

main

July 10, 2021

1 Generate supplementary data for DEGs

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

1.1 Function

```
[2]: def get_tissues_DEG(tissue):
    cols = ["Feature", "gencodeID", "ensemblID", "Symbol", "logFC",
            "AveExpr", "t", "P.Value", "adj.P.Val", "Type"]
    tissue_map = {"Caudate": "caudate", "Dentate Gyrus": "dentateGyrus",
                  "DLPFC": "dlpfc", "Hippocampus": "hippocampus"}

    gg = pd.read_csv("../../../s/_m/genes/diffExpr_EAvsAA_full.txt" %
    ↪tissue_map[tissue],
                    sep='\t', index_col=0)
    gg["Feature"] = gg.index; gg["Type"] = "Gene"
    tt = pd.read_csv("../../../s/_m/transcripts/diffExpr_EAvsAA_full.txt" %
    ↪tissue_map[tissue],
                    sep='\t', index_col=0)\
        .rename(columns={"gene_id": "gencodeID", "gene_name": "Symbol"})
    tt["ensemblID"] = tt.gencodeID.str.replace("\\.*", "", regex=True)
    tt["Feature"] = tt.index; tt["Type"] = "Transcript"
    ee = pd.read_csv("../../../s/_m/exons/diffExpr_EAvsAA_full.txt" %
    ↪tissue_map[tissue],
                    sep='\t', index_col=0)
    ee["Feature"] = ee.index; ee["Type"] = "Exon"
    jj = pd.read_csv("../../../s/_m/junctions/diffExpr_EAvsAA_full.txt" %
    ↪tissue_map[tissue],
                    sep='\t', index_col=0)\
        .drop(["Symbol"], axis=1)\
        .rename(columns={"newGeneID": "gencodeID", "newGeneSymbol":
    ↪"Symbol"})
    jj["ensemblID"] = jj.gencodeID.str.replace("\\.*", "", regex=True)
    jj["Feature"] = jj.index; jj["Type"] = "Junction"
    df = pd.concat([gg.reset_index().loc[:, cols], tt.reset_index().loc[:,
    ↪cols],
                    ee.reset_index().loc[:, cols], jj.reset_index().loc[:,
    ↪cols]], axis=0)
```

```
df["Tissue"] = tissue
return df
```

1.2 Main

```
[3]: caudate = get_tissues_DEG("Caudate")
      dlpfc = get_tissues_DEG("DLPFC")
      gyrus = get_tissues_DEG("Dentate Gyrus")
      hippo = get_tissues_DEG("Hippocampus")
```

```
[4]: df = pd.concat([caudate, dlpfc, gyrus, hippo], axis=0)
      df.shape
```

```
[4]: (2256376, 11)
```

```
[5]: df.groupby(["Tissue", "Type"]).size()
```

```
[5]: Tissue      Type
Caudate      Exon      347051
           Gene       22374
           Junction  149331
           Transcript  96554
DLPFC        Exon      332329
           Gene       22398
           Junction  141266
           Transcript  77750
Dentate Gyrus Exon      301641
           Gene       21140
           Junction  121134
           Transcript  61669
Hippocampus  Exon      327876
           Gene       22269
           Junction  136285
           Transcript  75309
dtype: int64
```

1.2.1 Save files

```
[6]: df.to_csv("diffExpr_ancestry_full_4regions.tsv", sep='\t', index=False)
```

```
[7]: df[(df["adj.P.Val"] < 0.05)].to_csv("diffExpr_ancestry_FDR05_4regions.tsv",
      ↪sep='\t', index=False)
```