main

July 14, 2021

1 Examine permutation based on female sample size

[1]:
```
import re
import pandas as pd
from glob import iglob
```

1.1 Main

[2]:
```
df = pd.DataFrame()
for filename in iglob("../../_m/permutation_*/diffExpr_szVctl_FDR05.txt"):
    m = re.search("\d+", filename)
    dt = pd.read_csv(filename, sep='\t', index_col=0)
    dt["Permutation"] = m.group(0)
    df = pd.concat([df, dt], axis=0)
df.to_csv("permutations.csv")
df.shape
```

[2]:
```
(4024, 17)
```

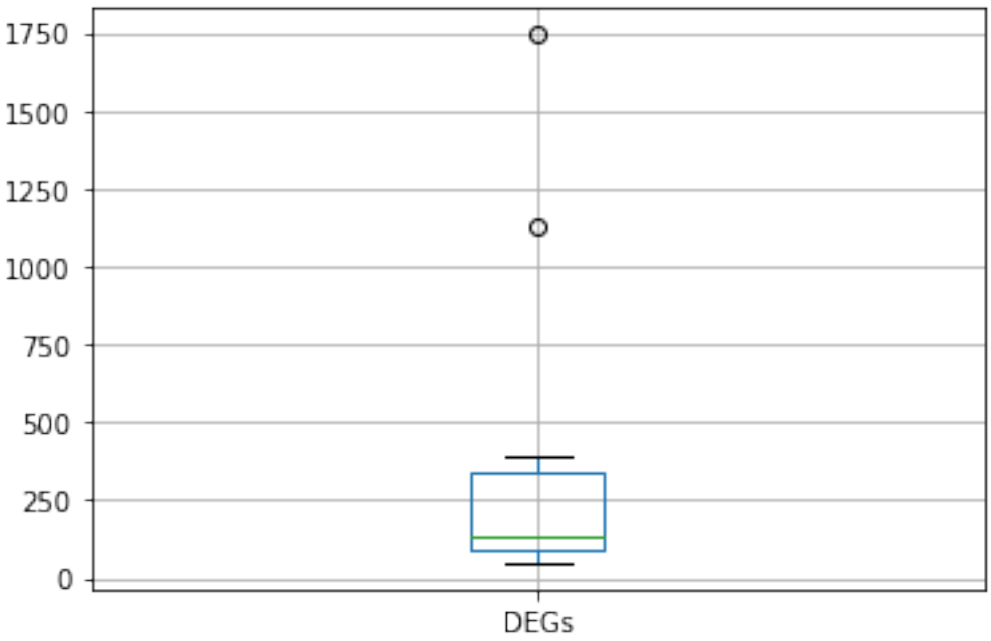
[3]:
```
xx = df.groupby("Permutation").size().reset_index().rename(columns={0:"DEGs"})
print("Median: %f" % xx.DEGs.median())
xx.DEGs.describe()
```

```
Median: 136.500000
```

[3]:
```
count      10.000000
mean      402.400000
std       571.606158
min        48.000000
25%        89.000000
50%       136.500000
75%       340.750000
max      1744.000000
Name: DEGs, dtype: float64
```

[4]:
```
xx.boxplot()
```

[4]:
```
<AxesSubplot:>
```



[5]:
```
female = pd.read_csv("../../../female_analysis/_m/genes/diffExpr_szVctl_FDR05.
 ↪txt", sep='\t', index_col=0)
print("There are %d DEGs with females!" % female.shape[0])
```

```
There are 922 DEGs with females!
```

[ ]:
```

```