == 1) & (df['Dir_%s' % tissue2] == -1)) |
            ((df['Dir_%s' % tissue1] == -1) & (df['Dir_%s' % tissue2] == 1))]
    return df.loc[:, ['Feature', 'Dir_%s' % tissue1, 'Dir_%s' % tissue2]]\
        .merge(get_deg(tissue1), on='Feature').drop('Dir', axis=1)

```

1.1 BrainSeq Tissue Comparison

```

[5]: cd = extract_disconcordant('caudate', 'dlpfc')
cd.to_csv("disconcordant_genes_%s_%s.csv" % ('caudate', 'dlpfc'), index=False)
print(cd.shape)
cd

```

(8, 8)

```

[5]:

```

	Feature	Dir_caudate	Dir_dlpfc	ensemblID	Symbol	\
0	ENSG000000277883.1	1.0	-1.0	ENSG000000277883	NLRP3P1	
1	ENSG000000070601.9	1.0	-1.0	ENSG000000070601	FRMPD1	
2	ENSG000000006625.17	-1.0	1.0	ENSG000000006625	GGCT	
3	ENSG000000100302.6	1.0	-1.0	ENSG000000100302	RASD2	
4	ENSG000000198034.10	-1.0	1.0	ENSG000000198034	RPS4X	
5	ENSG000000027075.13	-1.0	1.0	ENSG000000027075	PRKCH	
6	ENSG000000161544.9	1.0	-1.0	ENSG000000161544	CYGB	
7	ENSG000000133574.9	-1.0	1.0	ENSG000000133574	GIMAP4	

	adj.P.Val	logFC	t
0	6.562210e-07	0.404942	6.067922
1	1.220536e-02	0.205316	3.352644
2	1.784174e-02	-0.106137	-3.200556
3	2.019820e-02	0.179923	3.146998
4	2.940114e-02	-0.060516	-2.981437
5	3.445104e-02	-0.130863	-2.912396
6	4.099586e-02	0.176233	2.835293
7	4.672012e-02	-0.156888	-2.774555

```
[6]: ch = extract_disconcordant('caudate', 'hippo')
ch.to_csv("disconcordant_genes_%s_%s.csv" % ('caudate', 'hippo'), index=False)
print(ch.shape)
ch
```

(7, 8)

```
[6]:
```

	Feature	Dir_caudate	Dir_hippo	ensemblID	Symbol	\
0	ENSG000000130558.19	-1.0	1.0	ENSG000000130558	OLFM1	
1	ENSG000000106952.7	-1.0	1.0	ENSG000000106952	TNFSF8	
2	ENSG000000196455.7	1.0	-1.0	ENSG000000196455	PIK3R4	
3	ENSG000000138207.13	1.0	-1.0	ENSG000000138207	RBP4	
4	ENSG000000104044.15	-1.0	1.0	ENSG000000104044	OCA2	
5	ENSG000000139719.9	1.0	-1.0	ENSG000000139719	VPS33A	
6	ENSG000000123329.17	-1.0	1.0	ENSG000000123329	ARHGAP9	

	adj.P.Val	logFC	t
0	0.000019	-0.228691	-5.292123
1	0.000043	-0.412720	-5.081967
2	0.000243	0.111391	4.617778
3	0.001274	0.220448	4.142458
4	0.006757	-0.289518	-3.585903
5	0.019055	0.068472	3.169776
6	0.040745	-0.166918	-2.839868

```
[7]: print("There are %d genes where caudate is different from both DLPFC and_
→hippocampus!\n" %
len(set(ch.ensemblID ) & set(cd.ensemblID)))

cd[(cd['ensemblID'].isin(list(set(ch.ensemblID ) & set(cd.ensemblID))))]
```

There are 0 genes where caudate is different from both DLPFC and hippocampus!

