

CSCI5481_HWK2

October 10, 2018

1 CSCI 5481 Homework 2

Prof: Knights

Author: Christopher White

Date: October 11, 2018

```
In [4]: import os
        from nwalgn import *
```

1.1 Run Command Line

```
In [5]: !python nwalgn.py -q './Homework 2 - sequences/Human_HOX.fa' -r './Homework 2 - sequences/Human_HOX.fa'
```

```
In [6]: !python nwalgn.py -q './Homework 2 - sequences/Human_HOX.fa' -r './Homework 2 - sequences/Human_HOX.fa'
```

```
In [7]: !python nwalgn.py -q './Homework 2 - sequences/Human_PAX.fa' -r './Homework 2 - sequences/Human_PAX.fa'
```

```
In [8]: !python nwalgn.py -q './Homework 2 - sequences/Human_PAX.fa' -r './Homework 2 - sequences/Human_PAX.fa'
```

1.2 Read in Sequences and Calculate Scores

```
In [33]: human_HOX = read_seq('./Homework 2 - sequences/Human_HOX.fa')
        fly_HOX = read_seq('./Homework 2 - sequences/Fly_HOX.fa')
        human_PAX = read_seq('./Homework 2 - sequences/Human_PAX.fa')
        fly_PAX = read_seq('./Homework 2 - sequences/Fly_PAX.fa')

        score_hox,_,_ = find_alignments(human_HOX, fly_HOX)
        score_pax,_,_ = find_alignments(human_PAX, fly_PAX)

        print('HOX Score: {}'.format(score_hox))
        print('PAX Score: {}'.format(score_pax))
```

HOX Score: -494.0

PAX Score: -850.0

1.3 Run Randomized Alignments

```
In [41]: from random import sample
import numpy as np
import pandas as pd
from multiprocessing import Pool

In [50]: # Create a wrapper for multiprocessing
def random_permute_HOX(i):
    score, _, _ = find_alignments(sample(human_HOX, k=len(human_HOX)),
                                   sample(fly_HOX, k=len(fly_HOX)))

    return score

def random_permute_PAX(i):
    score, _, _ = find_alignments(sample(human_PAX, k=len(human_PAX)),
                                   sample(fly_PAX, k=len(fly_PAX)))

    return score

In [63]: NUM_ITERS = 10000

p = Pool(processes=8)
score_hox_iter = p.map(random_permute_HOX, [i for i in range(NUM_ITERS)])
score_pax_iter = p.map(random_permute_PAX, [i for i in range(NUM_ITERS)])

score_table = pd.DataFrame({'HOX':score_hox_iter, 'PAX':score_pax_iter})

In [64]: score_table.head()

Out [64]:
```

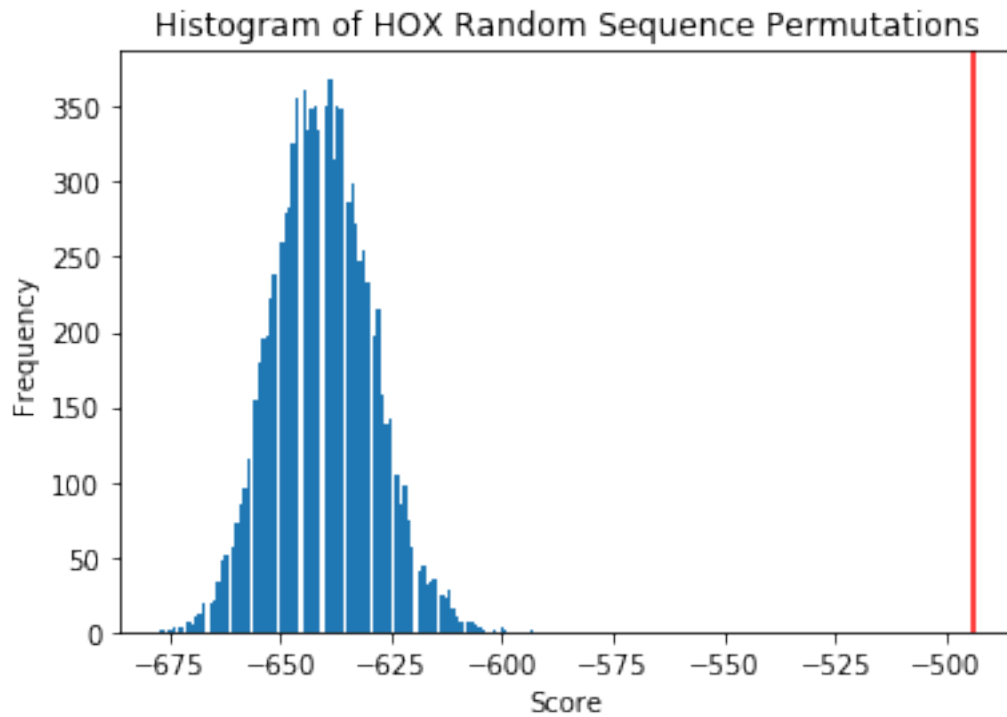
	HOX	PAX
0	-636.0	-1300.0
1	-656.0	-1302.0
2	-624.0	-1313.0
3	-645.0	-1296.0
4	-629.0	-1308.0

1.4 Plot Histograms

```
In [53]: import matplotlib as mpl
import matplotlib.pyplot as plt

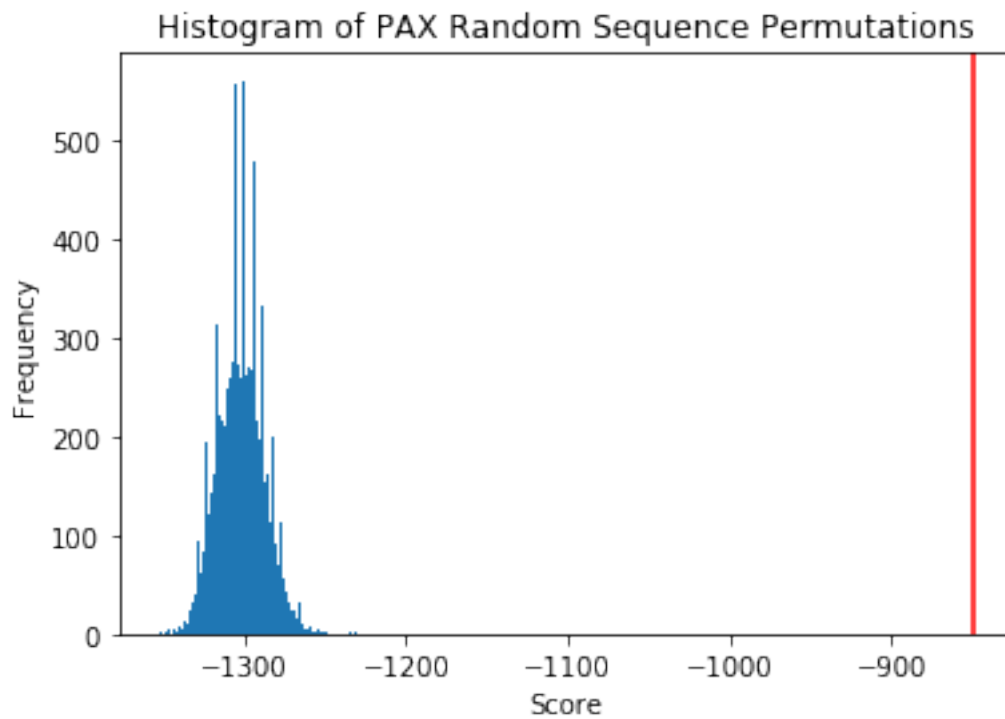
In [69]: hax = score_table['HOX'].plot.hist(bins=100)
hax.set_title('Histogram of HOX Random Sequence Permutations')
hax.set_xlabel('Score')
hax.vlines(score_hox, 0, 1, transform=hax.get_xaxis_transform(), colors='r')

Out [69]: <matplotlib.collections.LineCollection at 0x117843f28>
```



```
In [70]: pax = score_table['PAX'].plot.hist(bins=100)
pax.set_title('Histogram of PAX Random Sequence Permutations')
pax.set_xlabel('Score')
pax.vlines(score_pax, 0, 1, transform=pax.get_xaxis_transform(), colors='r')
```

```
Out[70]: <matplotlib.collections.LineCollection at 0x11792f4a8>
```



```
In [67]: print('HOX Mean: {}\\nHOX Std: {}'.format(score_table['HOX'].mean(), score_table['HOX'].std())  
         print('PAX Mean: {}\\nPAX Std: {}'.format(score_table['PAX'].mean(), score_table['PAX'].std()))
```

```
HOX Mean: -640.4855  
HOX Std: 11.026737896806221  
PAX Mean: -1302.0103  
PAX Std: 14.26688993175573
```

```
In [68]: p.close()
```