main

August 12, 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_CtrlvsSZ_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]:
```
(131, 17)
```

[3]:
```
xx = df.groupby("Permutation").size()\
       .reset_index().rename(columns={0:"DEGs"})\
       .merge(pd.DataFrame({"Permutation": [str(x).zfill(2) for x in␣
 ↪range(1,11)]}),
              on="Permutation", how="outer")\
       .fillna(0).sort_values("Permutation")
print("Median: %f" % xx.DEGs.median())
xx.DEGs.describe()
```

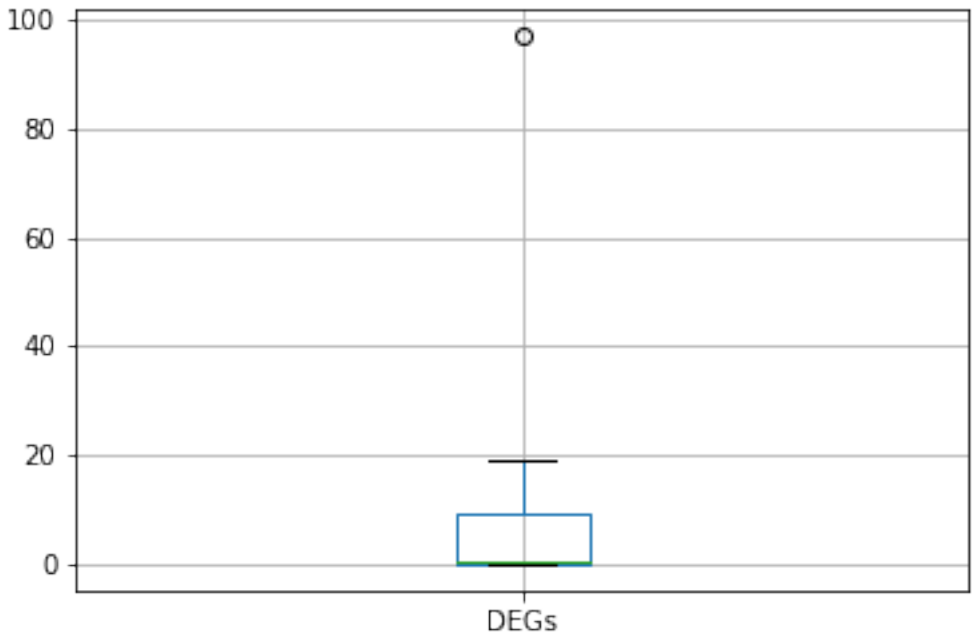
```
Median: 0.500000
```

[3]:
```
count    10.000000
mean     13.100000
std      30.189954
min       0.000000
25%       0.000000
50%       0.500000
75%       9.500000
max      97.000000
```

```
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 1 DEGs with females!
```

[ ]:
```

```