main

March 7, 2022

1 TWAS tissue comparison

```
[1]: import numpy as np
     import pandas as pd
     from matplotlib import pyplot as plt
     from matplotlib_venn import venn3, venn3_circles
```

1.1 Prepare data

```
[2]: def limiting_features(set_dict, f1, f2):
         xx = len(set_dict[f1] & set_dict[f2]) / len(set_dict[f2]) * 100
         print("Comparing %s with %s: %0.2f%%" % (f1, f2, xx))
         print("Features in common: %d" % len(set_dict[f1] & set_dict[f2]))
```

1.1.1 Load PGC2+CLOZUK

```
[3]: pgc2_file = '/ceph/projects/v4_phase3_paper/inputs/sz_gwas/pgc2_clozuk/
     ↪map_phase3/_m/libd_hg38_pgc2sz_snps.tsv'
     pgc2_df = pd.read_csv(pgc2_file, sep='\t', low_memory=False, index_col=0)
```

```
/home/jbenja13/.local/lib/python3.9/site-packages/numpy/lib/arraysetops.py:583:
FutureWarning: elementwise comparison failed; returning scalar instead, but in
the future will perform elementwise comparison
  mask |= (ar1 == a)
```

1.1.2 Load TWAS associations

```
[4]: caudate_file = '/ceph/projects/v4_phase3_paper/analysis/twas_ea/'+\
                    'gene_weights/fusion_pgc2/summary_stats/_m/fusion_associations.
     ↪txt'
     caudate0 = pd.read_csv(caudate_file, sep='\t')
     caudate = caudate0[(caudate0['FDR'] <= 0.05)].copy()
     print(caudate.shape[0])
```

```
481
```

```
[5]: dlpfc_file = '/ceph/users/jbenja13/phase3_paper/phase2/twas/extract_twas/_m/
     ↪dlpfc_twas_assocations_fusion.csv'
```

```
dlpfc0 = pd.read_csv(dlpfc_file, low_memory=False)
dlpfc0 = dlpfc0[(dlpfc0['feature'] == 'gene') & (dlpfc0['region'] == 'DLPFC')].
 ↪drop('FILE', axis=1).copy()
dlpfc0 = dlpfc0.loc[:, ['ID', 'genesymbol', 'CHR', 'P0', 'P1', 'HSQ', 'BEST.
 ↪GWAS.ID', 'BEST.GWAS.Z',
                        'EQTL.ID', 'EQTL.R2', 'EQTL.Z', 'EQTL.GWAS.Z', 'NSNP',␣
 ↪'NWGT', 'MODEL',
                        'MODELCV.R2', 'MODELCV.PV', 'TWAS.Z', 'TWAS.P', 'TWAS.
 ↪FDR', 'TWAS.Bonf',
                        'BEST.GWAS.pos_hg19', 'BEST.GWAS.pos_hg38']]\
            .rename(columns={'ID': 'FILE', 'genesymbol': 'ID', 'TWAS.FDR':␣
 ↪'FDR'})
dlpfc0.FILE = dlpfc0.FILE.str.replace('\\..*', '', regex=True)
dlpfc = dlpfc0[(dlpfc0['FDR'] <= 0.05)]
print(dlpfc.shape[0])
```

406

```
[6]: hippo_file = '/ceph/users/jbenja13/phase3_paper/phase2/twas/extract_twas/_m/
      ↪hippo_twas_assocations_fusion.csv'
     hippo0 = pd.read_csv(hippo_file, low_memory=False)
     hippo0 = hippo0[(hippo0['feature'] == 'gene') & (hippo0['region'] == 'HIPPO')].
      ↪drop('FILE', axis=1).copy()
     hippo0 = hippo0.loc[:, ['ID', 'genesymbol', 'CHR', 'P0', 'P1', 'HSQ', 'BEST.
      ↪GWAS.ID', 'BEST.GWAS.Z',
                             'EQTL.ID', 'EQTL.R2', 'EQTL.Z', 'EQTL.GWAS.Z', 'NSNP',␣
      ↪'NWGT', 'MODEL',
                             'MODELCV.R2', 'MODELCV.PV', 'TWAS.Z', 'TWAS.P', 'TWAS.
      ↪FDR', 'TWAS.Bonf',
                             'BEST.GWAS.pos_hg19', 'BEST.GWAS.pos_hg38']]\
                 .rename(columns={'ID': 'FILE', 'genesymbol': 'ID', 'TWAS.FDR':␣
      ↪'FDR'})
     hippo0.FILE = hippo0.FILE.str.replace('\\..*', '', regex=True)
     hippo = hippo0[(hippo0['FDR'] <= 0.05)]
     print(hippo.shape[0])
```

270

```
[7]: ## Caudate
     new_caudate0 = pd.merge(caudate0, pgc2_df, left_on='BEST.GWAS.ID',␣
      ↪right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
     new_caudate0['GWAS.SNP'] = ['Risk SNP' if x <= 5e-8 else 'Other' for x in␣
      ↪new_caudate0['P']]
     ## DLPFC
     new_dlpfc0 = pd.merge(dlpfc0, pgc2_df, left_on='BEST.GWAS.ID', right_on='rsid',␣
      ↪suffixes=['_TWAS', '_PGC2'])
```

```python
new_dlpfc0['GWAS.SNP'] = ['Risk SNP' if x <= 5e-8 else 'Other' for x in
 →new_dlpfc0['P']]
## Hippocampus
new_hippo0 = pd.merge(hippo0, pgc2_df, left_on='BEST.GWAS.ID', right_on='rsid',
 →suffixes=['_TWAS', '_PGC2'])
new_hippo0['GWAS.SNP'] = ['Risk SNP' if x <= 5e-8 else 'Other' for x in
 →new_hippo0['P']]
```

```python
[8]: tt = pd.merge(new_caudate0[['FILE', 'ID', 'TWAS.Z', 'FDR', 'GWAS.SNP']].
      →set_index(['FILE', 'ID']).add_prefix('Caudate_'),
                 new_dlpfc0[['FILE', 'ID', 'TWAS.Z', 'FDR', 'GWAS.SNP']].
      →set_index(['FILE', 'ID']).add_prefix('DLPFC_'),
                 left_index=True, right_index=True, how='outer')\
          .merge(new_hippo0[['FILE', 'ID', 'TWAS.Z', 'FDR', 'GWAS.SNP']].
      →set_index(['FILE', 'ID']).add_prefix('HIPPO_'),
                 left_index=True, right_index=True, how='outer')\
          .reset_index().rename(columns={'FILE': 'Geneid', 'ID': 'Symbol'})
     tt.sort_values("Caudate_FDR").to_csv('TWAS_gene_tissue_summary.csv',
      →index=False, header=True)
     tt.head(2)
```

/home/jbenja13/.local/lib/python3.9/site-
packages/pandas/core/indexes/multi.py:3587: RuntimeWarning: The values in the
array are unorderable. Pass `sort=False` to suppress this warning.
  uniq_tuples = lib.fast_unique_multiple([self._values, other._values],
sort=sort)

```
[8]:            Geneid    Symbol  Caudate_TWAS.Z  Caudate_FDR Caudate_GWAS.SNP  \
     0  ENSG00000000457     SCYL3        1.880729     0.269026            Other
     1  ENSG00000000460   C1orf112       -1.843971     0.283386            Other

        DLPFC_TWAS.Z  DLPFC_FDR DLPFC_GWAS.SNP  HIPPO_TWAS.Z  HIPPO_FDR  \
     0           NaN        NaN            NaN           NaN        NaN
     1           NaN        NaN            NaN           NaN        NaN

        HIPPO_GWAS.SNP
     0             NaN
     1             NaN
```

1.2 Plot Venn

1.2.1 Genes

Heritable comparison

```python
[9]: tissues = {
         'Caudate': set(caudate0.FILE),
         'DLPFC': set(dlpfc0.FILE),
         'Hippocampus': set(hippo0.FILE),
```
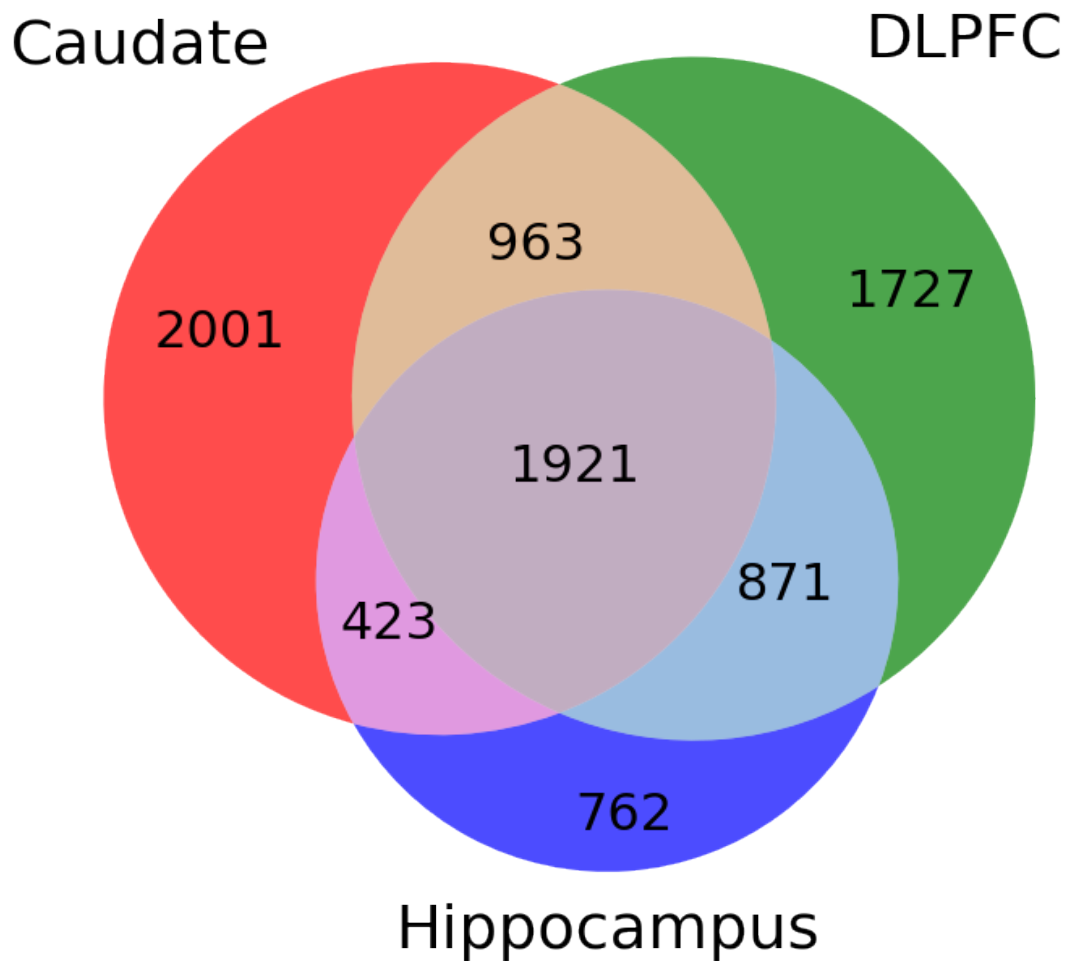
```
        }
```

```
[10]: plt.rcParams.update({'font.size': 32, 'font.weight': 'normal'})
       plt.figure(figsize=(12,12))

       v = venn3([tissues['Caudate'], tissues['DLPFC'], tissues['Hippocampus']],
                 ('Caudate', 'DLPFC', 'Hippocampus'))

       v.get_patch_by_id('100').set_alpha(0.7)
       v.get_patch_by_id('010').set_alpha(0.7)
       v.get_patch_by_id('001').set_alpha(0.7)

       plt.savefig('twas_tissue_comparison_allFeatures.png')
       plt.savefig('twas_tissue_comparison_allFeatures.pdf')
       plt.savefig('twas_tissue_comparison_allFeatures.svg')
       plt.show()
```

```
[11]: limiting_features(tissues, 'Caudate', 'Hippocampus')
      limiting_features(tissues, 'DLPFC', 'Caudate')
      limiting_features(tissues, 'DLPFC', 'Hippocampus')
```

```
Comparing Caudate with Hippocampus: 58.94%
Features in common: 2344
Comparing DLPFC with Caudate: 54.33%
Features in common: 2884
Comparing DLPFC with Hippocampus: 70.20%
Features in common: 2792
```

