

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]:
```

	Feature	Symbol	ensemblID	logFC	\
ENST00000367055.8	ENST00000367055.8	SOD2	ENSG00000112096	3.886009	

```
ENST00000264366.10  ENST00000264366.10  ANK2  ENSG00000145362  5.252028
```

```

adj.P.Val      Type
ENST00000367055.8  0.088229  transcript
ENST00000264366.10  0.158492  transcript

```

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]:
chr21:8401487-8445715(-)  chr21:8401487-8445715(-)  NaN  NaN
chr21:8401417-8445654(-)  chr21:8401417-8445654(-)  NaN  NaN
chr14:49566087-49586737(+)  chr14:49566087-49586737(+)  NaN  NaN
chr21:8401395-8445553(-)  chr21:8401395-8445553(-)  NaN  NaN
chr21:8401970-8446050(+)  chr21:8401970-8446050(+)  NaN  NaN

logFC  adj.P.Val  Type
chr21:8401487-8445715(-)  6.462429  0.000002  junction
chr21:8401417-8445654(-)  3.818954  0.000026  junction
chr14:49566087-49586737(+)  -4.428507  0.000129  junction
chr21:8401395-8445553(-)  5.809398  0.000363  junction
chr21:8401970-8446050(+)  6.047697  0.000673  junction
```

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