

# main

September 4, 2021

## 1 Examine integration results for TWAS associations

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

```
[2]: df = pd.read_csv("../manuscript_supp_data/_m/
↳BrainSeq_Phase3_Caudate_GWAS_DE_eQTL_Integration.txt.gz", sep='\t')
df = df[(df["Feature_Type"] == "Gene")].copy()
df["ensemblID"] = df.gene_id.str.replace("\\.*", "", regex=True)
df = df.groupby("gene_id").first().reset_index()
df.head()
```

```
[2]:
```

	gene_id	Symbol	variant_id	A1	A2	risk_allele	OR	\
0	ENSG00000028116.16	VRK2	chr2:57874972:T:A	T	A	A	0.91021	
1	ENSG00000066248.14	NGEF	chr2:232843683:G:A	G	A	A	0.91758	
2	ENSG00000100403.11	ZC3H7B	chr22:41357599:G:A	G	A	A	0.94102	
3	ENSG00000104765.15	BNIP3L	chr8:26368096:G:A	G	A	A	0.93463	
4	ENSG00000105708.8	ZNF14	chr19:19623068:T:C	T	C	T	1.06130	

	P	pval_nominal	adj.P.Val	logFC	t	eqtl_slope	\
0	1.370000e-08	6.451330e-04	0.010761	-0.127117	-3.431328	0.150703	
1	7.130000e-13	2.415450e-04	0.009144	0.104229	3.493550	-0.144167	
2	1.760000e-08	1.177050e-07	0.012553	-0.055238	-3.367649	0.258939	
3	2.300000e-08	1.403510e-04	0.010653	-0.063722	-3.435641	-0.101083	
4	4.300000e-08	3.436760e-06	0.016532	0.074769	3.254277	0.231031	

	de_dir	eqtl_gwas_dir	agree_direction	Feature_Type	ensemblID
0	Down	Up	No	Gene	ENSG00000028116
1	Up	Down	No	Gene	ENSG00000066248
2	Down	Up	No	Gene	ENSG00000100403
3	Down	Down	Yes	Gene	ENSG00000104765
4	Up	Up	Yes	Gene	ENSG00000105708

```
[3]: twas_df = pd.read_csv("../twas/feature_comparison/manuscript_supp_data/_m/
↳BrainSeq_Phase3_Caudate_TWAS_associations_allFeatures.txt.gz", sep='\t')
twas_df = twas_df[(twas_df["Type"] == "Gene")].copy()
twas_df.head()
```

```
[3]:
```

	Feature	ensemblID	ID	HSQ	BEST.GWAS.ID \
0	ENSG000000138944	ENSG000000138944	KIAA1644	0.185313	chr22:43809985:A:G
1	ENSG000000185052	ENSG000000185052	SLC24A3	0.178962	chr20:18949619:C:T
2	ENSG000000188730	ENSG000000188730	VWC2	0.463808	chr7:49469103:C:T
3	ENSG000000123352	ENSG000000123352	SPATS2	0.078277	chr12:49368582:G:A
4	ENSG000000156486	ENSG000000156486	KCNS2	0.128937	chr8:98176494:A:G

	EQTL.ID	TWAS.Z	TWAS.P	FDR	Bonferroni	Type
0	chr22:44052458:G:A	-0.025951	0.979296	0.992839	1.0	Gene
1	chr20:19234998:A:G	0.210426	0.833335	0.940940	1.0	Gene
2	chr7:49579650:A:G	0.302719	0.762104	0.913105	1.0	Gene
3	chr12:49672264:G:A	-0.029423	0.976527	0.991655	1.0	Gene
4	chr8:98426617:T:C	1.620339	0.105159	0.371742	1.0	Gene

```
[4]: len(set(df.ensemblID) & set(twas_df.ensemblID))
```

```
[4]: 26
```

```
[5]: df[(df["ensemblID"].isin(list(set(df.ensemblID) & set(twas_df.ensemblID))))]
```

```
[5]:
```

	gene_id	Symbol	variant_id	A1	A2 \
0	ENSG000000028116.16	VRK2	chr2:57874972:T:A	T	A
4	ENSG000000105708.8	ZNF14	chr19:19623068:T:C	T	C
5	ENSG000000105894.11	PTN	chr7:137363413:T:C	T	C
6	ENSG000000112511.17	PHF1	chr6:33343202:G:A	G	A
7	ENSG000000114054.13	PCCB	chr3:136088767:G:T	G	T
8	ENSG000000120733.13	KDM3B	chr5:138362610:C:T	C	T
9	ENSG000000124613.8	ZNF391	chr6:26990749:T:C	T	C
10	ENSG000000132563.15	REEP2	chr5:138362610:C:T	C	T
11	ENSG000000137185.11	ZSCAN9	chr6:27731058:C:T	C	T
14	ENSG000000148842.17	CNNM2	chr10:102825368:C:A	C	A
15	ENSG000000149527.17	PLCH2	chr1:2440958:A:G	A	G
17	ENSG000000161896.11	IP6K3	chr6:33541236:A:G	A	G
18	ENSG000000164088.17	PPM1M	chr3:52253452:G:T	G	T
19	ENSG000000166166.12	TRMT61A	chr14:103541799:CAGG:C	CAGG	C
20	ENSG000000171045.14	TSNARE1	chr8:142261784:C:T	C	T
21	ENSG000000176998.4	HCG4	chr6:29298706:G:A	G	A
23	ENSG000000198911.11	SREBF2	chr22:41830391:C:T	C	T
24	ENSG000000204256.12	BRD2	chr6:32669525:G:A	G	A
25	ENSG000000204356.12	NELFE	chr6:31868665:A:G	A	G
26	ENSG000000204463.12	BAG6	chr6:31222761:G:A	G	A
28	ENSG000000204789.4	ZNF204P	chr6:27343879:A:G	A	G
30	ENSG000000228223.2	HCG11	chr6:26466161:G:A	G	A
31	ENSG000000228944.1	ENSG000000228944.1	chr7:24695385:G:C	G	C
32	ENSG000000244731.7	C4A	chr6:31482137:T:C	T	C
33	ENSG000000249484.8	LINC01470	chr5:152727548:A:G	A	G
34	ENSG000000253553.5	ENSG000000253553.5	chr8:88271460:A:G	A	G

	risk_allele	OR	P	pval_nominal	adj.P.Val	logFC	\
0	A	0.91021	1.370000e-08	6.451330e-04	1.076112e-02	-0.127117	
4	T	1.06130	4.300000e-08	3.436760e-06	1.653178e-02	0.074769	
5	C	0.94234	2.770000e-09	2.497190e-04	7.972676e-03	-0.086066	
6	A	0.94764	4.060000e-08	1.541960e-04	2.281934e-02	-0.046170	
7	T	0.94493	5.050000e-09	2.185890e-14	3.397384e-02	-0.058710	
8	C	1.05600	1.860000e-08	1.455140e-04	1.518040e-02	-0.031591	
9	C	0.94335	1.550000e-09	1.247880e-03	2.054017e-02	0.082216	
10	C	1.05600	1.860000e-08	3.010360e-07	6.655155e-09	0.137644	
11	C	1.26220	1.210000e-39	8.946610e-04	2.262693e-02	-0.077130	
14	C	1.06040	1.120000e-09	2.413790e-11	3.899296e-03	0.053701	
15	G	0.92873	4.630000e-11	2.811440e-09	1.909382e-04	-0.169308	
17	G	0.93912	2.350000e-10	1.762100e-04	2.435252e-03	-0.233018	
18	G	1.05600	1.350000e-08	6.597780e-04	8.758407e-06	-0.136491	
19	C	0.93885	4.218000e-09	1.193570e-04	4.958693e-02	-0.058611	
20	C	1.07400	1.150000e-08	4.723740e-05	4.121874e-03	-0.076791	
21	G	1.26360	2.500000e-39	4.443110e-05	1.090935e-02	0.231187	
23	C	1.05530	2.290000e-08	9.115040e-05	9.947151e-04	-0.075761	
24	G	1.21970	6.820000e-30	1.131810e-05	1.718619e-02	-0.051562	
25	A	1.07930	1.660000e-13	7.305820e-04	2.507792e-02	-0.046326	
26	A	0.92265	6.790000e-14	8.199240e-05	1.241493e-02	-0.044002	
28	G	0.93306	4.310000e-13	1.058040e-03	1.603175e-02	0.095817	
30	A	0.91432	1.020000e-14	1.374100e-05	1.181958e-03	0.114043	
31	C	0.91974	2.310000e-08	1.514450e-04	2.256942e-02	0.278474	
32	T	1.06010	4.120000e-08	1.330480e-04	7.880375e-03	0.358141	
33	A	1.06910	8.730000e-10	5.630250e-14	1.143030e-02	0.406227	
34	A	1.06200	4.410000e-08	3.018090e-04	2.909648e-02	0.202146	