TWAS significant

```
[12]: tissues = {
          'Caudate': set(caudate.FILE),
          'DLPFC': set(dlpfc.FILE),
          'Hippocampus': set(hippo.FILE),
      }
```

```
[13]: plt.rcParams.update({'font.size': 32, 'font.weight': 'normal'})
      plt.figure(figsize=(12,12))

      v = venn3([tissues['Caudate'], tissues['DLPFC'], tissues['Hippocampus']],
                ('Caudate', 'DLPFC', 'Hippocampus'))

      v.get_patch_by_id('100').set_alpha(0.7)
      v.get_patch_by_id('010').set_alpha(0.7)
      v.get_patch_by_id('001').set_alpha(0.7)

      plt.savefig('twas_tissue_comparison_fdr05.png')
      plt.savefig('twas_tissue_comparison_fdr05.pdf')
      plt.savefig('twas_tissue_comparison_fdr05.svg')
      plt.show()
```
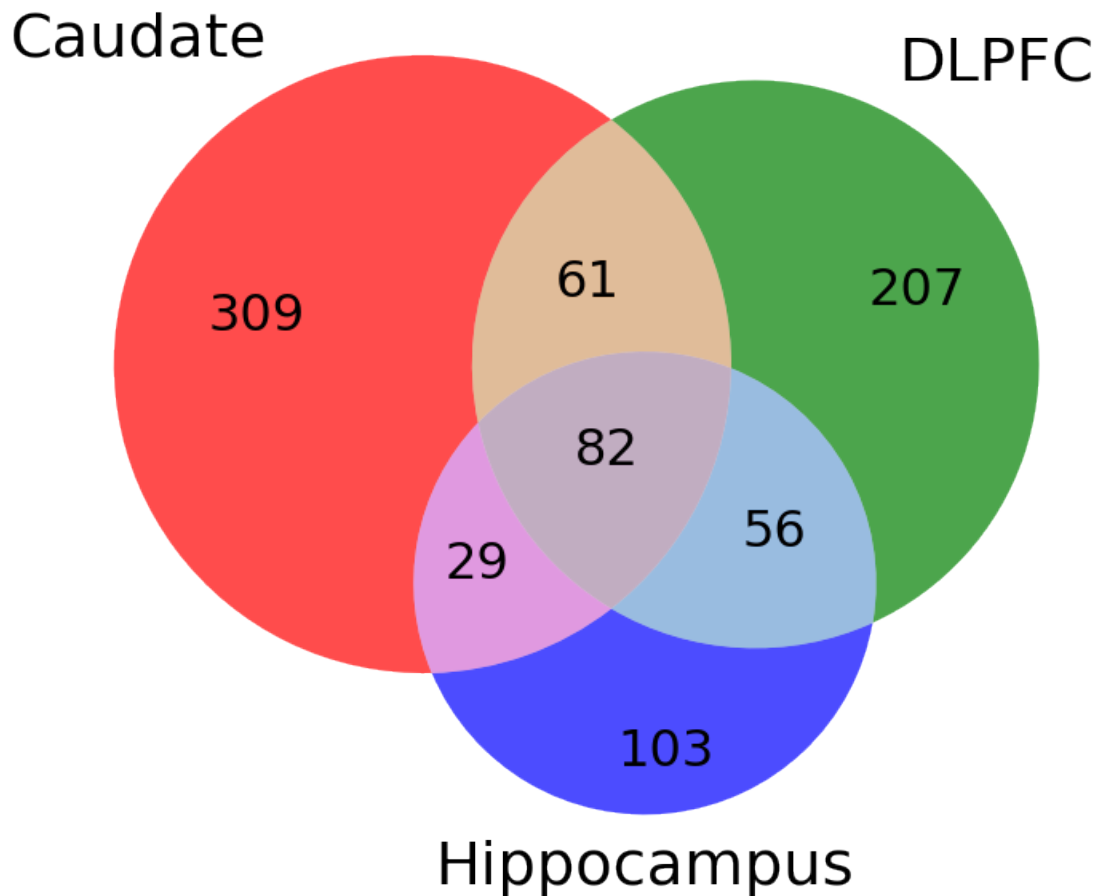
```
[14]:  overlapping_twas = new_caudate0.set_index('FILE').loc[set(caudate.FILE) &␣
       ↪set(dlpfc.FILE) & set(hippo.FILE), :]
       overlapping_twas.to_csv('overlapping_tissue_twasList.txt', sep='\t')
```

```
[15]:  overlapping_twas2 = caudate0.set_index('FILE').loc[set(caudate0.FILE) &␣
       ↪set(dlpfc0.FILE) & set(hippo0.FILE), :]
       overlapping_twas2.to_csv('overlapping_tissue_twasList_allFeatures.txt',␣
       ↪sep='\t')
       overlapping_twas2.shape
```

```
[15]:  (1921, 23)
```

```
[16]:  limiting_features(tissues, 'Caudate', 'Hippocampus')
       limiting_features(tissues, 'DLPFC', 'Caudate')
       limiting_features(tissues, 'DLPFC', 'Hippocampus')
```

```
Comparing Caudate with Hippocampus: 41.11%
Features in common: 111
Comparing DLPFC with Caudate: 29.73%
```

```
Features in common: 143
Comparing DLPFC with Hippocampus: 51.11%
Features in common: 138
```

```
[17]: caudate_only_genes = new_caudate0.set_index('FILE')\
                            .loc[tissues['Caudate'] - tissues['DLPFC'] -␣
      ↪tissues['Hippocampus'], :]
      caudate_only_genes.to_csv('caudate_only_twasList_genes.txt', sep='\t')
      print(caudate_only_genes.shape)
      caudate_only_genes.head()
```

```
(309, 46)
```

```
[17]:                       ID   CHR_TWAS         P0          P1        HSQ  \
      FILE
      ENSG00000242866     STRC         15   43599398    43618800.0  0.187533
      ENSG00000183379  SYNDIG1L        14   74405893    74426102.0  0.212943
      ENSG00000204435    CSNK2B         6   31665236    31670343.0  0.419359
      ENSG00000109323     MANBA         4  102631488   102760994.0  0.071179
      ENSG00000114446     IFT57         3  108160812   108222570.0  0.255847

                         BEST.GWAS.ID   BEST.GWAS.Z            EQTL.ID  \
      FILE
      ENSG00000242866   chr15:43782086:A:G         4.689   chr15:43631757:A:C
      ENSG00000183379   chr14:74416653:A:C        -4.167   chr14:74428292:C:T
      ENSG00000204435    chr6:31348749:T:C       -11.911    chr6:31617307:C:T
      ENSG00000109323   chr4:102267552:C:T         8.284   chr4:102856235:G:A
      ENSG00000114446   chr3:108058204:C:A         4.264   chr3:108215455:C:T

                         EQTL.R2    EQTL.Z  …       OR        SE          P  \
      FILE                                  …
      ENSG00000242866   0.143977  6.061215  …  0.95077  0.010767  2.750000e-06
      ENSG00000183379   0.062318  4.791923  …  1.04060  0.009544  3.090000e-05
      ENSG00000204435   0.207261 -6.948345  …  1.20840  0.015892  1.040000e-32
      ENSG00000109323  -0.001608  3.594929  …  0.85248  0.019267  1.190000e-16
      ENSG00000114446  -0.003130  4.006017  …  0.95581  0.010600  2.010000e-05

                        hg19chrc  hg38chrc     hg38pos  pgc2_a1_same_as_our_counted  \
      FILE
      ENSG00000242866     chr15     chr15    43782086                         False
      ENSG00000183379     chr14     chr14    74416653                         False
      ENSG00000204435      chr6      chr6    31348749                         False
      ENSG00000109323      chr4      chr4   102267552                         False
      ENSG00000114446      chr3      chr3   108058204                         False

                             rsid  is_index_snp  GWAS.SNP
      FILE
      ENSG00000242866   rs8034451         False     Other
```

```
ENSG00000183379   rs11159081              False     Other
ENSG00000204435    rs9265994              False  Risk SNP
ENSG00000109323   rs13107325               True  Risk SNP
ENSG00000114446    rs3762679              False     Other

[5 rows x 46 columns]
```

1.3   Z score comparison

```
[18]: import warnings
      warnings.filterwarnings('ignore')

      from plotnine import *
      from scipy.stats import spearmanr
```

```
[19]: def save_plot(p, fn):
          for ext in ['png', 'pdf', 'svg']:
              p.save(fn + '.' + ext)
```

1.3.1   DLPFC and Caudate

```
[20]: df1 = dlpfc0.merge(caudate0, on=['FILE'],
                         suffixes=['_dlpfc', '_caudate'],
                         how='outer').fillna(0)

      pp = (ggplot(df1, aes(x='TWAS.Z_caudate', y='TWAS.Z_dlpfc')) +␣
       ↪geom_point(alpha=0.5, size=1.25) +
            labs(x='TWAS Zscore Caudate', y='TWAS Zscore DLPFC') + theme_light() +
            theme(axis_text=element_text(size=14), axis_title=element_text(size=16,␣
       ↪face='bold')))
      pp
```
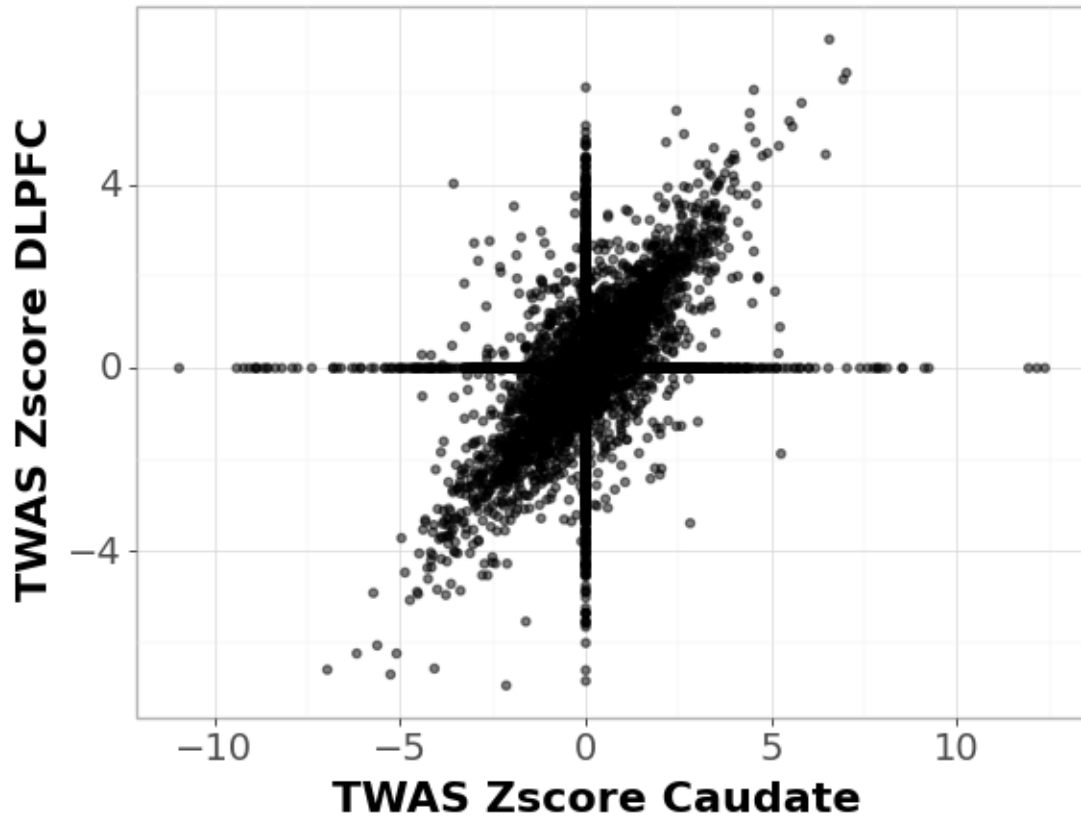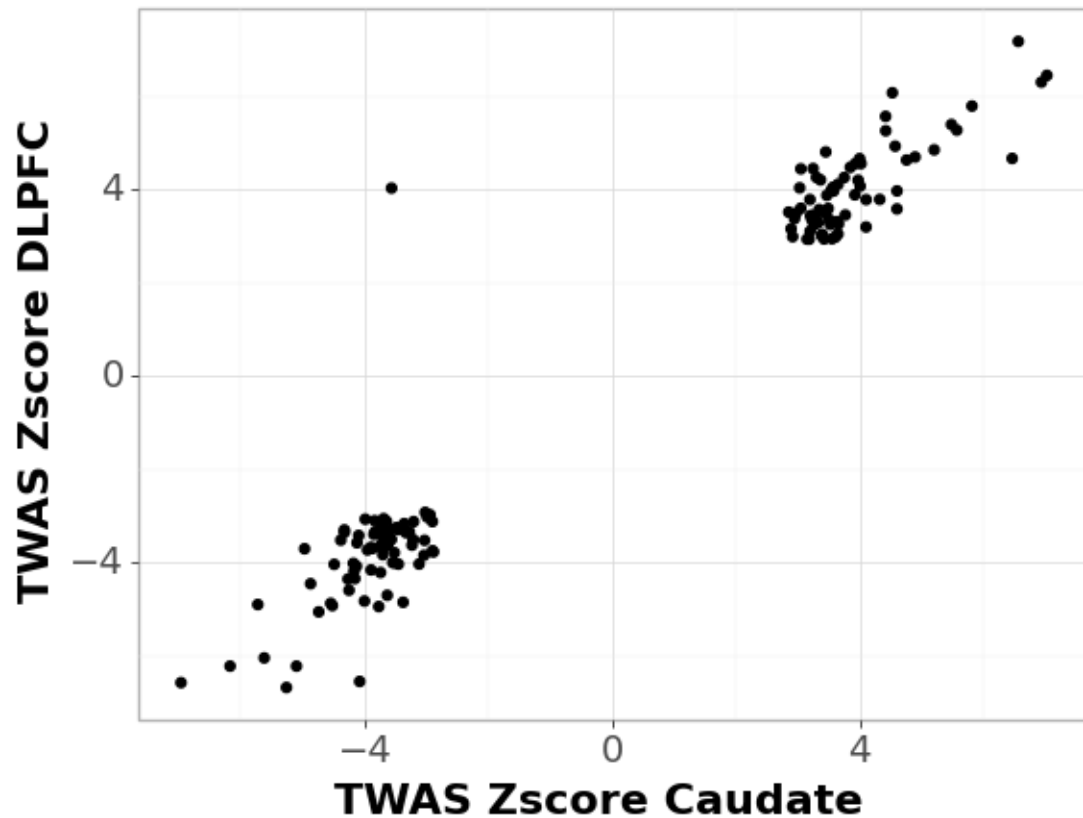
```
[20]: <ggplot: (8770683968121)>
```

```
[21]: save_plot(pp, 'twas_zscore_comparison_dlpfc_caudate_allFeatures')
```

```
[22]: df1 = dlpfc0.merge(caudate0, on=['FILE'],
                         suffixes=['_dlpfc', '_caudate'],
                         how='inner')
      spearmanr(df1['TWAS.Z_caudate'], df1['TWAS.Z_dlpfc'])
```

```
[22]: SpearmanrResult(correlation=0.7496074946623684, pvalue=0.0)
```

```
[23]: pp = (ggplot(df1, aes(x='TWAS.Z_caudate', y='TWAS.Z_dlpfc')) + geom_point() +
            labs(x='TWAS Zscore Caudate', y='TWAS Zscore DLPFC') + theme_light() +
            theme(axis_text=element_text(size=14), axis_title=element_text(size=16,␣
      ↪face='bold')))
      pp
```
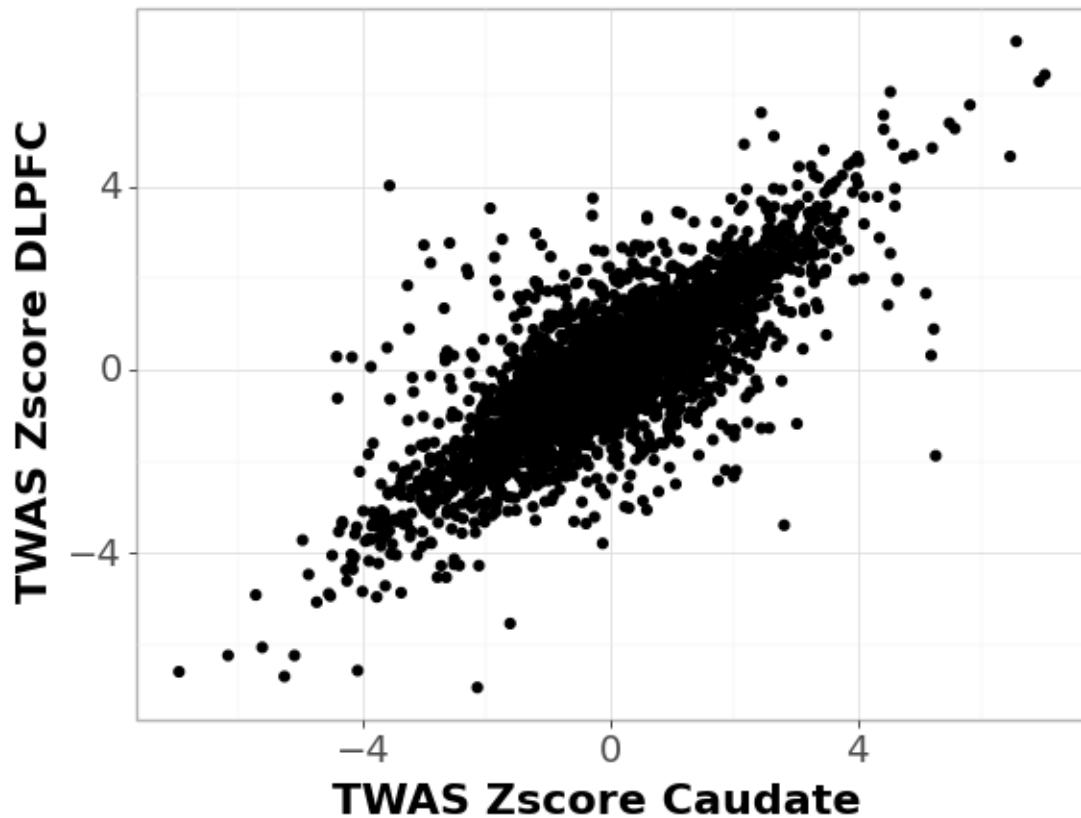
[23]: `<ggplot: (8770683994350)>`

[24]:
```
save_plot(pp, 'twas_zscore_comparison_dlpfc_caudate')
```

Significant TWAS

[25]:
```
df1 = dlpfc.merge(caudate, on=['FILE'], suffixes=['_dlpfc', '_caudate'])
spearmanr(df1['TWAS.Z_caudate'], df1['TWAS.Z_dlpfc'])
```

[25]: `SpearmanrResult(correlation=0.8894670802015994, pvalue=8.408451593959005e-50)`

[26]:
```
pp = (ggplot(df1, aes(x='TWAS.Z_caudate', y='TWAS.Z_dlpfc')) + geom_point() +
      labs(x='TWAS Zscore Caudate', y='TWAS Zscore DLPFC') + theme_light() +
      theme(axis_text=element_text(size=14), axis_title=element_text(size=16,
 ↪face='bold')))
pp
```

<ggplot: (8770681976330)>

```
save_plot(pp, 'twas_zscore_comparison_dlpfc_caudate_fdr05')
```

1.3.2 Hippocampus and Caudate

```
df2 = hippo0.merge(caudate0, on=['FILE'],
                   suffixes=['_hippo', '_caudate'],
                   how='outer').fillna(0)

pp = (ggplot(df2, aes(x='TWAS.Z_caudate', y='TWAS.Z_hippo')) + geom_point() +
      labs(x='TWAS Zscore Caudate', y='TWAS Zscore Hippocampus') +␣
 ↪theme_light() +
      theme(axis_text=element_text(size=14), axis_title=element_text(size=16,␣
 ↪face='bold')))
pp
```

11
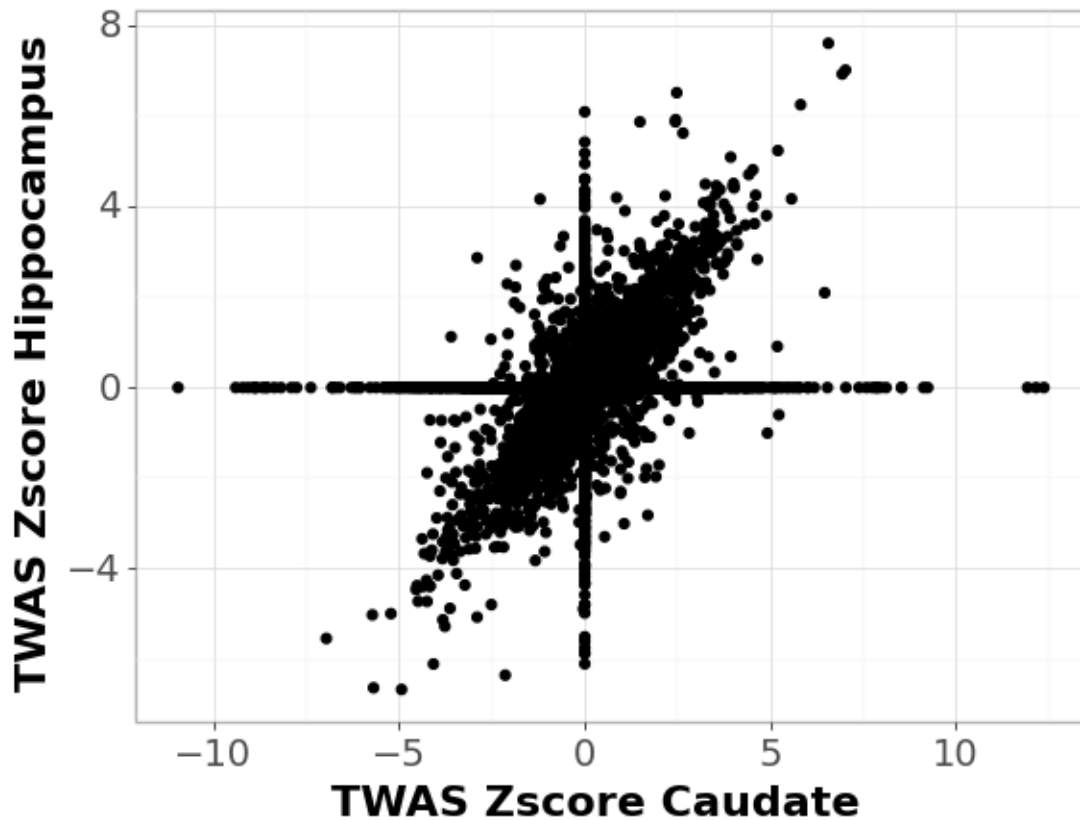
[28]: `<ggplot: (8770681919076)>`

[29]:
```
save_plot(pp, 'twas_zscore_comparison_hippo_caudate_allFeatures')
```

[30]:
```
df2 = hippo0.merge(caudate0, on=['FILE'],
                   suffixes=['_hippo', '_caudate'])
spearmanr(df2['TWAS.Z_caudate'], df2['TWAS.Z_hippo'])
```

[30]: `SpearmanrResult(correlation=0.7759226107710375, pvalue=0.0)`

[31]:
```
pp = (ggplot(df2, aes(x='TWAS.Z_caudate', y='TWAS.Z_hippo')) + geom_point() +
      labs(x='TWAS Zscore Caudate', y='TWAS Zscore Hippocampus') +
 →theme_light() +
      theme(axis_text=element_text(size=14), axis_title=element_text(size=16,
 →face='bold')))
pp
```
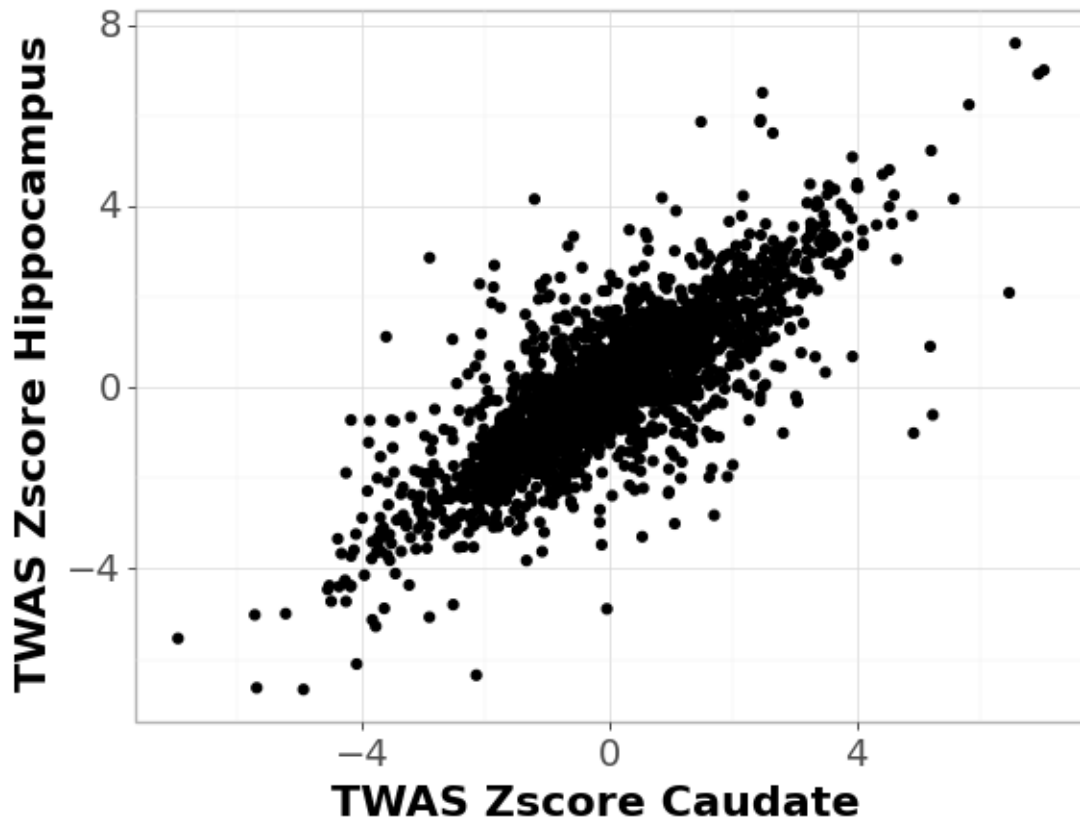
```
[31]: <ggplot: (8770681749342)>
```

```
[32]: save_plot(pp, 'twas_zscore_comparison_hippo_caudate')
```

Significant TWAS

```
[33]: df2 = hippo.merge(caudate, on=['FILE'],
                        suffixes=['_hippo', '_caudate'])
      spearmanr(df2['TWAS.Z_caudate'], df2['TWAS.Z_hippo'])
```

```
[33]: SpearmanrResult(correlation=0.8855124605124605, pvalue=4.438039606917797e-38)
```

```
[34]: pp = (ggplot(df2, aes(x='TWAS.Z_caudate', y='TWAS.Z_hippo')) + geom_point() +
            labs(x='TWAS Zscore Caudate', y='TWAS Zscore Hippocampus') +␣
      ↪theme_light() +
            theme(axis_text=element_text(size=14), axis_title=element_text(size=16,␣
      ↪face='bold')))
      pp
```
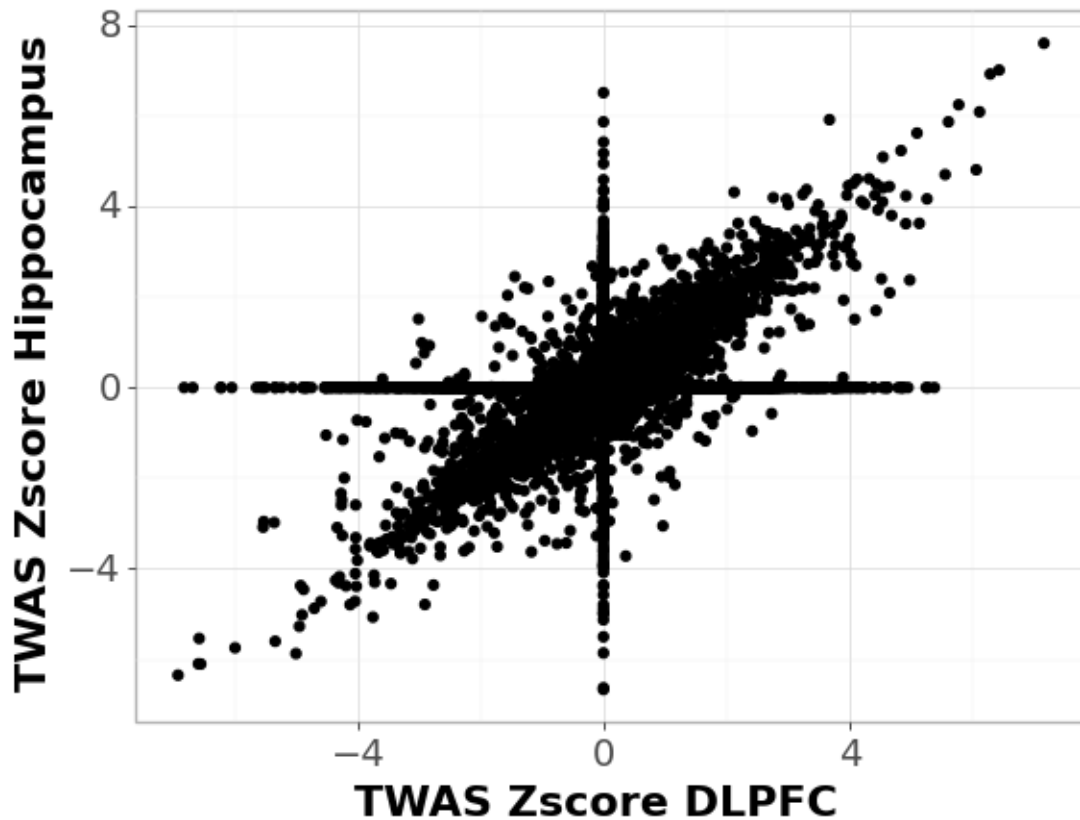
`<ggplot: (8770682052837)>`
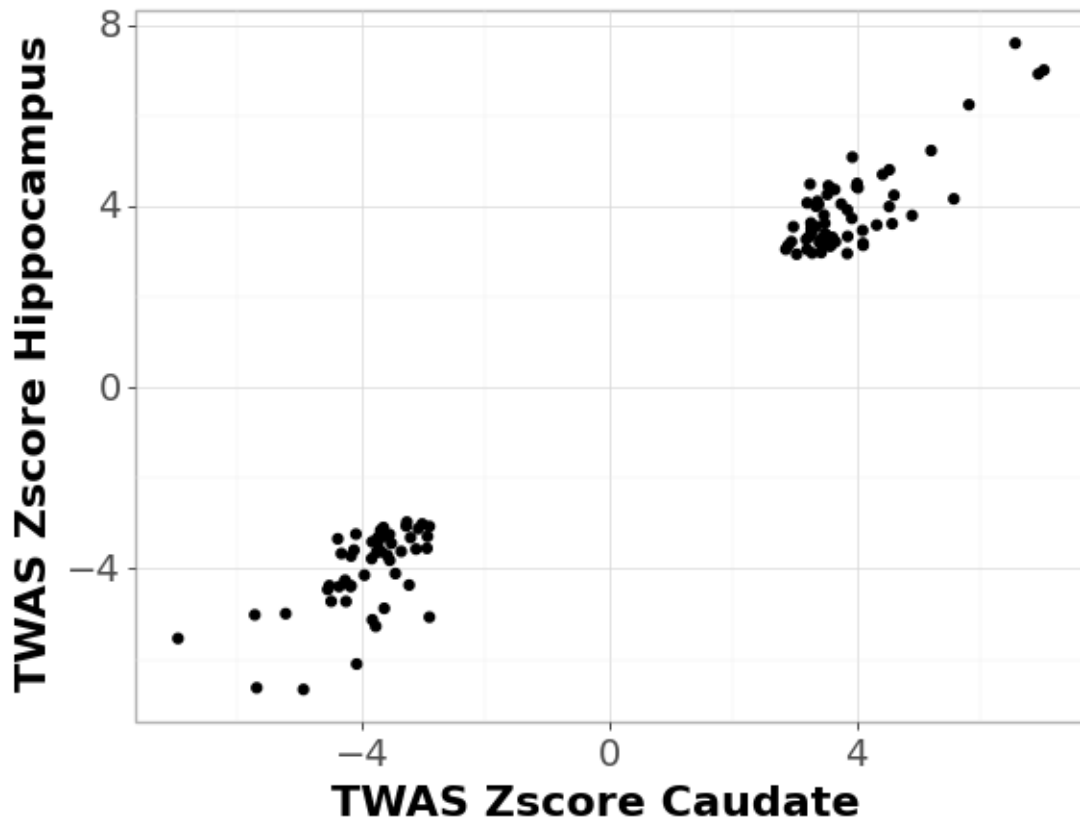
```
save_plot(pp, 'twas_zscore_comparison_hippo_caudate_fdr05')
```

1.3.3 DLPFC and Hippocampus

```
df3 = hippo0.merge(dlpfc0, on=['FILE'], how='outer',
                   suffixes=['_hippo', '_dlpfc']).fillna(0)
pp = (ggplot(df3, aes(x='TWAS.Z_dlpfc', y='TWAS.Z_hippo')) + geom_point() +
      labs(x='TWAS Zscore DLPFC', y='TWAS Zscore Hippocampus') + theme_light() +
      theme(axis_text=element_text(size=14), axis_title=element_text(size=16,
  ↪face='bold')))
pp
```

[36]: `<ggplot: (8770682253719)>`

[37]:
```
save_plot(pp, 'twas_zscore_comparison_hippo_dlpfc_allFeatures')
```

[38]:
```
df3 = hippo0.merge(dlpfc0, on=['FILE'],
                   suffixes=['_hippo', '_dlpfc'])
spearmanr(df3['TWAS.Z_dlpfc'], df3['TWAS.Z_hippo'])
```

