

# main

March 26, 2023

## 1 Comparison with other datasets

```
[1]: import functools, session_info
import numpy as np
import pandas as pd
from plotnine import *
from pyhere import here
from scipy.stats import binom_test, fisher_exact, linregress

from warnings import filterwarnings
from matplotlib.cbook import mplDeprecation
filterwarnings('ignore', category=mplDeprecation)
filterwarnings('ignore', category=UserWarning, module='plotnine.*')
filterwarnings('ignore', category=DeprecationWarning, module='plotnine.*')
```

```
[2]: import matplotlib.pyplot as plt
import matplotlib as mpl
```

### 1.1 BrainSeq functions

```
[3]: @functools.lru_cache()
def get_biomart():
    return pd.read_csv('../_h/biomart.csv', index_col=0)

@functools.lru_cache()
def get_deg(tissue):
    fname = f'../_h/{tissue}/_m/genes/diffExpr_maleVfemale_full.txt'
    dft = pd.read_csv(fname, sep='\t', index_col=0)
    dft['Feature'] = dft.index
    dft['Direction'] = np.sign(dft['t'])
    dft['ensemblID'] = dft.gencodeID.str.replace('\\.*', '', regex=True)
    return dft[['Feature', 'ensemblID', 'Symbol', 'adj.P.Val', 'logFC', 't',
    ↪ 'Direction']]

@functools.lru_cache()
```

```

def get_deg_sig(tissue, fdr):
    dft = get_deg(tissue)
    return dft[(dft['adj.P.Val'] < fdr)]

@functools.lru_cache()
def merge_dataframes(tissue1, tissue2, fdr=1):
    return get_deg_sig(tissue1, fdr)\
        .merge(get_deg_sig(tissue2, fdr), on='Feature',
               suffixes=[f'_{tissue1}', f'_{tissue2}'])

@functools.lru_cache()
def brainseq_shared_degs():
    fn = "../summary_table/_m/differential_expression_analysis_4features_sex.
    ↪txt.gz"
    df0 = pd.read_csv(fn, sep='\t')
    df = df0.loc[(df0["Type"] == "Gene") & (df0["adj.P.Val"] < 0.05),
                 ["ensemblID", "Tissue"]].copy()
    df["DE"] = 1
    return df.pivot(index="ensemblID", columns="Tissue", values="DE").fillna(0)

```

## 1.2 Trabzuni comparison

```

[4]: @functools.lru_cache()
def get_symbols(tissue, fdr):
    return get_deg_sig(tissue, fdr)[['Symbol']]

def tissue_annotation(tissue):
    return {'dlpfc': 'DLPFC', 'hippocampus': 'Hippocampus',
           'caudate': 'Caudate', 'cmc_dlpfc': 'CMC DLPFC'}[tissue]

```

```

[5]: def heatmap(data, row_labels, col_labels, ax=None,
                cbar_kw={}, cbarlabel="", **kwargs):
    """
    Create a heatmap from a numpy array and two lists of labels.

    Parameters
    -----
    data
        A 2D numpy array of shape (N, M).
    row_labels
        A list or array of length N with the labels for the rows.
    col_labels
        A list or array of length M with the labels for the columns.
    """

```

```

ax
    A `matplotlib.axes.Axes` instance to which the heatmap is plotted. If
    not provided, use current axes or create a new one. Optional.
cbar_kw
    A dictionary with arguments to `matplotlib.figure.colorbar`. Optional.
cbarlabel
    The label for the colorbar. Optional.
**kwargs
    All other arguments are forwarded to `imshow`.
"""

if not ax:
    ax = plt.gca()

# Plot the heatmap
im = ax.imshow(data, **kwargs)

# Create colorbar
cbar = ax.figure.colorbar(im, ax=ax, **cbar_kw)
cbar.ax.set_ylabel(cbarlabel, rotation=-90, va="bottom")

# We want to show all ticks...
ax.set_xticks(np.arange(data.shape[1]))
ax.set_yticks(np.arange(data.shape[0]))
# ... and label them with the respective list entries.
ax.set_xticklabels(col_labels)
ax.set_yticklabels(row_labels)

# Let the horizontal axes labeling appear on top.
ax.tick_params(top=True, bottom=False, labelsize=18,
               labeltop=True, labelbottom=False)

# Rotate the tick labels and set their alignment.
plt.setp(ax.get_xticklabels(), rotation=-30, ha="right",
         rotation_mode="anchor")

# Turn spines off and create white grid.
for edge, spine in ax.spines.items():
    spine.set_visible(False)

ax.set_xticks(np.arange(data.shape[1]+1)-.5, minor=True)
ax.set_yticks(np.arange(data.shape[0]+1)-.5, minor=True)
ax.grid(which="minor", color="w", linestyle='-', linewidth=3)
ax.tick_params(which="minor", bottom=False, left=False)

return im, cbar

```

```

def annotate_heatmap(im, data=None, valfmt="{x:.2f}",
                    textcolors=("black", "white"),
                    threshold=None, **textkw):
    """
    A function to annotate a heatmap.

    Parameters
    -----
    im
        The AxesImage to be labeled.
    data
        Data used to annotate. If None, the image's data is used. Optional.
    valfmt
        The format of the annotations inside the heatmap. This should either
        use the string format method, e.g. "$ {x:.2f}", or be a
        `matplotlib.ticker.Formatter`. Optional.
    textcolors
        A pair of colors. The first is used for values below a threshold,
        the second for those above. Optional.
    threshold
        Value in data units according to which the colors from textcolors are
        applied. If None (the default) uses the middle of the colormap as
        separation. Optional.
    **kwargs
        All other arguments are forwarded to each call to `text` used to create
        the text labels.
    """

    if not isinstance(data, (list, np.ndarray)):
        data = im.get_array()

    # Normalize the threshold to the images color range.
    if threshold is not None:
        threshold = im.norm(threshold)
    else:
        threshold = im.norm(data.max())/2.

    # Set default alignment to center, but allow it to be
    # overwritten by textkw.
    kw = dict(horizontalalignment="center",
                verticalalignment="center")
    kw.update(textkw)

    # Get the formatter in case a string is supplied
    if isinstance(valfmt, str):
        valfmt = mpl.ticker.StrMethodFormatter(valfmt)

```

```

# Loop over the data and create a `Text` for each "pixel".
# Change the text's color depending on the data.
texts = []
for i in range(data.shape[0]):
    for j in range(data.shape[1]):
        kw.update(color=textcolors[int(im.norm(data[i, j]) > threshold)])
        text = im.axes.text(j, i, valfmt(data[i, j], None), **kw)
        texts.append(text)

return texts

```

```

[6]: config_trabzuni = {
    'genes': here('input/public_results/extract_geneLists/_m/
↳ trabzuni_sex_diff_genes_logFC.csv'),
    'annot': here('input/public_results/_m/trabzuni/trabzuni_annotation.txt'),
}

zz = dict()
for tissue in ['caudate', 'dlpfc', 'hippocampus']:
    ww = []; yy = []
    fdr = 0.05
    trabzuni = pd.read_csv(config_trabzuni['genes'], index_col=0)\
        .drop(['average ALL'], axis=1)
    annot = pd.read_csv(config_trabzuni['annot'], sep='\t', index_col=0)
    for col in trabzuni.drop('Chromosome', axis=1).columns:
        trabzuni.loc[trabzuni[col] != 0, col] = 1
        tot = len(set(trabzuni[(trabzuni[col] != 0)].index))
        overlap = len(set(get_symbols(tissue, fdr).Symbol) &
            set(trabzuni[(trabzuni[col] != 0)].index))
        xx = overlap / tot
        ww.append(xx)
        yy.append(annot.to_dict()['TISSUE'][col.split(' ')[0]])
        print("There is %d (%.1f%%) overlap between %s and %s %s!" %
            (overlap, xx*100, tissue_annotation(tissue),
            annot.to_dict()['TISSUE'][col.split(' ')[0]],
            col.split(" ")[1]))
    print("")
    zz[tissue_annotation(tissue)] = ww

```

```

There is 23 (71.9%) overlap between Caudate and cerebellum (N=128)!
There is 17 (77.3%) overlap between Caudate and frontal cortex (N=122)!
There is 18 (85.7%) overlap between Caudate and hippocampus (N=114)!
There is 7 (87.5%) overlap between Caudate and hypothalamus (N=13)!
There is 20 (90.9%) overlap between Caudate and medulla (N=109)!
There is 16 (84.2%) overlap between Caudate and occipital cortex (N=125)!
There is 14 (77.8%) overlap between Caudate and putamen (N=121)!

```

There is 17 (89.5%) overlap between Caudate and substantia nigra (N=96)!

There is 19 (82.6%) overlap between Caudate and temporal cortex (N=114)!

There is 21 (77.8%) overlap between Caudate and thalamus (N=107)!

There is 20 (87.0%) overlap between Caudate and white matter (N=120)!

There is 20 (62.5%) overlap between DLPFC and cerebellum (N=128)!

There is 17 (77.3%) overlap between DLPFC and frontal cortex (N=122)!

There is 18 (85.7%) overlap between DLPFC and hippocampus (N=114)!

There is 7 (87.5%) overlap between DLPFC and hypothalamus (N=13)!

There is 20 (90.9%) overlap between DLPFC and medulla (N=109)!

There is 16 (84.2%) overlap between DLPFC and occipital cortex (N=125)!

There is 14 (77.8%) overlap between DLPFC and putamen (N=121)!

There is 17 (89.5%) overlap between DLPFC and substantia nigra (N=96)!

There is 19 (82.6%) overlap between DLPFC and temporal cortex (N=114)!

There is 19 (70.4%) overlap between DLPFC and thalamus (N=107)!

There is 19 (82.6%) overlap between DLPFC and white matter (N=120)!

There is 22 (68.8%) overlap between Hippocampus and cerebellum (N=128)!

There is 17 (77.3%) overlap between Hippocampus and frontal cortex (N=122)!

There is 18 (85.7%) overlap between Hippocampus and hippocampus (N=114)!

There is 7 (87.5%) overlap between Hippocampus and hypothalamus (N=13)!

There is 20 (90.9%) overlap between Hippocampus and medulla (N=109)!

There is 16 (84.2%) overlap between Hippocampus and occipital cortex (N=125)!

There is 14 (77.8%) overlap between Hippocampus and putamen (N=121)!

There is 17 (89.5%) overlap between Hippocampus and substantia nigra (N=96)!

There is 19 (82.6%) overlap between Hippocampus and temporal cortex (N=114)!

There is 20 (74.1%) overlap between Hippocampus and thalamus (N=107)!