```
[7]: Empty DataFrame
Columns: [Feature, Dir_caudate, Dir_dlpfc, ensemblID, Symbol, adj.P.Val, logFC,
t]
Index: []
```

```
[8]: dh = extract_disconcordant('dlpfc', 'hippo')
dh.to_csv("disconcordant_genes_%s_%s.csv" % ('dlpfc', 'hippo'), index=False)
print(dh.shape)
dh
```

(3, 8)

```
[8]:
```

	Feature	Dir_dlpfc	Dir_hippo	ensemblID	Symbol	\
0	ENSG000000078902.15	-1.0	1.0	ENSG000000078902	TOLLIP	
1	ENSG000000114670.13	-1.0	1.0	ENSG000000114670	NEK11	

2	ENSG00000138207.13	1.0	-1.0	ENSG00000138207	RBP4
---	--------------------	-----	------	-----------------	------

	adj.P.Val	logFC	t
0	0.019953	-0.082392	-3.188057
1	0.025698	-0.078108	-3.078731
2	0.040661	0.132448	2.866446

```
[9]: print("There are %d genes where hippocampus is different from both caudate and_
↳DLPFC!\n" %
len(set(dh.ensemblID ) & set(ch.ensemblID)))

dh[(dh['ensemblID'].isin(list(set(dh.ensemblID ) & set(ch.ensemblID))))]
```

There are 1 genes where hippocampus is different from both caudate and DLPFC!

```
[9]:
```

	Feature	Dir_dlpfc	Dir_hippo	ensemblID	Symbol	\
2	ENSG00000138207.13	1.0	-1.0	ENSG00000138207	RBP4	

	adj.P.Val	logFC	t
2	0.040661	0.132448	2.866446

```
[10]: print("There are %d genes where DLPFC is different from both caudate and_
↳hippocampus!\n" %
len(set(dh.ensemblID ) & set(cd.ensemblID)))
```

There are 0 genes where DLPFC is different from both caudate and hippocampus!

```
[11]: cg = extract_disconcordant('caudate', 'gyrus')
cg.to_csv("disconcordant_genes_%s_%s.csv" % ('caudate', 'gyrus'), index=False)
print(cg.shape)
cg
```

(13, 8)

```
[11]:
```

	Feature	Dir_caudate	Dir_gyrus	ensemblID	Symbol	\
0	ENSG000000277883.1	1.0	-1.0	ENSG000000277883	NLRP3P1	
1	ENSG000000075234.16	-1.0	1.0	ENSG000000075234	TTC38	
2	ENSG000000134265.12	1.0	-1.0	ENSG000000134265	NAPG	
3	ENSG000000101224.17	-1.0	1.0	ENSG000000101224	CDC25B	
4	ENSG000000068831.18	1.0	-1.0	ENSG000000068831	RASGRP2	
5	ENSG000000256537.4	1.0	-1.0	ENSG000000256537	SMIM10L1	
6	ENSG000000125845.6	-1.0	1.0	ENSG000000125845	BMP2	
7	ENSG000000105700.10	-1.0	1.0	ENSG000000105700	KXD1	
8	ENSG000000189376.11	1.0	-1.0	ENSG000000189376	C8orf76	
9	ENSG000000228624.7	-1.0	1.0	ENSG000000228624	NaN	
10	ENSG000000106069.20	1.0	-1.0	ENSG000000106069	CHN2	
11	ENSG000000071991.8	1.0	-1.0	ENSG000000071991	CDH19	

12	ENSG00000138640.14	-1.0	1.0	ENSG00000138640	FAM13A
----	--------------------	------	-----	-----------------	--------

	adj.P.Val	logFC	t
0	6.562210e-07	0.404942	6.067922
1	7.654728e-06	-0.301819	-5.508393
2	1.247357e-03	0.080082	4.149428
3	1.598725e-03	-0.118644	-4.069179
4	2.788749e-03	0.139554	3.891750
5	2.984501e-03	0.111219	3.870464
6	1.131502e-02	-0.209347	-3.383479
7	1.440904e-02	-0.071441	-3.288598
8	1.626238e-02	0.145713	3.239705
9	1.774571e-02	-0.110123	-3.202949
10	2.118743e-02	0.144669	3.124922
11	4.012544e-02	0.190290	2.847467
12	4.715735e-02	-0.078567	-2.769428

```
[12]: dg = extract_disconcordant('dlpfc', 'gyrus')
dg.to_csv("disconcordant_genes_%s_%s.csv" % ('dlpfc', 'gyrus'), index=False)
print(dg.shape)
dg
```

(3, 8)

```
[12]:
```

	Feature	Dir_dlpfc	Dir_gyrus	ensemblID	Symbol	\
0	ENSG00000147251.15	1.0	-1.0	ENSG00000147251	DOCK11	
1	ENSG00000188386.6	-1.0	1.0	ENSG00000188386	PPP3R2	
2	ENSG00000280294.1	-1.0	1.0	ENSG00000280294	NaN	

	adj.P.Val	logFC	t
0	0.000729	0.138798	4.368086
1	0.013399	-0.240428	-3.343980
2	0.022723	-0.118239	-3.132760

```
[13]: print("There are %d genes where dentate gyrus is different from both caudate_
↪and DLPFC!\n" %
len(set(cg.ensemblID ) & set(dg.ensemblID)))

cg[(cg['ensemblID'].isin(list(set(dg.ensemblID ) & set(cg.ensemblID))))]
```

There are 0 genes where dentate gyrus is different from both caudate and DLPFC!

```
[13]: Empty DataFrame
Columns: [Feature, Dir_caudate, Dir_gyrus, ensemblID, Symbol, adj.P.Val, logFC,
t]
Index: []
```

```
[14]: hg = extract_disconcordant('hippo', 'gyrus')
hg.to_csv("disconcordant_genes_%s_%s.csv" % ('hippo', 'gyrus'), index=False)
print(hg.shape)
hg
```

(4, 8)

```
[14]:
```

	Feature	Dir_hippo	Dir_gyrus	ensemblID	Symbol	\
0	ENSG000000101224.17	-1.0	1.0	ENSG000000101224	CDC25B	
1	ENSG000000075234.16	-1.0	1.0	ENSG000000075234	TTC38	
2	ENSG000000165410.14	-1.0	1.0	ENSG000000165410	CFL2	
3	ENSG000000214140.10	-1.0	1.0	ENSG000000214140	PRCD	

	adj.P.Val	logFC	t
0	0.000833	-0.137511	-4.293268
1	0.012194	-0.165073	-3.367048
2	0.031319	-0.111948	-2.967453
3	0.043470	-0.165288	-2.808149

```
[15]: print("There are %d genes where dentate gyrus is different from both DLPFC and_
↪hippocampus!\n" %
len(set(hg.ensemblID ) & set(dg.ensemblID)))

hg[(hg['ensemblID'].isin(list(set(hg.ensemblID ) & set(dg.ensemblID))))]
```

There are 0 genes where dentate gyrus is different from both DLPFC and hippocampus!

```
[15]: Empty DataFrame
Columns: [Feature, Dir_hippo, Dir_gyrus, ensemblID, Symbol, adj.P.Val, logFC, t]
Index: []
```

```
[16]: print("There are %d genes where dentate gyrus is different from both caudate,
↪DLPFC, and hippocampus!\n" %
len(set(cg.ensemblID) & set(dg.ensemblID) & set(hg.ensemblID)))

cg[(cg['ensemblID'].isin(list(set(cg.ensemblID) & set(dg.ensemblID) & set(hg.
↪ensemblID))))]
```

There are 0 genes where dentate gyrus is different from both caudate, DLPFC, and hippocampus!

```
[16]: Empty DataFrame
Columns: [Feature, Dir_caudate, Dir_gyrus, ensemblID, Symbol, adj.P.Val, logFC, t]
Index: []
```

[]: