

main

November 24, 2021

## 1 Summary of by sex schizophrenia results

```
[1]: import pandas as pd
```

### 1.1 Functions

```
[2]: def map_tissue(tissue):  
    return {"caudate": "Caudate", "dlpfc": "DLPFC",  
           "hippocampus": "Hippocampus"}[tissue]  
  
def get_degs(tissue):  
    f_file = "../../../%s/female_analysis/metrics_summary/" % tissue +\  
            "_m/female_specific_DE_4features.txt"  
    m_file = "../../../%s/male_analysis/metrics_summary/" % tissue +\  
            "_m/male_specific_DE_4features.txt"  
    female = pd.read_csv(f_file, sep='\t').drop(["Male_Pval", "Male_FDR"],  
↪axis=1)  
    female["Sex"] = "Female"  
    male = pd.read_csv(m_file, sep='\t').drop(["Female_Pval", "Female_FDR"],  
↪axis=1)  
    male["Sex"] = "Male"  
    df = pd.concat([female, male], axis=0)  
    df["Tissue"] = map_tissue(tissue)  
    return df
```

### 1.2 Examine data

#### 1.2.1 Extract DEs

```
[3]: dt = pd.DataFrame()  
for tissue in ["caudate", "dlpfc", "hippocampus"]:  
    dt = pd.concat([dt, get_degs(tissue)])  
dt.shape
```

```
[3]: (10835, 11)
```

### 1.2.2 print summary

#### Features

```
[4]: dt.groupby(["Tissue", "Sex", "Type"]).size()
```

```
[4]: Tissue      Sex      Type      count
      Caudate    Female exon         19
      Caudate    Female gene         30
      Caudate    Female junction      17
      Caudate    Female transcript     11
      Caudate    Male   exon        5834
      Caudate    Male   gene        1858
      Caudate    Male   junction     2064
      Caudate    Male   transcript    480
      DLPFC      Male   exon         116
      DLPFC      Male   gene         122
      DLPFC      Male   junction       20
      DLPFC      Male   transcript     22
      Hippocampus Female junction       5
      Hippocampus Male   exon        113
      Hippocampus Male   gene        104
      Hippocampus Male   junction       3
      Hippocampus Male   transcript     17

dtype: int64
```

#### ensemblID

```
[5]: dt2 = dt.groupby(["Tissue", "Sex", "Type", "ensemblID"]).first().reset_index()
      dt2.groupby(["Tissue", "Sex", "Type"]).size()
```

```
[5]: Tissue      Sex      Type      count
      Caudate    Female exon         15
      Caudate    Female gene         30
      Caudate    Female junction      14
      Caudate    Female transcript     11
      Caudate    Male   exon        1830
      Caudate    Male   gene        1858
      Caudate    Male   junction     1061
      Caudate    Male   transcript    460
      DLPFC      Male   exon          80
      DLPFC      Male   gene         122
      DLPFC      Male   junction       17
      DLPFC      Male   transcript     22
      Hippocampus Female junction       5
      Hippocampus Male   exon         50
      Hippocampus Male   gene        104
      Hippocampus Male   junction       3
      Hippocampus Male   transcript     15

dtype: int64
```

### 1.2.3 Save results

```
[6]: dt.to_csv("differential_expression_schizophrenia_by_sex_4features.txt.gz",  
      ↪sep='\t', index=False)  
dt.head(5)
```

```
[6]:
```

	Feature	gencodeID	Symbol	ensemblID	Chrom	\
0	ENSG00000070915.9	ENSG00000070915.9	SLC12A3	ENSG00000070915	chr16	
1	ENSG00000111181.12	ENSG00000111181.12	SLC6A12	ENSG00000111181	chr12	
2	ENSG00000263006.6	ENSG00000263006.6	ROCK1P1	ENSG00000263006	chr18	
3	ENSG00000224273.2	ENSG00000224273.2	NaN	ENSG00000224273	chr7	
4	ENSG00000249669.8	ENSG00000249669.8	CARMN	ENSG00000249669	chr5	

  

	logFC	t	adj.P.Val	Type	Sex	Tissue
0	0.909953	4.668428	0.006380	gene	Female	Caudate
1	-0.432628	-4.405278	0.009994	gene	Female	Caudate
2	0.536616	4.306700	0.012060	gene	Female	Caudate
3	0.674541	4.289365	0.012708	gene	Female	Caudate
4	-0.527707	-4.202454	0.015476	gene	Female	Caudate

```
[ ]:
```