There is 20 (87.0%) overlap between Hippocampus and white matter (N=120)!

```
[7]: dfz = pd.DataFrame.from_dict(zz)
dfz['Tissue'] = trabzuni.drop('Chromosome', axis=1).columns
dfz = dfz.set_index('Tissue')
dfz
```

```
[7]:
```

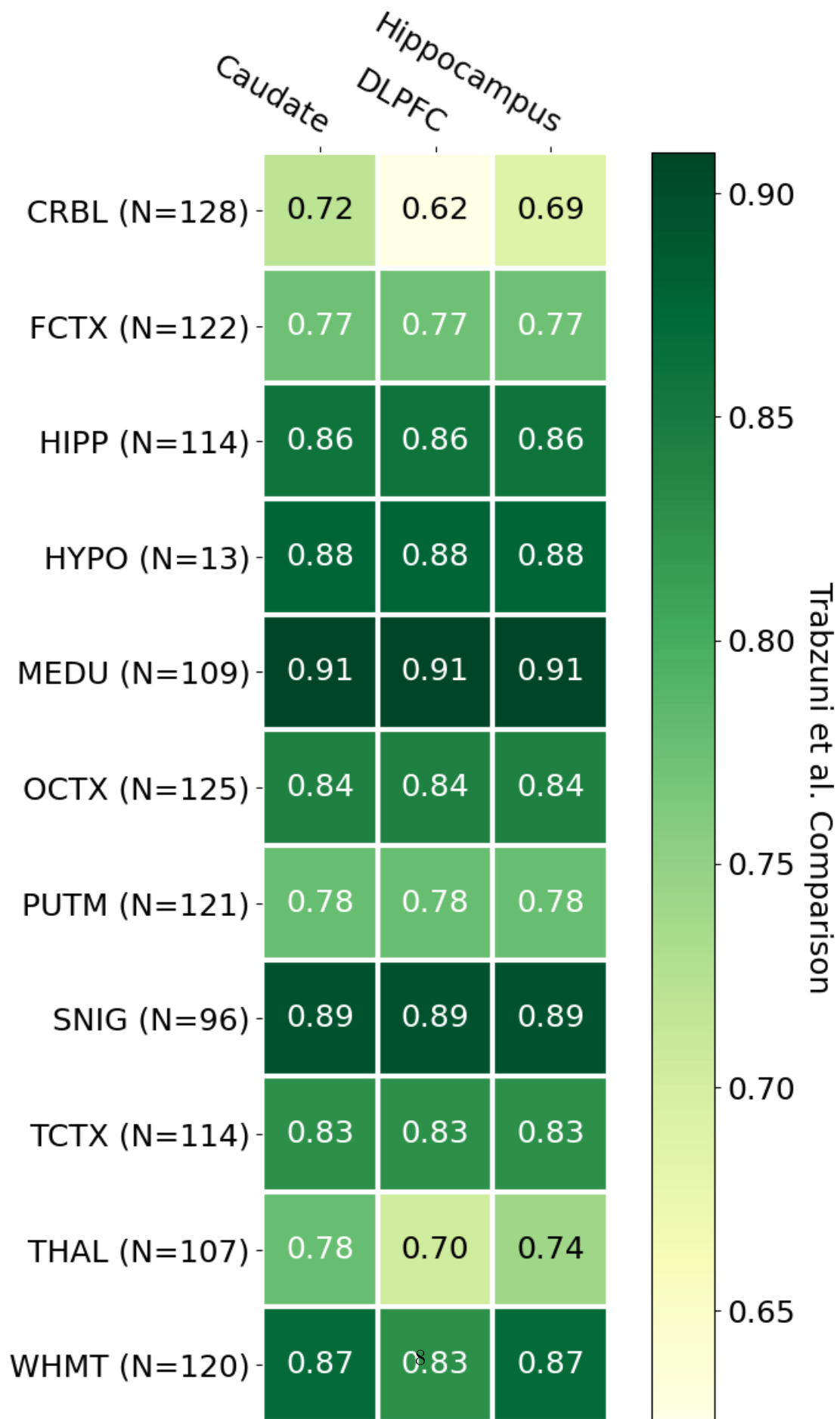
	Caudate	DLPFC	Hippocampus
Tissue			
CRBL (N=128)	0.718750	0.625000	0.687500
FCTX (N=122)	0.772727	0.772727	0.772727
HIPP (N=114)	0.857143	0.857143	0.857143
HYP0 (N=13)	0.875000	0.875000	0.875000
MEDU (N=109)	0.909091	0.909091	0.909091
OCTX (N=125)	0.842105	0.842105	0.842105
PUTM (N=121)	0.777778	0.777778	0.777778
SNIG (N=96)	0.894737	0.894737	0.894737
TCTX (N=114)	0.826087	0.826087	0.826087
THAL (N=107)	0.777778	0.703704	0.740741