[38]: `SpearmanrResult(correlation=0.846982546568797, pvalue=0.0)`

[39]:
```
pp = (ggplot(df3, aes(x='TWAS.Z_dlpfc', y='TWAS.Z_hippo')) + geom_point() +
      labs(x='TWAS Zscore DLPFC', y='TWAS Zscore Hippocampus') + theme_light() +
      theme(axis_text=element_text(size=14), axis_title=element_text(size=16,␣
 ↪face='bold')))
pp
```
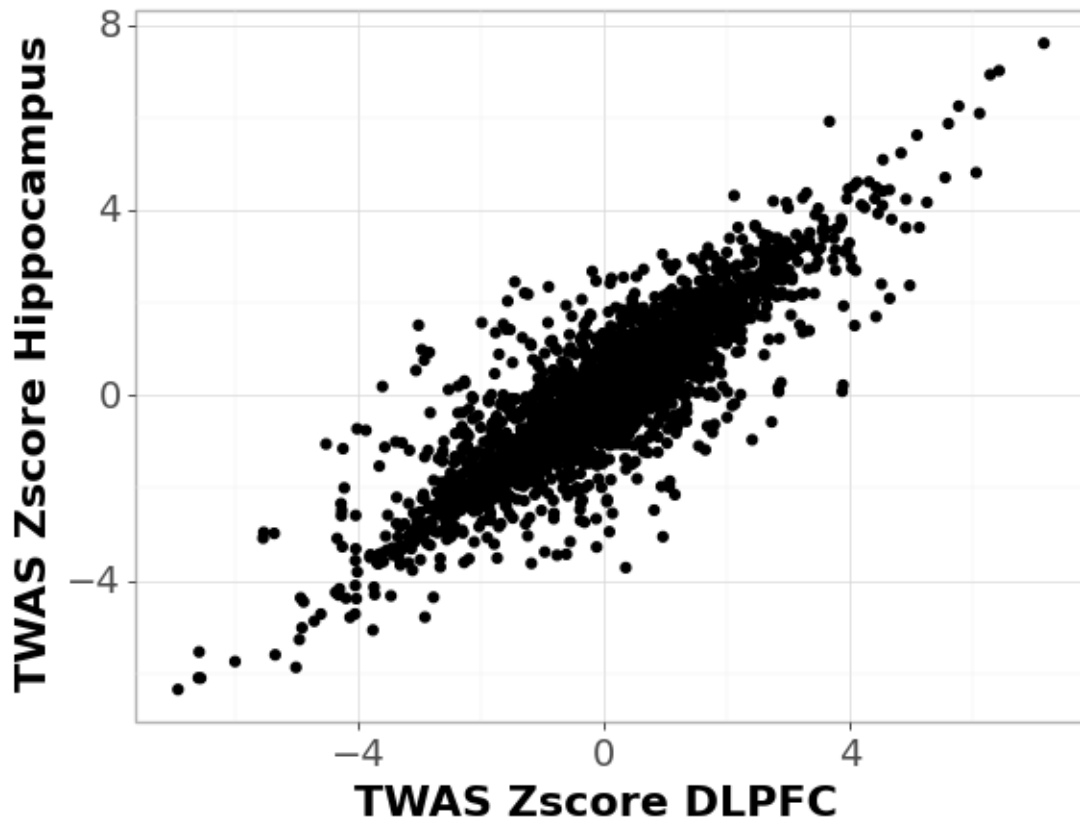
[39]: `<ggplot: (8770681717514)>`

[40]:
```
save_plot(pp, 'twas_zscore_comparison_hippo_dlpfc')
```

Significant TWAS

[41]:
```
df3 = dlpfc.merge(hippo, on=['FILE'], suffixes=['_dlpfc', '_hippo'])
spearmanr(df3['TWAS.Z_hippo'], df3['TWAS.Z_dlpfc'])
```

[41]: `SpearmanrResult(correlation=0.907421085257396, pvalue=4.667456229995493e-53)`

[42]:
```
pp = (ggplot(df3, aes(x='TWAS.Z_dlpfc', y='TWAS.Z_hippo')) + geom_point() +
      labs(x='TWAS Zscore DLPFC', y='TWAS Zscore Hippocampus') + theme_light() +
      theme(axis_text=element_text(size=14), axis_title=element_text(size=16,
   ↪face='bold')))
pp
```
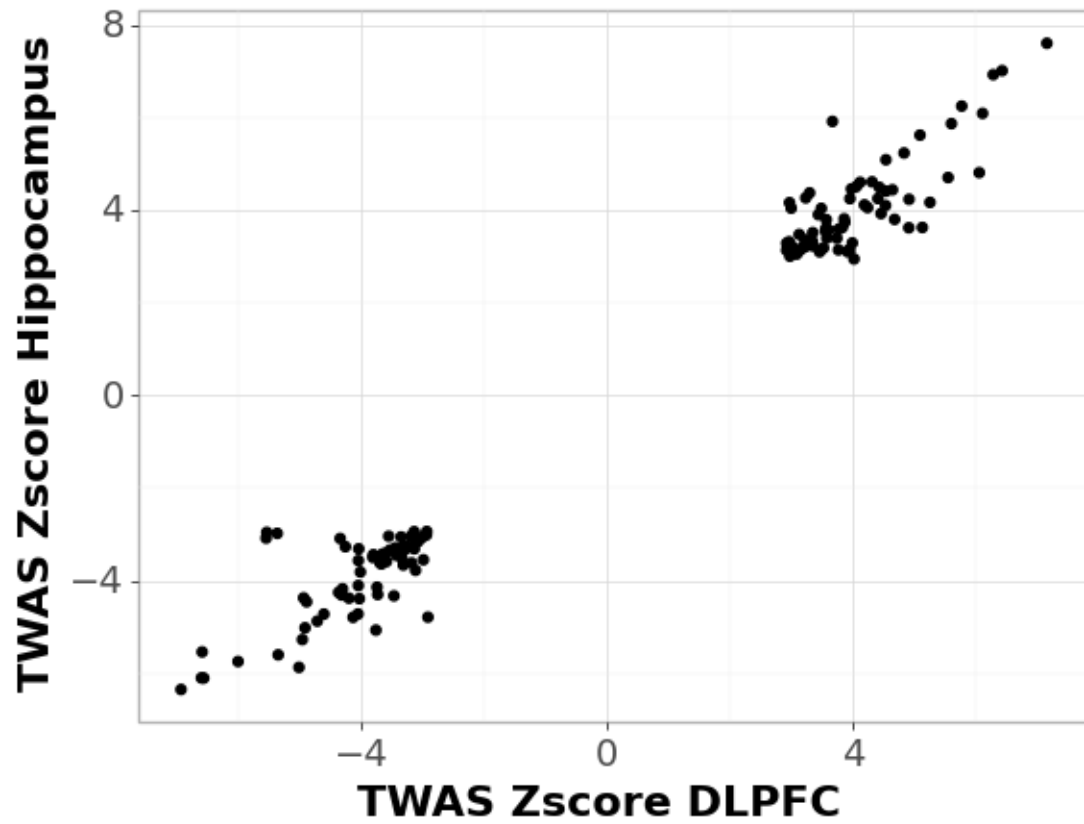
[42]: <ggplot: (8770682041881)>

[43]: `save_plot(pp, 'twas_zscore_comparison_hippo_dlpfc_fdr05')`