

main

December 10, 2020

## 1 Extract unique female specific SZ-associated genes

```
[1]: import functools
import numpy as np
import pandas as pd
from scipy.stats import mannwhitneyu
from statsmodels.stats.multitest import fdrcorrection

[2]: @functools.lru_cache()
def get_res_df():
    return pd.read_csv('.././.././.././interaction_sex_sz/cmc_dlpfc/_m/genes/
↳residualized_expression.tsv', sep='\t').T

@functools.lru_cache()
def get_pheno_df():
    return pd.read_csv('/ceph/users/jbenja13/projects/sex_sz_ria/input/
↳commonMind/phenotypes/combine_files/_m/CMC_phenotypes_all.csv').
↳set_index("RNAseq:Sample_RNA_ID")

@functools.lru_cache()
def get_res_pheno_df():
    return pd.merge(get_pheno_df(), get_res_df(), left_index=True,
↳right_index=True)

[3]: def get_de(feature):
    f = pd.read_csv('.././.././female_analysis/_m/%s/diffExpr_szVctl_full.txt' %
↳feature, sep='\t')\
        .rename(columns={'gene_id': 'gencodeID'})
    f['ensemblID'] = f.gencodeID.str.replace("\\.*", "")
    f.set_index('ensemblID', inplace=True)
    m = pd.read_csv('.././.././male_analysis/_m/%s/diffExpr_szVctl_full.txt' %
↳feature, sep='\t')\
        .rename(columns={'gene_id': 'gencodeID'})
    m['ensemblID'] = m.gencodeID.str.replace("\\.*", "")
    m.set_index('ensemblID', inplace=True)
```

```

    a = pd.read_csv('/ceph/projects/v3_phase3_paper/inputs/cmc/_m/
    ↪CMC_MSSM-Penn-Pitt_DLPFC_mRNA_IlluminaHiSeq2500_gene-adjustedSVA-differentialExpression-inc
    ↪tsv', sep='\t')\
        .rename(columns={"MAPPED_genes": 'gene_name'}).set_index('genes')
    return f, m, a

def get_unique(x, y, thres=0.05):
    return x.merge(pd.DataFrame(index = list(set(x[(x['adj.P.Val'] <= thres)].
    ↪index) -
                                                set(y[(y['adj.P.Val'] <= thres)].
    ↪index)))),
                    left_index=True, right_index=True)

def subset_sz_female():
    df = get_res_pheno_df()
    ctl = df[(df['Dx'] == 'Control') & (df['Sex'] == 'XX')].copy()
    sz = df[(df['Dx'] == 'SCZ') & (df['Sex'] == 'XX')].copy()
    return ctl, sz

def add_pvals_adjustPval(df):
    ctl, sz = subset_sz_female()
    f_pval = []
    for gene_id in df.Feature:
        stat, pval = mannwhitneyu(ctl[gene_id], sz[gene_id])
        f_pval.append(pval)
    fdr_f = fdr correction(f_pval)
    return pd.concat([df.set_index('Feature'),
                      pd.DataFrame({'Female_Pval': f_pval,
                                   'Female_FDR': fdr_f[1]},
                                   index=df.Feature)], axis=1)

```

## 1.1 Genes

```

[4]: f, m, a = get_de('genes')
    m['Feature'] = m.gencodeID
    m['ensemblID'] = m.index
    #genes = get_unique(get_unique(f, m), a)
    genes = get_unique(m, f).rename(columns={'chromosome_name': 'Chrom',
                                             'hgnc_symbol': 'Symbol'})
    genes = genes[['Feature', 'gencodeID', 'Symbol', 'ensemblID',
                  'Chrom', 'logFC', 't', 'adj.P.Val']].sort_values('adj.P.Val')
    genes.Chrom = 'chr'+genes.Chrom
    genes.shape

```

[4]: (495, 8)

```
[5]: genes = add_pvals_adjustPval(genes)
genes = genes[~(genes['Female_Pval'] <= 0.05)] ## Stringents
genes['Type'] = 'gene'
genes.shape
```

[5]: (172, 10)

```
[6]: genes.head(2)
```

```
[6]:
```

	gencodeID	Symbol	ensemblID	Chrom	\
Feature					
ENSG00000119411.10	ENSG00000119411.10	BSPRY	ENSG00000119411	chr9	
ENSG00000159871.14	ENSG00000159871.14	LYPD5	ENSG00000159871	chr19	

  

	logFC	t	adj.P.Val	Female_Pval	Female_FDR	\
Feature						
ENSG00000119411.10	0.280110	5.861029	0.000028	0.439030	0.454209	
ENSG00000159871.14	0.213822	5.885126	0.000028	0.362537	0.391824	

  

	Type
Feature	
ENSG00000119411.10	gene
ENSG00000159871.14	gene

## 1.2 DE summary

### 1.2.1 DE (feature)

```
[7]: gg = len(set(genes['gencodeID']))

print("\nGene:\t\t%d" % (gg))
```

Gene: 172

```
[8]: genes.to_csv('male_specific_DE_genes.txt', sep='\t', index=True, header=True)
```

## 1.3 Number of DEGs on allosomes

```
[9]: genes[(genes['Chrom'].isin(['chrX', 'chrY']))].groupby(['Type', 'Chrom']).size()
```

```
[9]: Type  Chrom
gene  chrX      2
dtype: int64
```

[ ]: