main

August 6, 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]: (10958, 11)
```

1.2.2 print summary

Features

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

```
[4]: Tissue       Sex     Type
     Caudate      Female  exon            19
                          gene           124
                          junction        17
                          transcript      11
                  Male    exon          5834
                          gene          1858
                          junction      2064
                          transcript     480
     DLPFC        Male    exon           116
                          gene           122
                          junction        49
                          transcript      22
     Hippocampus  Female  junction         5
                  Male    exon           113
                          gene           104
                          junction         3
                          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
     Caudate      Female  exon            15
                          gene           124
                          junction        14
                          transcript      11
                  Male    exon          1830
                          gene          1858
                          junction      1061
                          transcript     460
     DLPFC        Male    exon            80
                          gene           122
                          junction        41
                          transcript      22
     Hippocampus  Female  junction         5
                  Male    exon            50
                          gene           104
                          junction         3
                          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  ENSG00000111181.12  ENSG00000111181.12  SLC6A12  ENSG00000111181  chr12
     1   ENSG00000070915.9   ENSG00000070915.9  SLC12A3  ENSG00000070915  chr16
     2   ENSG00000263006.6   ENSG00000263006.6  ROCK1P1  ENSG00000263006  chr18
     3   ENSG00000141750.6   ENSG00000141750.6    STAC2  ENSG00000141750  chr17
     4   ENSG00000249669.8   ENSG00000249669.8    CARMN  ENSG00000249669   chr5


           logFC         t  adj.P.Val  Type     Sex   Tissue
     0  -0.456838 -4.917349   0.001106  gene  Female  Caudate
     1   0.848495  4.756160   0.001680  gene  Female  Caudate
     2   0.537733  4.498877   0.003265  gene  Female  Caudate
     3   0.264285  4.392359   0.004231  gene  Female  Caudate
     4  -0.536152 -4.285193   0.005433  gene  Female  Caudate
```

```
[ ]:
```