main

September 4, 2021

1 Generate supplemental data of all significant integration analysis

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

[2]:
```
genes = pd.read_csv('../../_m/genes/integration_by_symbol.txt', sep='\t').
↪rename(columns={'gene': 'Feature_ID'})
genes['Feature_Type'] = 'Gene'
genes.head()
```

[2]:

	gene_id	Symbol	variant_id	A1	A2	risk_allele	OR	\
0	ENSG00000148842.17	CNNM2	chr10:102825368:C:A	C	A	C	1.0604	
1	ENSG00000148842.17	CNNM2	chr10:102825648:A:T	A	T	A	1.0635	
2	ENSG00000148842.17	CNNM2	chr10:102832918:G:C	G	C	G	1.0644	
3	ENSG00000148842.17	CNNM2	chr10:102835663:T:G	T	G	T	1.0617	
4	ENSG00000148842.17	CNNM2	chr10:102835962:C:T	C	T	C	1.0633	

	P	pval_nominal	adj.P.Val	logFC	t	eqtl_slope	\
0	1.120000e-09	2.413790e-11	0.003899	0.053701	3.799656	-0.169331	
1	2.300000e-10	1.944530e-05	0.003899	0.053701	3.799656	-0.119926	
2	1.330000e-10	3.352800e-06	0.003899	0.053701	3.799656	-0.127135	
3	4.970000e-10	2.310390e-10	0.003899	0.053701	3.799656	-0.163240	
4	2.590000e-10	3.624190e-09	0.003899	0.053701	3.799656	-0.152542	

	de_dir	eqtl_gwas_dir	agree_direction	Feature_Type
0	Up	Down	No	Gene
1	Up	Down	No	Gene
2	Up	Down	No	Gene
3	Up	Down	No	Gene
4	Up	Down	No	Gene

[3]:
```
trans = pd.read_csv('../../_m/transcripts/integration_by_symbol.txt', sep='\t').
↪rename(columns={'gene': 'Feature_ID', 'gene_name': 'Symbol'})
trans['Feature_Type'] = 'Transcript'
trans.head()
```

[3]:
```
              gene_id Symbol           variant_id A1 A2 risk_allele      OR  \
0  ENST00000369878.8  CNNM2  chr10:102825368:C:A  C  A           C  1.0604
1  ENST00000369878.8  CNNM2  chr10:102825648:A:T  A  T           A  1.0635
2  ENST00000369878.8  CNNM2  chr10:102832918:G:C  G  C           G  1.0644
3  ENST00000369878.8  CNNM2  chr10:102835663:T:G  T  G           T  1.0617
4  ENST00000369878.8  CNNM2  chr10:102835962:C:T  C  T           C  1.0633

              P  pval_nominal  adj.P.Val     logFC         t  eqtl_slope  \
0  1.120000e-09  4.928250e-11   0.000852  0.084962  4.753266   -0.179745
1  2.300000e-10  7.617130e-05   0.000852  0.084962  4.753266   -0.119358
2  1.330000e-10  1.354810e-05   0.000852  0.084962  4.753266   -0.128029
3  4.970000e-10  9.108410e-11   0.000852  0.084962  4.753266   -0.179316
4  2.590000e-10  7.879660e-09   0.000852  0.084962  4.753266   -0.160431

  de_dir eqtl_gwas_dir agree_direction Feature_Type
0     Up          Down              No   Transcript
1     Up          Down              No   Transcript
2     Up          Down              No   Transcript
3     Up          Down              No   Transcript
4     Up          Down              No   Transcript
```

[4]:
```
exons = pd.read_csv('../../_m/exons/integration_by_symbol.txt', sep='\t').
↪rename(columns={'gene': 'Feature_ID'})
exons['Feature_Type'] = 'Exon'
exons.head()
```

