

```
In [79]: import pandas as pd
pd.set_option('display.max_colwidth', 100)
import numpy as np
import gseapy as gp

# names = gp.get_library_name()
gobp = 'GO_Biological_Process_2021'
gomf = 'GO_Molecular_Function_2021'
gocc = 'GO_Cellular_Component_2021'

d = pd.read_csv(r'fdr5.csv', index_col=0)
g = d.loc[d['PPDE'] > 0.95].sort_values('PPDE', ascending=False).iloc[:1000].s

r = gp.enrichr(
    gene_list=g.index.to_list(),
    gene_sets=[gocc, gomf, gobp],
    cutoff=0.05,
    organism='Human',
).results
```

I used the data from part 1 of homework 3 that uses only PBMC data.

1. Top 3 GO Bio Processes

```
In [82]: r.loc[r['Gene_set'] == 'GO_Biological_Process_2021'].iloc[0:3,[1, 4]]
```

```
Out[82]:
```

	Term	Adjusted P-value
933	aerobic electron transport chain (GO:0019646)	1.884799e-16
934	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	1.884799e-16
935	mitochondrial electron transport, NADH to ubiquinone (GO:0006120)	1.313649e-09

2. Top 3 GO Molecular Functions

```
In [89]: r.loc[r['Gene_set'] == 'GO_Molecular_Function_2021'].iloc[0:3,[1, 4]]
```

```
Out[89]:
```

	Term	Adjusted P-value
301	oxidoreduction-driven active transmembrane transporter activity (GO:0015453)	5.046719e-15
302	NADH dehydrogenase (quinone) activity (GO:0050136)	4.727279e-10
303	NADH dehydrogenase (ubiquinone) activity (GO:0008137)	4.727279e-10

3. Top 3 GO Cellular Components

```
In [90]: r.loc[r['Gene_set'] == 'GO_Cellular_Component_2021'].iloc[0:3,[1, 4]]
```

```
Out[90]:
```

	Term	Adjusted P-value
0	mitochondrial membrane (GO:0031966)	4.288697e-23
1	mitochondrial inner membrane (GO:0005743)	3.534770e-21
2	organelle inner membrane (GO:0019866)	2.588383e-19

4. Does the selected GO term make sense given the list of DE genes?

Look at GO term aerobic electron transport chain (GO:0019646)

- This makes sense based on the name. Covid sufferers have trouble turning over oxygen, so it would make sense that RNA transcripts related to aerobic oxidation in the ETC will be DE. Likely these genes will be highly up-regulated because the cell is trying to boost these metabolic pathways to provide enough oxygen for the cells.
- A google scholar search for "covid-19" aerobic "electron transport chain" with year filter since 2021 yields 601 results which suggests that at least electron transport chain and covid-19 are strongly correlated.
- This paper <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7934846/> states the following:

Our observations demonstrate that, within 7 days of infection, morbidity, wasting, and failure to thrive in these animals is associated with significant decrease in expression of genes related to key cellular metabolic processes regulating aerobic cellular respiration and energy production.

However, their statement that expression of genes decreases actually disagrees with my hypothesis above. There is some effect from SARS-CoV-2 that causes the cells to express these important proteins at a much lower level.

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```