

main

July 10, 2021

1 Feature summary analysis

```
[1]: import numpy as np
import pandas as pd
```

1.1 Summary plots

1.1.1 Genes

```
[2]: genes = pd.read_csv('../_m/genes/diffExpr_maleVfemale_full.txt', sep='\t')
genes = genes[(genes['adj.P.Val'] < 0.05)].sort_values('adj.P.Val')
genes.head()
```

```
[2]:
```

	gene_id	logFC	AveExpr	t	P.Value	adj.P.Val	\
0	ENSG00000174796.12	-0.142817	3.601461	-5.005702	7.696597e-07	0.01523	
14	ENSG00000034053.14	0.151370	7.379052	4.427216	1.180359e-05	0.01523	
13	ENSG00000122435.9	-0.195373	3.526808	-4.453764	1.039225e-05	0.01523	
12	ENSG00000188404.8	-0.473004	0.894959	-4.456125	1.032278e-05	0.01523	
11	ENSG00000198818.9	-0.152785	3.979944	-4.478548	9.339884e-06	0.01523	

	z.std	ensembl_gene_id	position	Direction	hgnc_symbol	\
0	-4.942908	ENSG00000174796	12	DOWN	THAP6	
14	4.381184	ENSG00000034053	14	UP	APBA2	
13	-4.408847	ENSG00000122435	9	DOWN	TRMT13	
12	-4.410299	ENSG00000188404	8	DOWN	SELL	
11	-4.431917	ENSG00000198818	9	DOWN	SFT2D1	

	percentage_gene_gc_content	gene_biotype	chromosome_name
0	40.52	protein_coding	4
14	48.63	protein_coding	15
13	35.80	protein_coding	1
12	38.85	protein_coding	1
11	43.21	protein_coding	6

```
[3]: genes.sort_values('adj.P.Val').to_csv('chromosome_DEG.csv', index=False,
↪header=True)
```

1.2 DE summary

1.2.1 DE (feature)

```
[4]: gg = len(set(genes['gene_id']))  
  
print("\nGene:\t\t%d" % (gg))
```

Gene: 1315

1.2.2 Feature effect size summary

```
[5]: feature_list = ['Genes', 'Transcript', 'Exons', 'Junctions']  
feature_df = [genes]  
ii = 0  
  
ff = feature_df[ii]  
half = len(set(ff[(np.abs(ff['logFC']) >= 0.5)].gene_id))  
one = len(set(ff[(np.abs(ff['logFC']) >= 1)].gene_id))  
print("\nThere are %d unique %s with abs(log2FC) >= 0.5" % (half, feature_list[ii]))  
print("There are %d unique %s with abs(log2FC) >= 1" % (one, feature_list[ii]))
```

There are 7 unique Genes with abs(log2FC) >= 0.5

There are 0 unique Genes with abs(log2FC) >= 1

1.3 Autosomal only

```
[6]: genes.chromosome_name.fillna('?', inplace=True)  
  
auto = genes[(genes.chromosome_name.str.contains('\d+'))].copy()\br/>        .rename(columns={'chromosome_name': 'chr', 'hgnc_symbol': 'gene_name'})  
auto = auto[['gene_id', 'chr', 'gene_name', 'logFC', 'adj.P.Val']]  
auto.head()
```

```
[6]:
```

	gene_id	chr	gene_name	logFC	adj.P.Val
0	ENSG00000174796.12	4	THAP6	-0.142817	0.01523
14	ENSG00000034053.14	15	APBA2	0.151370	0.01523
13	ENSG00000122435.9	1	TRMT13	-0.195373	0.01523
12	ENSG00000188404.8	1	SELL	-0.473004	0.01523
11	ENSG00000198818.9	6	SFT2D1	-0.152785	0.01523

```
[7]: auto.sort_values('adj.P.Val').to_csv('autosomal_DEG.csv', index=False,  
        header=True)
```

1.4 DE summary

1.4.1 DE (feature)

```
[8]: gg = len(set(auto['gene_id']))  
  
print("\nGene:\t\t%d" % (gg))
```

Gene: 1254

1.4.2 Feature effect size summary

```
[9]: feature_list = ['Genes', 'Transcript', 'Exons', 'Junctions']  
feature_df = [auto]  
ii = 0  
  
ff = feature_df[ii]  
half = len(set(ff[(np.abs(ff['logFC']) >= 0.5)].gene_id))  
one = len(set(ff[(np.abs(ff['logFC']) >= 1)].gene_id))  
print("\nThere are %d unique %s with abs(log2FC) >= 0.5" % (half,   
    ↪feature_list[ii]))  
print("There are %d unique %s with abs(log2FC) >= 1" % (one, feature_list[ii]))
```

There are 7 unique Genes with abs(log2FC) >= 0.5

There are 0 unique Genes with abs(log2FC) >= 1

```
[ ]:
```