[4]:
```
    gene_id Symbol           variant_id A1 A2 risk_allele      OR  \
0  e598887  CNNM2  chr10:102825368:C:A  C  A           C  1.0604
1  e598887  CNNM2  chr10:102825648:A:T  A  T           A  1.0635
2  e598887  CNNM2  chr10:102832918:G:C  G  C           G  1.0644
3  e598887  CNNM2  chr10:102835663:T:G  T  G           T  1.0617
4  e598887  CNNM2  chr10:102835962:C:T  C  T           C  1.0633

              P  pval_nominal  adj.P.Val    logFC         t  eqtl_slope  \
0  1.120000e-09  1.275560e-10   0.001824  0.08217  4.275115   -0.172984
1  2.300000e-10  8.685790e-05   0.001824  0.08217  4.275115   -0.117049
2  1.330000e-10  5.547170e-05   0.001824  0.08217  4.275115   -0.117347
3  4.970000e-10  5.187010e-10   0.001824  0.08217  4.275115   -0.169565
4  2.590000e-10  9.561270e-09   0.001824  0.08217  4.275115   -0.157295

  de_dir eqtl_gwas_dir agree_direction Feature_Type
0     Up          Down              No         Exon
1     Up          Down              No         Exon
2     Up          Down              No         Exon
3     Up          Down              No         Exon
4     Up          Down              No         Exon
```

[5]:
```
juncs = pd.read_csv('../../_m/junctions/integration_by_symbol.txt', sep='\t').
↪rename(columns={'gene': 'Feature_ID'})
juncs['Feature_Type'] = 'Junction'
juncs.head()
```

[5]:

	gene_id	Symbol	variant_id	A1	A2	risk_allele	OR	P	pval_nominal	adj.P.Val	logFC	t	eqtl_slope	de_dir	eqtl_gwas_dir	agree_direction	Feature_Type
0	chr11:47180691-47182402(-)	PACSIN3	chr11:46729945:T:TAGG	T	TAGG	TAGG	0.925705	3.875000e-09	0.000221	0.022512	-0.186281	-3.419199	0.208793	Down	Up	No	Junction
1	chr11:47180691-47182402(-)	PACSIN3	chr11:46945082:T:C	T	C	C	0.896840	3.720000e-08	0.000115	0.022512	-0.186281	-3.419199	0.421745	Down	Up	No	Junction
2	chr11:113412884-113415420(-)	DRD2	chr11:113447023:G:C	G	C	G	1.066000	4.300000e-10	0.000011	0.004208	0.302443	4.079084	-0.260710	Up	Down	No	Junction
3	chr11:113412884-113415420(-)	DRD2	chr11:113457959:A:C	A	C	A	1.065800	3.690000e-10	0.000024	0.004208	0.302443	4.079084	-0.252745	Up	Down	No	Junction
4	chr11:113412884-113415420(-)	DRD2	chr11:113463378:C:T	C	T	C	1.064600	8.050000e-10	0.000005	0.004208	0.302443	4.079084	-0.272205	Up	Down	No	Junction

[6]:
```
df = pd.concat([genes, trans, exons, juncs], axis=0, sort=False)
print(df.shape)
df.head(2)
```

```
(13169, 17)
```

[6]:

	gene_id	Symbol	variant_id	A1	A2	risk_allele	OR	P	pval_nominal	adj.P.Val	logFC	t	eqtl_slope	de_dir	eqtl_gwas_dir	agree_direction	Feature_Type
0	ENSG00000148842.17	CNNM2	chr10:102825368:C:A	C	A	C	1.0604	1.120000e-09	2.413790e-11	0.003899	0.053701	3.799656	-0.169331	Up	Down	No	Gene
1	ENSG00000148842.17	CNNM2	chr10:102825648:A:T	A	T	A	1.0635	2.300000e-10	1.944530e-05	0.003899	0.053701	3.799656	-0.119926	Up	Down	No	Gene

[7]:
```
df.to_csv('BrainSeq_Phase3_Caudate_GWAS_DE_eQTL_Integration.txt.gz',␣
↪index=False, header=True, sep='\t')
```