

# main

July 15, 2021

## 1 Summary analysis of interaction in Caudate Nucleus

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

### 1.1 Summary plots

#### 1.1.1 Genes

```
[2]: genes = pd.read_csv('../_m/genes/diffExpr_interaction_full.txt',  
                        sep='\t', index_col=0)  
genes = genes[(genes['adj.P.Val'] < 0.20)].sort_values('adj.P.Val')  
genes['Feature'] = genes.index  
genes = genes[['Feature', 'Symbol', 'ensemblID', 'logFC', 'adj.P.Val']]  
genes['Type'] = 'gene'  
genes.head()
```

```
[2]: Empty DataFrame  
Columns: [Feature, Symbol, ensemblID, logFC, adj.P.Val, Type]  
Index: []
```

#### 1.1.2 Transcripts

```
[3]: trans = pd.read_csv('../_m/transcripts/diffExpr_interaction_full.txt',  
                        sep='\t', index_col=0)  
trans = trans[(trans['adj.P.Val'] < 0.20)].sort_values('adj.P.Val')  
trans['Feature'] = trans.index  
trans['ensemblID'] = trans.gene_id.str.replace('\\.\d+', '')  
trans = trans[['Feature', 'Symbol', 'ensemblID', 'logFC', 'adj.P.Val']]  
#trans.rename(columns={'gene_name': 'Symbol'}, inplace=True)  
trans['Type'] = 'transcript'  
trans.head()
```

<ipython-input-1-8f63412c9da4>:5: FutureWarning: The default value of regex will change from True to False in a future version.

```
trans['ensemblID'] = trans.gene_id.str.replace('\\.\d+', '')
```

```
[3]: Empty DataFrame  
Columns: [Feature, Symbol, ensemblID, logFC, adj.P.Val, Type]
```

Index: []

### 1.1.3 Exons

```
[4]: exons = pd.read_csv('../_m/exons/diffExpr_interaction_full.txt',
                        sep='\t', index_col=0)
exons = exons[(exons['adj.P.Val'] < 0.20)].sort_values('adj.P.Val')
exons['Feature'] = exons.index
exons = exons[['Feature', 'Symbol', 'ensemblID', 'logFC', 'adj.P.Val']]
exons['Type'] = 'exon'
exons.head()
```

```
[4]: Empty DataFrame
Columns: [Feature, Symbol, ensemblID, logFC, adj.P.Val, Type]
Index: []
```

### 1.1.4 Junctions

```
[5]: juncs = pd.read_csv('../_m/junctions/diffExpr_interaction_full.txt',
                        sep='\t', index_col=0)
juncs = juncs[(juncs['adj.P.Val'] < 0.05)].sort_values('adj.P.Val')
juncs['Feature'] = juncs.index
juncs = juncs[['Feature', 'Symbol', 'ensemblID', 'logFC', 'adj.P.Val']]
juncs['Type'] = 'junction'
juncs.head()
```

```
[5]:
```

	Feature	Symbol	\
chr18:31592940-31592993(+)	chr18:31592940-31592993(+)	NaN	
chrX:140784309-140784456(+)	chrX:140784309-140784456(+)	NaN	
chr21:8219273-8402302(-)	chr21:8219273-8402302(-)	NaN	
chr6:31816438-31828631(+)	chr6:31816438-31828631(+)	NaN	
chr10:17881182-17885268(+)	chr10:17881182-17885268(+)	MRC1	

  

	ensemblID	logFC	adj.P.Val	Type
chr18:31592940-31592993(+)	NaN	3.982802	0.005746	junction
chrX:140784309-140784456(+)	NaN	3.240533	0.034982	junction
chr21:8219273-8402302(-)	NaN	-3.091047	0.034982	junction
chr6:31816438-31828631(+)	NaN	-3.978074	0.034982	junction
chr10:17881182-17885268(+)	ENSG00000260314	-1.087409	0.034982	junction

**Junction (chr21:8402231-8402287) within RNA45SN3, there is lots of splicing recorded in sorted BAM.**