WHMT (N=120) 0.869565 0.826087 0.869565

```
[8]: dfz.to_csv('brainseq_overlap_trabzuni_fraction.txt', sep='\t')
```

```
[9]: %matplotlib inline
fig, ax = plt.subplots(figsize=(8, 12))
plt.rcParams.update({'font.size': 18})
im, cbar = heatmap(dfz, dfz.index, dfz.columns, ax=ax,
                  cmap="YlGn", cbarlabel="Trabzuni et al. Comparison")
texts = annotate_heatmap(im, valfmt="{x:.2f}")

fig.tight_layout()
plt.savefig('brainseq_overlap_trabzuni_heatmap.png')
plt.savefig('brainseq_overlap_trabzuni_heatmap.pdf')
plt.savefig('brainseq_overlap_trabzuni_heatmap.svg')
plt.show()
```





```
[10]: df = brainseq_shared_degs()
df.to_csv('brainseq_deg_across_tissues_comparison.csv')
df.head(2)
```

```
[10]: Tissue          Caudate  DLPFC  Hippocampus
ensemblID
ENSG000000002586      1.0    1.0          1.0
ENSG000000003137      1.0    0.0          0.0
```

```
[11]: tt0 = get_biomart()\
      .merge(df, left_on='ensembl_gene_id', right_index=True)\
      .merge(trabzuni, left_on='external_gene_name', right_index=True,\
      how='right')\
      .set_index('external_gene_name')

tt1 = tt0[['Caudate', 'DLPFC', 'Hippocampus']].fillna(0)
tt2 = tt0[['CRBL (N=128)', 'FCTX (N=122)', 'HIPP (N=114)', 'HYPO (N=13)',
          'MEDU (N=109)', 'OCTX (N=125)', 'PUTM (N=121)', 'SNIG (N=96)',
          'TCTX (N=114)', 'THAL (N=107)', 'WHMT (N=120)']].fillna(0)
tt = pd.concat([tt1, tt2], axis=1)
tt.to_csv('brainseq_overlap_trabzuni_genes.csv')
tt.head(2)
```

```
[11]:          Caudate  DLPFC  Hippocampus  CRBL (N=128)  FCTX (N=122) \
external_gene_name
ABCA6           0.0    0.0           0.0           0.0           1.0
ALG12           0.0    0.0           0.0           0.0           0.0

          HIPP (N=114)  HYPO (N=13)  MEDU (N=109)  OCTX (N=125) \
external_gene_name
ABCA6           0.0           0.0           0.0           0.0
ALG12           0.0           0.0           0.0           0.0

          PUTM (N=121)  SNIG (N=96)  TCTX (N=114)  THAL (N=107) \
external_gene_name
ABCA6           0.0           0.0           0.0           0.0
ALG12           0.0           0.0           0.0           1.0

          WHMT (N=120)
external_gene_name
ABCA6           0.0
ALG12           0.0
```

### 1.3 Mayne comparison

```
[12]: mayne_file = here('input/public_results/extract_geneLists/_m/
    ↪mayne_sex_diff_genes.csv')
mayne = pd.read_csv(mayne_file)
mayne.groupby('Version').size()
```

```
[12]: Version
all      16
autosomes 215
noY      244
dtype: int64
```

```
[13]: zz = dict()
for tissue in ['caudate', 'dlpfc', 'hippocampus']:
    ww = []; yy = []
    fdr = 0.05
    mayne = pd.read_csv(mayne_file)
    mayne['Mayne'] = 1
    for ver in ['all', 'autosomes', 'noY']:
        mm = mayne.loc[(mayne['Version'] == ver), ['Ensembl', 'Mayne']]
        tot = len(set(mm.Ensembl))
        overlap = len(set(get_deg_sig(tissue, fdr).ensemblID) &
                        set(mm.Ensembl))
        xx = overlap / tot
        ww.append(xx)
        print("There is %d (%.1f%%) overlap between %s and %s from Mayne!" %
              (overlap, xx* 100, tissue_annotation(tissue), ver))
    print("")
    zz[tissue_annotation(tissue)] = ww
```

There is 13 (86.7%) overlap between Caudate and all from Mayne!  
There is 14 (6.5%) overlap between Caudate and autosomes from Mayne!  
There is 28 (11.5%) overlap between Caudate and noY from Mayne!

There is 13 (86.7%) overlap between DLPFC and all from Mayne!  
There is 2 (0.9%) overlap between DLPFC and autosomes from Mayne!  
There is 10 (4.1%) overlap between DLPFC and noY from Mayne!

There is 13 (86.7%) overlap between Hippocampus and all from Mayne!  
There is 0 (0.0%) overlap between Hippocampus and autosomes from Mayne!  
There is 9 (3.7%) overlap between Hippocampus and noY from Mayne!

```
[14]: dfm = pd.DataFrame.from_dict(zz)
dfm['Method'] = mayne.Version.unique()
dfm = dfm.set_index('Method')
```

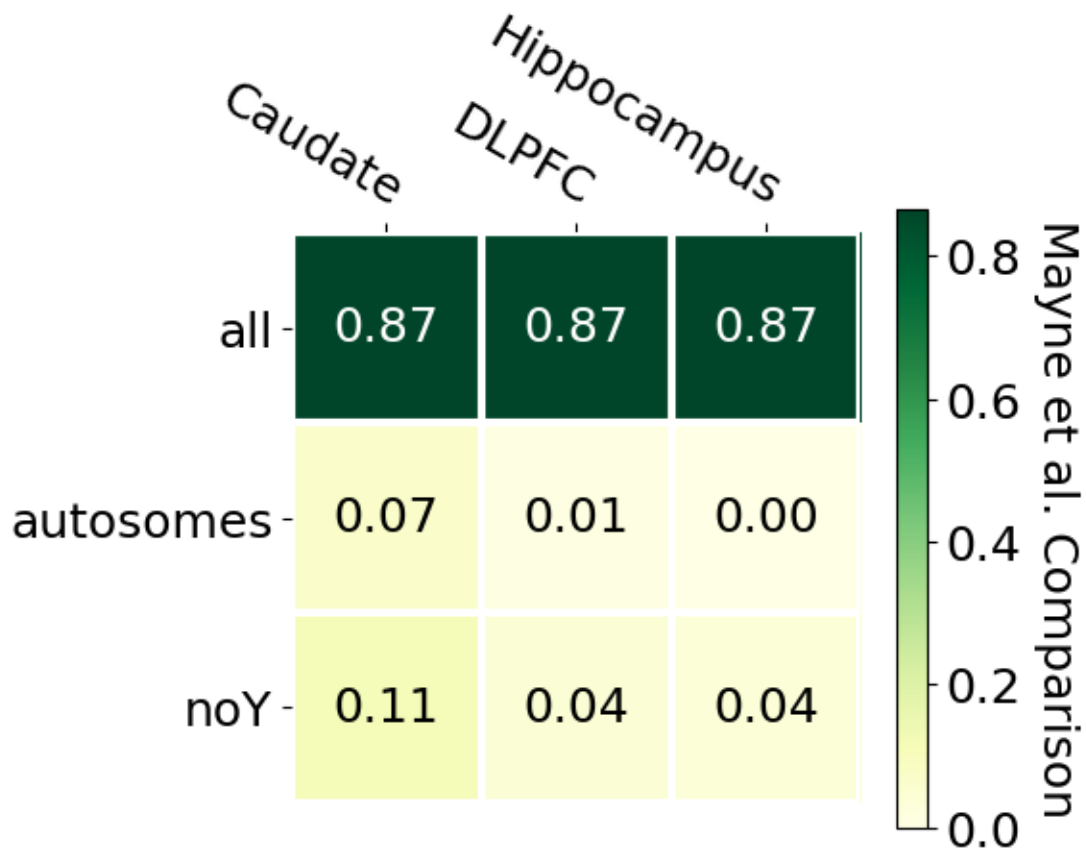
```
dfm.to_csv('brainseq_overlap_mayne_fraction.txt', sep='\t')
dfm
```

```
[14]:
```

