main

September 7, 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        rsid A1 A2       OR  \
     0  ENSG00000028116.16    VRK2   chr2:57944085:A:C   rs1518393  A  C  0.94149
     1  ENSG00000066248.14    NGEF  chr2:232925939:G:T   rs1878287  G  T  0.92380
     2  ENSG00000100403.11  ZC3H7B  chr22:41357599:G:A  rs11090045  G  A  0.94102
     3  ENSG00000104765.15  BNIP3L   chr8:26411675:C:T   rs1042992  C  T  0.92930
     4   ENSG00000105708.8   ZNF14  chr19:19629920:C:G    rs880090  C  G  1.06150

                   P  pval_nominal  adj.P.Val      logFC         t  eqtl_slope  \
     0  7.780000e-10  4.532180e-04   0.010761  -0.127117 -3.431328           1
     1  1.230000e-13  2.215200e-04   0.009144   0.104229  3.493550          -1
     2  1.760000e-08  1.177050e-07   0.012553  -0.055238 -3.367649           1
     3  3.670000e-09  1.163410e-04   0.010653  -0.063722 -3.435641          -1
     4  2.440000e-08  6.133540e-04   0.016532   0.074769  3.254277           1

        de_dir  eqtl_gwas_dir agree_direction  pgc2_a1_same_as_our_counted  \
     0      -1              1              No                        False
     1       1             -1              No                        False
     2      -1              1              No                        False
     3      -1             -1             Yes                        False
     4       1              1             Yes                        False

        Feature_Type        ensemblID
     0          Gene  ENSG00000028116
     1          Gene  ENSG00000066248
     2          Gene  ENSG00000100403
     3          Gene  ENSG00000104765
```

```
4           Gene   ENSG00000105708
```

```python
[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") & (twas_df["FDR"] < 0.05)].copy()
     twas_df.head()
```

```
[3]:           Feature         ensemblID        ID       HSQ         BEST.GWAS.ID  \
     15  ENSG00000132563  ENSG00000132563     REEP2  0.077983  chr5:138519616:T:C
     34  ENSG00000171246  ENSG00000171246     NPTX1  0.199819  chr17:80587803:G:A
     52  ENSG00000165434  ENSG00000165434    PGM2L1  0.160964  chr11:74445493:G:A
     63  ENSG00000065833  ENSG00000065833       ME1  0.090729   chr6:83697747:C:T
     70  ENSG00000146828  ENSG00000146828   SLC12A9  0.082561  chr7:100717894:A:G

                   EQTL.ID     TWAS.Z        TWAS.P       FDR  Bonferroni  Type
     15  chr5:138453285:C:A   3.355851  7.912109e-04  0.014566    1.000000  Gene
     34   chr17:80546983:C:T   3.547619  3.887305e-04  0.008697    1.000000  Gene
     52  chr11:74393463:T:C  -3.322061  8.935515e-04  0.015812    1.000000  Gene
     63   chr6:83457558:A:G  -5.406853  6.414180e-08  0.000006    0.000527  Gene
     70  chr7:100846338:G:A   3.167074  1.539812e-03  0.023946    1.000000  Gene
```

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

```
[4]: 16
```

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

```
[5]:             gene_id             Symbol         variant_id          rsid  \
     6   ENSG00000112511.17            PHF1    chr6:33427422:G:A     rs9461856
     7   ENSG00000114054.13            PCCB   chr3:136569563:G:A     rs7432375
     8   ENSG00000120733.13           KDM3B   chr5:138439892:C:G      rs982085
     9    ENSG00000124613.8           ZNF391    chr6:27424023:C:T    rs34071253
     10  ENSG00000132563.15           REEP2   chr5:138439892:C:G      rs982085
     14  ENSG00000148842.17           CNNM2  chr10:102852578:T:A    rs11191419
     17  ENSG00000161896.11            IP6K3    chr6:33773939:A:G     rs4711350
     19  ENSG00000166166.12          TRMT61A  chr14:103809712:C:T    rs12880821
     23  ENSG00000198911.11           SREBF2   chr22:41938949:C:G     rs1047997
     24  ENSG00000204256.12             BRD2    chr6:32669525:G:A     rs2395231
     26  ENSG00000204463.12             BAG6    chr6:31358852:C:T     rs2523594
     30   ENSG00000228223.2            HCG11    chr6:26466161:G:A     rs1977199
     31   ENSG00000228944.1  ENSG00000228944.1    chr7:24695385:G:C   rs112509803
     32   ENSG00000244731.7              C4A    chr6:31793436:G:A     rs2607014
     33   ENSG00000249484.8         LINC01470   chr5:152797561:A:G   rs111294930
     34   ENSG00000253553.5  ENSG00000253553.5    chr8:88310775:A:C     rs3844198

         A1 A2      OR             P  pval_nominal    adj.P.Val      logFC  \
     6    G  A  0.93180  1.680000e-13  2.090630e-04  2.281934e-02  -0.046170
```

```
7    G  A  1.07020  4.070000e-12  3.669320e-14  3.397384e-02 -0.058710
8    C  G  1.05710  1.030000e-08  8.821430e-05  1.518040e-02 -0.031591
9    C  T  1.26640  6.220000e-41  4.376050e-04  2.054017e-02  0.082216
10   C  G  1.05710  1.030000e-08  3.553170e-06  6.655155e-09  0.137644
14   T  A  1.08750  2.090000e-16  2.573560e-10  3.899296e-03  0.053701
17   A  G  0.92386  2.300000e-10  3.005850e-10  2.435252e-03 -0.233018
19   C  T  0.93470  4.690000e-12  3.767620e-04  4.958693e-02 -0.058611
23   C  G  1.08230  8.080000e-11  6.029740e-04  9.947151e-04 -0.075761
24   G  A  1.21970  6.820000e-30  1.131810e-05  1.718619e-02 -0.051562
26   C  T  0.91831  2.410000e-16  7.590350e-04  1.241493e-02 -0.044002
30   G  A  0.91432  1.020000e-14  1.374100e-05  1.181958e-03  0.114043
31   G  C  0.91974  2.310000e-08  1.514450e-04  2.256942e-02  0.278474
32   G  A  1.20320  1.200000e-31  1.732000e-33  7.880375e-03  0.358141
33   A  G  1.08580  9.040000e-12  6.338820e-14  1.143030e-02  0.406227
34   A  C  1.06450  1.210000e-08  7.121140e-04  2.909648e-02  0.202146
```

	t	eqtl_slope	de_dir	eqtl_gwas_dir	agree_direction \
6	-3.123441	-1	-1	-1	Yes
7	-2.946922	-1	-1	-1	Yes
8	-3.287837	-1	-1	-1	Yes
9	3.166580	-1	1	-1	No
10	6.993625	1	1	1	Yes
14	3.799656	-1	1	-1	No
17	-3.959823	1	-1	1	No
19	-2.775075	-1	-1	-1	Yes
23	-4.243388	-1	-1	-1	Yes
24	-3.239197	-1	-1	-1	Yes
26	-3.373864	1	-1	1	No
30	4.193600	1	1	1	Yes
31	3.127826	1	1	1	Yes
32	3.550690	1	1	1	Yes
33	3.407085	1	1	1	Yes
34	3.016220	-1	1	-1	No

	pgc2_a1_same_as_our_counted	Feature_Type	ensemblID
6	False	Gene	ENSG00000112511
7	False	Gene	ENSG00000114054
8	False	Gene	ENSG00000120733
9	False	Gene	ENSG00000124613
10	False	Gene	ENSG00000132563
14	False	Gene	ENSG00000148842
17	False	Gene	ENSG00000161896
19	False	Gene	ENSG00000166166
23	False	Gene	ENSG00000198911
24	False	Gene	ENSG00000204256
26	False	Gene	ENSG00000204463
30	False	Gene	ENSG00000228223

```
31              False     Gene   ENSG00000228944
32              False     Gene   ENSG00000244731
33              False     Gene   ENSG00000249484
34              False     Gene   ENSG00000253553
```

[ ]: