

main

September 7, 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	rsid	A1	\
0	ENSG00000137185.11	ZSCAN9	chr6:28651576:A:G	rs34724414	A	
1	ENSG00000137185.11	ZSCAN9	chr6:28662914:G:A	rs13194504	G	
2	ENSG00000137185.11	ZSCAN9	chr6:28468237:TAAACATA:T	rs138024639	T	
3	ENSG00000137185.11	ZSCAN9	chr6:28594470:TG:T	rs34119086	T	
4	ENSG00000137185.11	ZSCAN9	chr6:28608366:T:A	rs4947336	T	

	A2	OR	P	pval_nominal	adj.P.Val	logFC	\
0	G	1.275700	1.120000e-43	0.001482	0.022627	-0.07713	
1	A	1.275700	1.180000e-43	0.001482	0.022627	-0.07713	
2	TAAACATA	0.770435	2.595000e-43	0.000917	0.022627	-0.07713	
3	TG	0.787730	5.296000e-42	0.001482	0.022627	-0.07713	
4	A	1.268900	7.480000e-42	0.001482	0.022627	-0.07713	

	t	eqtl_slope	de_dir	eqtl_gwas_dir	agree_direction	\
0	-3.126621	-1	-1	-1	Yes	
1	-3.126621	-1	-1	-1	Yes	
2	-3.126621	-1	-1	-1	Yes	
3	-3.126621	-1	-1	-1	Yes	
4	-3.126621	-1	-1	-1	Yes	

	pgc2_a1_same_as_our_counted	Feature_Type
0	False	Gene
1	False	Gene
2	True	Gene
3	True	Gene
4	False	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	rsid	A1	A2	OR	\
0	ENST00000418983.1	HCG4	chr6:29854872:C:T	rs2517857	C	T	1.166700	
1	ENST00000418983.1	HCG4	chr6:29638984:G:A	rs3095268	G	A	1.146100	
2	ENST00000418983.1	HCG4	chr6:29956359:AG:A	rs5875221	A	AG	0.859074	
3	ENST00000418983.1	HCG4	chr6:29852420:A:G	rs9258609	A	G	1.159100	
4	ENST00000418983.1	HCG4	chr6:29850949:A:G	rs2523761	A	G	1.156400	

	P	pval_nominal	adj.P.Val	logFC	t	eqtl_slope	\
0	1.170000e-22	7.560360e-05	0.028722	0.251395	3.568189	1	
1	4.090000e-22	1.498240e-04	0.028722	0.251395	3.568189	1	
2	1.624000e-21	3.937150e-05	0.028722	0.251395	3.568189	1	
3	1.860000e-21	1.103540e-04	0.028722	0.251395	3.568189	1	
4	2.000000e-21	1.462340e-09	0.028722	0.251395	3.568189	1	

	de_dir	eqtl_gwas_dir	agree_direction	pgc2_a1_same_as_our_counted	\
0	1	1	Yes	False	
1	1	1	Yes	False	
2	1	1	Yes	True	
3	1	1	Yes	False	
4	1	1	Yes	False	

	Feature_Type
0	Transcript
1	Transcript
2	Transcript
3	Transcript
4	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	rsid	A1	A2	OR	\
0	e378566	ZSCAN26	chr6:28744470:A:G	rs1233578	A	G	1.22280	
1	e378558	ZSCAN26	chr6:28744470:A:G	rs1233578	A	G	1.22280	
2	e378566	ZSCAN26	chr6:28716406:T:G	rs9295778	T	G	0.82449	
3	e378558	ZSCAN26	chr6:28716406:T:G	rs9295778	T	G	0.82449	
4	e378566	ZSCAN26	chr6:28721895:C:A	rs9468413	C	A	0.82511	

	P	pval_nominal	adj.P.Val	logFC	t	eqtl_slope	\
0	4.320000e-44	7.294750e-10	0.018675	-0.084735	-3.408383	-1	

1	4.320000e-44	1.080300e-09	0.019079	-0.084425	-3.399939	-1
2	4.520000e-42	6.790160e-10	0.018675	-0.084735	-3.408383	-1
3	4.520000e-42	8.899240e-10	0.019079	-0.084425	-3.399939	-1
4	6.860000e-42	6.790160e-10	0.018675	-0.084735	-3.408383	-1

	de_dir	eqtl_gwas_dir	agree_direction	pgc2_a1_same_as_our_counted	\
0	-1	-1	Yes	False	
1	-1	-1	Yes	False	
2	-1	-1	Yes	False	
3	-1	-1	Yes	False	
4	-1	-1	Yes	False	

	Feature_Type
0	Exon
1	Exon
2	Exon
3	Exon
4	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	rsid	\
0	chr6:28271935-28272021(+)	ZSCAN26	chr6:28744470:A:G	rs1233578	
1	chr6:28271935-28272021(+)	ZSCAN26	chr6:28468237:TAAACATA:T	rs138024639	
2	chr6:28271935-28272021(+)	ZSCAN26	chr6:28716406:T:G	rs9295778	
3	chr6:28271935-28272021(+)	ZSCAN26	chr6:28721895:C:A	rs9468413	
4	chr6:28271935-28272021(+)	ZSCAN26	chr6:28426903:C:T	rs13201681	

	A1	A2	OR	P	pval_nominal	adj.P.Val	logFC	\
0	A	G	1.222800	4.320000e-44	9.439080e-11	0.018932	-0.26177	
1	T	TAAACATA	0.770435	2.595000e-43	2.249520e-04	0.018932	-0.26177	
2	T	G	0.824490	4.520000e-42	1.737900e-10	0.018932	-0.26177	
3	C	A	0.825110	6.860000e-42	1.737900e-10	0.018932	-0.26177	
4	C	T	1.265200	1.320000e-41	2.249520e-04	0.018932	-0.26177	

	t	eqtl_slope	de_dir	eqtl_gwas_dir	agree_direction	\
0	-3.493012	-1	-1	-1	Yes	
1	-3.493012	-1	-1	-1	Yes	
2	-3.493012	-1	-1	-1	Yes	
3	-3.493012	-1	-1	-1	Yes	
4	-3.493012	-1	-1	-1	Yes	

	pgc2_a1_same_as_our_counted	Feature_Type
0	False	Junction

1	True	Junction
2	False	Junction
3	False	Junction
4	False	Junction

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

(13169, 18)

```
[6]:
```

	gene_id	Symbol	variant_id	rsid	A1	A2	OR	\
0	ENSG00000137185.11	ZSCAN9	chr6:28651576:A:G	rs34724414	A	G	1.2757	
1	ENSG00000137185.11	ZSCAN9	chr6:28662914:G:A	rs13194504	G	A	1.2757	

	P	pval_nominal	adj.P.Val	logFC	t	eqtl_slope	\
0	1.120000e-43	0.001482	0.022627	-0.07713	-3.126621	-1	
1	1.180000e-43	0.001482	0.022627	-0.07713	-3.126621	-1	

	de_dir	eqtl_gwas_dir	agree_direction	pgc2_a1_same_as_our_counted	\
0	-1	-1	Yes	False	
1	-1	-1	Yes	False	

	Feature_Type
0	Gene
1	Gene

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