	Caudate	DLPFC	Hippocampus
Method			
all	0.866667	0.866667	0.866667
autosomes	0.065116	0.009302	0.000000
noY	0.114754	0.040984	0.036885

```
[15]: %matplotlib inline
fig, ax = plt.subplots(figsize=(6, 5))
plt.rcParams.update({'font.size': 18})
im, cbar = heatmap(dfm, dfm.index, dfm.columns, ax=ax,
                    cmap="YlGn", cbarlabel="Mayne et al. Comparison")
texts = annotate_heatmap(im, valfmt="{x:.2f}")

fig.tight_layout()
plt.savefig('brainseq_overlap_mayne_heatmap.png')
plt.savefig('brainseq_overlap_mayne_heatmap.pdf')
plt.savefig('brainseq_overlap_mayne_heatmap.svg')
plt.show()
```



```
[16]: df = brainseq_shared_degs()

mm = mayne.loc[(mayne['Version'] == 'all'), ['Ensembl', 'Mayne']]
mm = get_biomart().merge(df, left_on='ensembl_gene_id', right_index=True)\
    .merge(mm.set_index('Ensembl'), left_on="ensembl_gene_id",
           right_index=True, how='right')\
    .set_index('external_gene_name')
mm.to_csv('brainseq_overlap_mayne_all_genes.csv')
```

```
[17]: mm
```

```
[17]:
```

external_gene_name	ensembl_gene_id	entrezgene	\
KDM5D	ENSG00000012817	8284.0	
EIF1AY	ENSG000000198692	9086.0	
TXLNGY	ENSG000000131002	NaN	
NaN	ENSG000000274655	NaN	
PRKY	ENSG000000099725	5616.0	
NLGN4Y	ENSG000000165246	22829.0	
NaN	ENSG000000233864	NaN	
TXLNGY	ENSG000000131002	NaN	
UTY	ENSG000000183878	7404.0	
TTY14	ENSG000000176728	NaN	
TMSB4Y	ENSG000000154620	9087.0	
USP9Y	ENSG000000114374	8287.0	
KDM6A	ENSG000000147050	7403.0	
ZFY	ENSG000000067646	7544.0	
TBL1Y	ENSG000000092377	90665.0	
KDM5C	ENSG000000126012	8242.0	

external_gene_name	description	\
KDM5D	lysine demethylase 5D [Source:HGNC Symbol;Acc:...	
EIF1AY	eukaryotic translation initiation factor 1A Y-...	
TXLNGY	taxilin gamma pseudogene, Y-linked [Source:HGN...	
NaN		NaN
PRKY	protein kinase Y-linked (pseudogene) [Source:H...	
NLGN4Y	neuroligin 4 Y-linked [Source:HGNC Symbol;Acc:...	
NaN		NaN
TXLNGY	taxilin gamma pseudogene, Y-linked [Source:HGN...	
UTY	ubiquitously transcribed tetra-tryptophan rep...	
TTY14	testis-specific transcript, Y-linked 14 (non-p...	
TMSB4Y	thymosin beta 4 Y-linked [Source:HGNC Symbol;A...	
USP9Y	ubiquitin specific peptidase 9 Y-linked [Sourc...	
KDM6A	lysine demethylase 6A [Source:HGNC Symbol;Acc:...	

ZFY	zinc finger protein Y-linked [Source:HGNC Symb...
TBL1Y	transducin beta like 1 Y-linked [Source:HGNC S...
KDM5C	lysine demethylase 5C [Source:HGNC Symbol;Acc:...

	Caudate	DLPFC	Hippocampus	Mayne
external_gene_name				
KDM5D	1.0	1.0	1.0	1
EIF1AY	1.0	1.0	1.0	1
TXLNGY	1.0	1.0	1.0	1
NaN	NaN	NaN	NaN	1
PRKY	1.0	1.0	1.0	1
NLGN4Y	1.0	1.0	1.0	1
NaN	NaN	NaN	NaN	1
TXLNGY	1.0	1.0	1.0	1
UTY	1.0	1.0	1.0	1
TTY14	1.0	1.0	1.0	1
TMSB4Y	1.0	1.0	1.0	1
USP9Y	1.0	1.0	1.0	1
KDM6A	1.0	1.0	1.0	1
ZFY	1.0	1.0	1.0	1
TBL1Y	1.0	1.0	1.0	1
KDM5C	1.0	1.0	1.0	1

#### 1.4 Gershoni and Pietrokovski (GTEx) Comparison

```
[18]: gtex_file = here('input/public_results/extract_geneLists/_m/
↳gp_gtex_sex_diff_genes.csv')
gtex = pd.read_csv(gtex_file).rename(columns={'Ensembl': 'Gencode'}).
↳drop('Gene', axis=1)
gtex['Ensembl'] = gtex.Gencode.str.replace("\\\\.*", "", regex=True)
gtex.set_index('Ensembl', inplace=True)
gtex.shape
```

```
[18]: (18759, 47)
```

```
[19]: zz = dict()
for tissue in ['caudate', 'dlpfc', 'hippocampus']:
    ww = []; yy = []
    fdr = 0.05
    for col in gtex.drop(['Symbol', 'Gencode'], axis=1).columns:
        gtex.loc[gtex[col] != 0, col] = 1
        tot = len(set(gtex[gtex[col] != 0]).index)
        overlap = len(set(get_deg_sig(tissue, fdr).ensemblID) &
↳set(gtex[gtex[col] != 0]).index))
        xx = overlap / tot
        ww.append(xx)
```

```

print("There is %d (%.1f%%) overlap between %s and %s GTEEx!" %
      (overlap, xx* 100, tissue_annotation(tissue), col))
print("")
zz[tissue_annotation(tissue)] = ww

```

```

There is 32 (12.2%) overlap between Caudate and Adipose-Subcutaneous GTEEx!
There is 17 (34.0%) overlap between Caudate and Adipose-Visceral GTEEx!
There is 14 (82.4%) overlap between Caudate and Adrenal_Gland GTEEx!
There is 20 (62.5%) overlap between Caudate and Artery-Aorta GTEEx!
There is 15 (88.2%) overlap between Caudate and Artery-Coronary GTEEx!
There is 22 (59.5%) overlap between Caudate and Artery-Tibial GTEEx!
There is 10 (100.0%) overlap between Caudate and Bladder GTEEx!
There is 10 (76.9%) overlap between Caudate and Brain-Amygdala GTEEx!
There is 17 (11.0%) overlap between Caudate and Brain-Anterior_cingulate_cortex
GTEEx!
There is 13 (86.7%) overlap between Caudate and Brain-Caudate GTEEx!
There is 11 (91.7%) overlap between Caudate and Brain-Cerebellar_Hemisphere
GTEEx!
There is 16 (39.0%) overlap between Caudate and Brain-Cerebellum GTEEx!
There is 10 (100.0%) overlap between Caudate and Brain-Cortex GTEEx!
There is 12 (92.3%) overlap between Caudate and Brain-Frontal_Cortex GTEEx!
There is 11 (91.7%) overlap between Caudate and Brain-Hippocampus GTEEx!
There is 12 (85.7%) overlap between Caudate and Brain-Hypothalamus GTEEx!
There is 12 (85.7%) overlap between Caudate and Brain-Nucleus_accumbens GTEEx!
There is 12 (100.0%) overlap between Caudate and Brain-Putamen GTEEx!
There is 10 (76.9%) overlap between Caudate and Brain-Spinal_cord GTEEx!
There is 12 (100.0%) overlap between Caudate and Brain-Substantia_nigra GTEEx!
There is 195 (3.2%) overlap between Caudate and Breast-Mammary Tissue GTEEx!
There is 22 (91.7%) overlap between Caudate and Cells-
EBV.transformed_lymphocytes GTEEx!
There is 24 (77.4%) overlap between Caudate and Cells-Transformed GTEEx!
There is 15 (100.0%) overlap between Caudate and Colon-Sigmoid GTEEx!
There is 16 (88.9%) overlap between Caudate and Colon-Transverse GTEEx!
There is 15 (100.0%) overlap between Caudate and Esophagus-
Gastroesophageal_Junction GTEEx!
There is 22 (84.6%) overlap between Caudate and Esophagus-Mucosa GTEEx!
There is 20 (69.0%) overlap between Caudate and Esophagus-Muscularis GTEEx!
There is 12 (70.6%) overlap between Caudate and Heart-Atrial_Appendage GTEEx!
There is 23 (15.4%) overlap between Caudate and Heart-Left_Ventricle GTEEx!
There is 11 (68.8%) overlap between Caudate and Kidney-Cortex GTEEx!
There is 18 (72.0%) overlap between Caudate and Liver GTEEx!
There is 18 (81.8%) overlap between Caudate and Lung GTEEx!
There is 11 (100.0%) overlap between Caudate and Minor_Salivary_Gland GTEEx!
There is 33 (11.6%) overlap between Caudate and Muscle-Skeletal GTEEx!
There is 29 (56.9%) overlap between Caudate and Nerve-Tibial GTEEx!
There is 16 (94.1%) overlap between Caudate and Pancreas GTEEx!
There is 15 (57.7%) overlap between Caudate and Pituitary GTEEx!

```

There is 21 (14.2%) overlap between Caudate and Skin-Not\_Sun\_Exposed GTEx!  
 There is 32 (12.9%) overlap between Caudate and Skin-Sun\_Exposed GTEx!  
 There is 14 (82.4%) overlap between Caudate and Small Intestine-Terminal Ileum GTEx!  
 There is 16 (66.7%) overlap between Caudate and Spleen GTEx!  
 There is 19 (54.3%) overlap between Caudate and Stomach GTEx!  
 There is 25 (37.3%) overlap between Caudate and Thyroid GTEx!  
 There is 9 (81.8%) overlap between Caudate and Whole\_Blood GTEx!

There is 18 (6.8%) overlap between DLPFC and Adipose-Subcutaneous GTEx!  
 There is 13 (26.0%) overlap between DLPFC and Adipose-Visceral GTEx!  
 There is 13 (76.5%) overlap between DLPFC and Adrenal\_Gland GTEx!  
 There is 17 (53.1%) overlap between DLPFC and Artery-Aorta GTEx!  
 There is 14 (82.4%) overlap between DLPFC and Artery-Coronary GTEx!  
 There is 18 (48.6%) overlap between DLPFC and Artery-Tibial GTEx!  
 There is 10 (100.0%) overlap between DLPFC and Bladder GTEx!  
 There is 10 (76.9%) overlap between DLPFC and Brain-Amygdala GTEx!  
 There is 17 (11.0%) overlap between DLPFC and Brain-Anterior\_cingulate\_cortex GTEx!  
 There is 13 (86.7%) overlap between DLPFC and Brain-Caudate GTEx!  
 There is 11 (91.7%) overlap between DLPFC and Brain-Cerebellar\_Hemisphere GTEx!  
 There is 13 (31.7%) overlap between DLPFC and Brain-Cerebellum GTEx!  
 There is 10 (100.0%) overlap between DLPFC and Brain-Cortex GTEx!  
 There is 12 (92.3%) overlap between DLPFC and Brain-Frontal\_Cortex GTEx!  
 There is 10 (83.3%) overlap between DLPFC and Brain-Hippocampus GTEx!  
 There is 11 (78.6%) overlap between DLPFC and Brain-Hypothalamus GTEx!  
 There is 12 (85.7%) overlap between DLPFC and Brain-Nucleus\_accumbens GTEx!  
 There is 11 (91.7%) overlap between DLPFC and Brain-Putamen GTEx!  
 There is 9 (69.2%) overlap between DLPFC and Brain-Spinal\_cord GTEx!  
 There is 11 (91.7%) overlap between DLPFC and Brain-Substantia\_nigra GTEx!  
 There is 66 (1.1%) overlap between DLPFC and Breast-Mammary Tissue GTEx!  
 There is 16 (66.7%) overlap between DLPFC and Cells-EBV.transformed\_lymphocytes GTEx!  
 There is 16 (51.6%) overlap between DLPFC and Cells-Transformed GTEx!  
 There is 14 (93.3%) overlap between DLPFC and Colon-Sigmoid GTEx!  
 There is 13 (72.2%) overlap between DLPFC and Colon-Transverse GTEx!  
 There is 13 (86.7%) overlap between DLPFC and Esophagus-Gastroesophageal\_Junction GTEx!  
 There is 16 (61.5%) overlap between DLPFC and Esophagus-Mucosa GTEx!  
 There is 15 (51.7%) overlap between DLPFC and Esophagus-Muscularis GTEx!  
 There is 11 (64.7%) overlap between DLPFC and Heart-Atrial\_Appendage GTEx!  
 There is 16 (10.7%) overlap between DLPFC and Heart-Left\_Ventricle GTEx!  
 There is 11 (68.8%) overlap between DLPFC and Kidney-Cortex GTEx!  
 There is 14 (56.0%) overlap between DLPFC and Liver GTEx!  
 There is 16 (72.7%) overlap between DLPFC and Lung GTEx!  
 There is 10 (90.9%) overlap between DLPFC and Minor\_Salivary\_Gland GTEx!  
 There is 23 (8.1%) overlap between DLPFC and Muscle-Skeletal GTEx!  
 There is 20 (39.2%) overlap between DLPFC and Nerve-Tibial GTEx!

There is 13 (76.5%) overlap between DLPFC and Pancreas GTEx!  
 There is 14 (53.8%) overlap between DLPFC and Pituitary GTEx!  
 There is 16 (10.8%) overlap between DLPFC and Skin-Not\_Sun\_Exposed GTEx!  
 There is 19 (7.6%) overlap between DLPFC and Skin-Sun\_Exposed GTEx!  
 There is 13 (76.5%) overlap between DLPFC and Small Intestine-Terminal Ileum GTEx!  
 There is 14 (58.3%) overlap between DLPFC and Spleen GTEx!  
 There is 15 (42.9%) overlap between DLPFC and Stomach GTEx!  
 There is 17 (25.4%) overlap between DLPFC and Thyroid GTEx!  
 There is 8 (72.7%) overlap between DLPFC and Whole\_Blood GTEx!