	t	eqtl_slope	de_dir	eqtl_gwas_dir	agree_direction	Feature_Type	\
0	-3.431328	0.150703	Down	Up	No	Gene	
4	3.254277	0.231031	Up	Up	Yes	Gene	
5	-3.545516	0.122412	Down	Up	No	Gene	
6	-3.123441	-0.161077	Down	Down	Yes	Gene	
7	-2.946922	-0.246348	Down	Down	Yes	Gene	
8	-3.287837	-0.167676	Down	Down	Yes	Gene	
9	3.166580	-0.145940	Up	Down	No	Gene	
10	6.993625	0.145249	Up	Up	Yes	Gene	
11	-3.126621	-0.307218	Down	Down	Yes	Gene	
14	3.799656	-0.169331	Up	Down	No	Gene	
15	-4.727576	-0.174917	Down	Down	Yes	Gene	
17	-3.959823	0.131471	Down	Up	No	Gene	
18	-5.484394	-0.116511	Down	Down	Yes	Gene	
19	-2.775075	-0.127611	Down	Down	Yes	Gene	
20	-3.780711	0.177779	Down	Up	No	Gene	
21	3.425733	0.374783	Up	Up	Yes	Gene	
23	-4.243388	-0.177463	Down	Down	Yes	Gene	

24	-3.239197	-0.404029	Down	Down	Yes	Gene
25	-3.081897	0.090267	Down	Up	No	Gene
26	-3.373864	0.148294	Down	Up	No	Gene
28	3.267501	0.089913	Up	Up	Yes	Gene
30	4.193600	0.211170	Up	Up	Yes	Gene
31	3.127826	0.412318	Up	Up	Yes	Gene
32	3.550690	0.178003	Up	Up	Yes	Gene
33	3.407085	0.488819	Up	Up	Yes	Gene
34	3.016220	-0.220670	Up	Down	No	Gene

```

ensemblID
0  ENSG00000028116
4  ENSG000000105708
5  ENSG000000105894
6  ENSG000000112511
7  ENSG000000114054
8  ENSG000000120733
9  ENSG000000124613
10 ENSG000000132563
11 ENSG000000137185
14 ENSG000000148842
15 ENSG000000149527
17 ENSG000000161896
18 ENSG000000164088
19 ENSG000000166166
20 ENSG000000171045
21 ENSG000000176998
23 ENSG000000198911
24 ENSG000000204256
25 ENSG000000204356
26 ENSG000000204463
28 ENSG000000204789
30 ENSG000000228223
31 ENSG000000228944
32 ENSG000000244731
33 ENSG000000249484
34 ENSG000000253553

```

[ ]: