



Data interpretation:

Each row is a gene

Each column is our samples from people.

The colors show how active or inaction each gene is in the sample, the brighter yellow the gene is more active than usual. As for the ones that are dark-blue, that means they are less active.

The “Tree like” branches on the left and upper part are called dendrograms, these are used to show which genes are most similar to each other.

Code: (Done using python)

Each line explained:

seaborn–Used to graph heatmaps

Pandas –Used for reading CSVs and handling tables

Matplotlib.pyplot–used with seaborn

```
!pip install seaborn
```

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# loading data
df = pd.read_csv("normalized.csv")

df = df.set_index(df.columns[0])

df['variance'] = df.var(axis=1)
df_top = df.sort_values(by='variance', ascending=False).head(50)
df_top = df_top.drop(columns=['variance'])

sns.set(style="whitegrid")
sns.clustermap(
    df_top,
    cmap="viridis",
    figsize=(12, 10),
    standard_scale=0 # 0 = normalize genes (rows)
)

plt.title("Heatmap of Top 50 Most Variable Genes")
plt.show()
```