There is 22 (8.4%) overlap between Hippocampus and Adipose-Subcutaneous GTEx!  
 There is 15 (30.0%) overlap between Hippocampus and Adipose-Visceral GTEx!  
 There is 14 (82.4%) overlap between Hippocampus and Adrenal\_Gland GTEx!  
 There is 20 (62.5%) overlap between Hippocampus and Artery-Aorta GTEx!  
 There is 15 (88.2%) overlap between Hippocampus and Artery-Coronary GTEx!  
 There is 20 (54.1%) overlap between Hippocampus and Artery-Tibial GTEx!  
 There is 10 (100.0%) overlap between Hippocampus and Bladder GTEx!  
 There is 10 (76.9%) overlap between Hippocampus and Brain-Amygdala GTEx!  
 There is 17 (11.0%) overlap between Hippocampus and Brain-Anterior\_cingulate\_cortex GTEx!  
 There is 13 (86.7%) overlap between Hippocampus and Brain-Caudate GTEx!  
 There is 11 (91.7%) overlap between Hippocampus and Brain-Cerebellar\_Hemisphere GTEx!  
 There is 15 (36.6%) overlap between Hippocampus and Brain-Cerebellum GTEx!  
 There is 10 (100.0%) overlap between Hippocampus and Brain-Cortex GTEx!  
 There is 12 (92.3%) overlap between Hippocampus and Brain-Frontal\_Cortex GTEx!  
 There is 11 (91.7%) overlap between Hippocampus and Brain-Hippocampus GTEx!  
 There is 12 (85.7%) overlap between Hippocampus and Brain-Hypothalamus GTEx!  
 There is 12 (85.7%) overlap between Hippocampus and Brain-Nucleus\_accumbens GTEx!  
 There is 12 (100.0%) overlap between Hippocampus and Brain-Putamen GTEx!  
 There is 10 (76.9%) overlap between Hippocampus and Brain-Spinal\_cord GTEx!  
 There is 12 (100.0%) overlap between Hippocampus and Brain-Substantia\_nigra GTEx!  
 There is 35 (0.6%) overlap between Hippocampus and Breast-Mammary Tissue GTEx!  
 There is 21 (87.5%) overlap between Hippocampus and Cells-EBV.transformed\_lymphocytes GTEx!  
 There is 22 (71.0%) overlap between Hippocampus and Cells-Transformed GTEx!  
 There is 15 (100.0%) overlap between Hippocampus and Colon-Sigmoid GTEx!  
 There is 15 (83.3%) overlap between Hippocampus and Colon-Transverse GTEx!  
 There is 15 (100.0%) overlap between Hippocampus and Esophagus-Gastroesophageal\_Junction GTEx!  
 There is 19 (73.1%) overlap between Hippocampus and Esophagus-Mucosa GTEx!  
 There is 17 (58.6%) overlap between Hippocampus and Esophagus-Muscularis GTEx!  
 There is 12 (70.6%) overlap between Hippocampus and Heart-Atrial\_Appendage GTEx!  
 There is 20 (13.4%) overlap between Hippocampus and Heart-Left\_Ventricle GTEx!  
 There is 11 (68.8%) overlap between Hippocampus and Kidney-Cortex GTEx!



There is 16 (64.0%) overlap between Hippocampus and Liver GTEX!  
 There is 18 (81.8%) overlap between Hippocampus and Lung GTEX!  
 There is 11 (100.0%) overlap between Hippocampus and Minor\_Salivary\_Gland GTEX!  
 There is 25 (8.8%) overlap between Hippocampus and Muscle-Skeletal GTEX!  
 There is 23 (45.1%) overlap between Hippocampus and Nerve-Tibial GTEX!  
 There is 15 (88.2%) overlap between Hippocampus and Pancreas GTEX!  
 There is 15 (57.7%) overlap between Hippocampus and Pituitary GTEX!  
 There is 18 (12.2%) overlap between Hippocampus and Skin-Not\_Sun\_Exposed GTEX!  
 There is 23 (9.2%) overlap between Hippocampus and Skin-Sun\_Exposed GTEX!  
 There is 14 (82.4%) overlap between Hippocampus and Small Intestine-Terminal Ileum GTEX!  
 There is 16 (66.7%) overlap between Hippocampus and Spleen GTEX!  
 There is 17 (48.6%) overlap between Hippocampus and Stomach GTEX!  
 There is 19 (28.4%) overlap between Hippocampus and Thyroid GTEX!  
 There is 8 (72.7%) overlap between Hippocampus and Whole\_Blood GTEX!

```
[20]: dfg = pd.DataFrame.from_dict(zz)
dfg['Tissue'] = gtex.drop(['Symbol', 'Gencode'], axis=1).columns
dfg = dfg.set_index('Tissue')
dfg.to_csv('brainseq_overlap_gtex_fraction.txt', sep='\t')
dfg.head()
```

```
[20]:
```

	Caudate	DLPFC	Hippocampus
Tissue			
Adipose-Subcutaneous	0.121673	0.068441	0.083650
Adipose-Visceral	0.340000	0.260000	0.300000
Adrenal_Gland	0.823529	0.764706	0.823529
Artery-Aorta	0.625000	0.531250	0.625000
Artery-Coronary	0.882353	0.823529	0.882353

```
[21]: %matplotlib inline
fig, ax = plt.subplots(figsize=(36, 8))
plt.rcParams.update({'font.size': 12})
im, cbar = heatmap(dfg.T, dfg.columns, dfg.index, ax=ax,
                  cmap="YlGn",
                  cbarlabel="Gershoni and Pietrokovski Comparison")
texts = annotate_heatmap(im, valfmt="{x:.2f}")

fig.tight_layout()
plt.savefig('brainseq_overlap_gtex_heatmap_full.png')
plt.savefig('brainseq_overlap_gtex_heatmap_full.pdf')
plt.savefig('brainseq_overlap_gtex_heatmap_full.svg')
plt.show()
```



```
        right_index=True, how='outer')\  
        .set_index('external_gene_name')  
gg.to_csv('brainseq_overlap_gtex_genes.csv')
```

## 1.5 Session information

```
[24]: session_info.show()
```

```
[24]: <IPython.core.display.HTML object>
```