Pairwise_Results

October 31, 2023

[ ]:

```
# load libraries
import pandas as pd # standard
import numpy as np # standard

import matplotlib.pyplot as plt
import seaborn as sns

import thermogram_utilities

import warnings
warnings.filterwarnings("ignore")
```

[ ]:

```
# set up for visuals
df = pd.read_excel("/Users/avery/OneDrive/Documents/GitHub/
 ↪Clinical_TLB_2023-2024/lung_cancer_tlb.xlsx")
df['CancerType'] = np.where(df['CancerType'].isna(), 'Control',␣
 ↪df['CancerType'])

# get location of cut off values
lower_column_index = df.columns.get_loc("T51")
upper_column_index = df.columns.get_loc("T83.1")
label_column_index = df.columns.get_loc("CancerType")

column_indices = np.arange(lower_column_index, upper_column_index)
column_indices = np.append(column_indices, 0)
column_indices = np.append(column_indices, 1)




column_indices = np.append(column_indices, label_column_index)

df = df.iloc[:, column_indices]

df_long = pd.melt(df, id_vars=['sample_id', 'pub_id', 'CancerType'],␣
 ↪var_name='temp', value_name='dsp' )
```

```
median_df = thermogram_utilities.median_curve(df_long, 'CancerType', 'temp',␣
 ↪'dsp')

median_df['temperature'] = median_df['temperature'].str.replace('T', '').
 ↪astype(float)
```

[ ]:

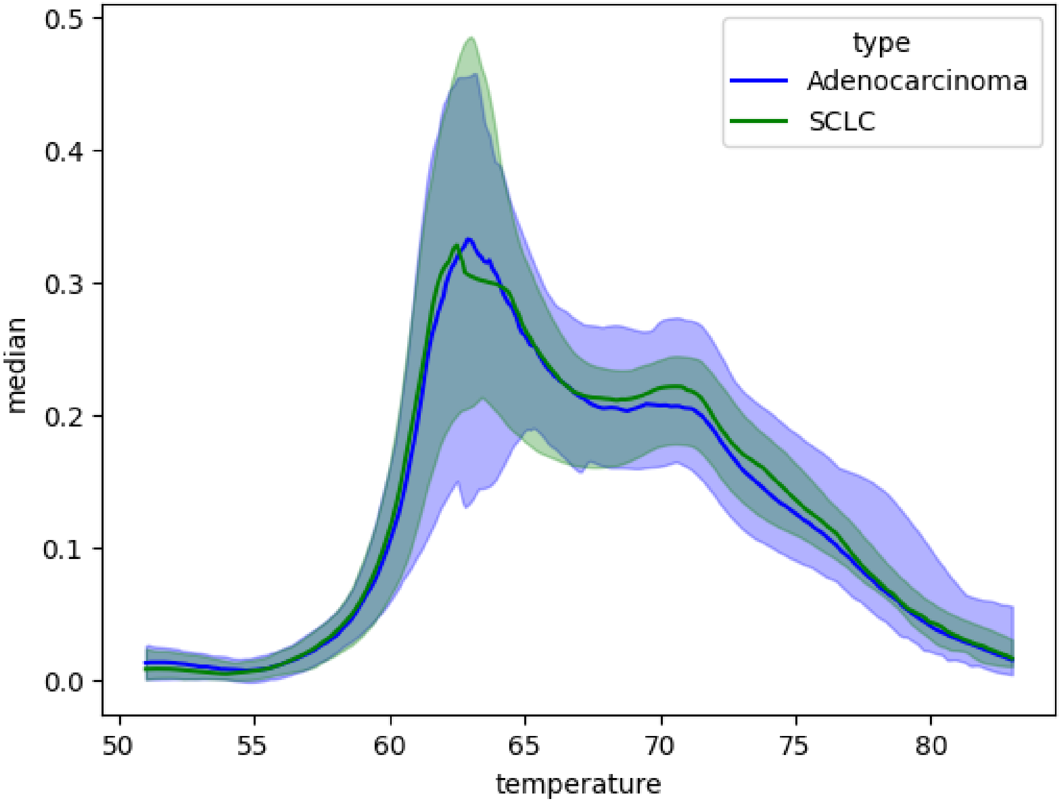
```
results_table = pd.DataFrame(columns=["Pair", "Weighted Accuracy", "AUC"])
df['CancerType'].value_counts()
```

[ ]:

```
CancerType
Adenocarcinoma    72
Control           51
Squamous          46
SCLC              16
NOS                8
Large cell         6
Name: count, dtype: int64
```

[ ]:

```
# adeno, sclc

graph_df = median_df[(median_df["type"] == "Adenocarcinoma") |␣
 ↪(median_df["type"] == "SCLC")]
colors = {'Adenocarcinoma': 'blue', 'SCLC': 'green'}  # Specify colors for each␣
 ↪type

# Create a line plot using Seaborn with matching colors
sns.lineplot(data=graph_df, x='temperature', y='median', hue='type',␣
 ↪palette=colors)

# Create separate ribbons for each "type" with matching colors
for type_name in graph_df['type'].unique():
    type_data = graph_df[graph_df['type'] == type_name]
    plt.fill_between(type_data["temperature"], type_data["lower_q"],␣
 ↪type_data["upper_q"], color=colors[type_name], alpha=0.3, label=type_name)
```

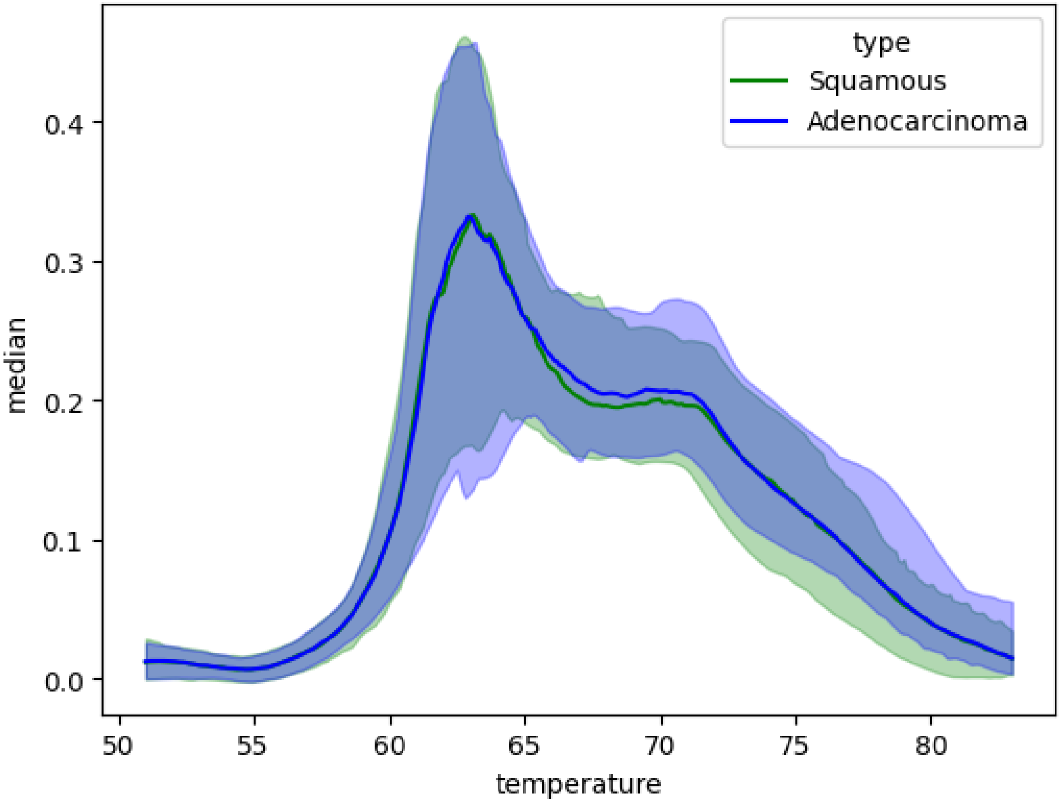


[ ]:
```
# adeno vs sclc results
adeno_sclc_1 = pd.read_excel("Adeno_SCLC.xlsx")
adeno_sclc_2 = pd.read_excel("Adeno_SCLC_1.xlsx")
adeno_sclc = pd.concat([adeno_sclc_1, adeno_sclc_2], ignore_index = True)

adeno_sclc['max_depth'] = np.where(pd.isna(adeno_sclc['max_depth']), "None",␣
 ↪adeno_sclc["max_depth"])
adeno_sclc['max_features'] = np.where(pd.isna(adeno_sclc['max_features']),␣
 ↪"None", adeno_sclc["max_features"])

adeno_sclc_results = adeno_sclc.groupby(['n_estimators', 'max_depth',␣
 ↪'max_features'], as_index = False).mean().sort_values("Weighted Accuracy",␣
 ↪ascending=False)
```

[ ]:
```
new_row = pd.DataFrame([["Adeno - SCLC", adeno_sclc_results.iloc[0]["Weighted␣
 ↪Accuracy"], adeno_sclc_results.iloc[0]["AUC"]]], columns=["Pair", "Weighted␣
 ↪Accuracy", "AUC"])

results_table = pd.concat([results_table, new_row], ignore_index=True)
```

[ ]:

```
# adeno, squamous

graph_df = median_df[(median_df["type"] == "Adenocarcinoma") |␣
 ↪(median_df["type"] == "Squamous")]
colors = {'Adenocarcinoma': 'blue', 'Squamous': 'green'}  # Specify colors for␣
 ↪each type

# Create a line plot using Seaborn with matching colors
sns.lineplot(data=graph_df, x='temperature', y='median', hue='type',␣
 ↪palette=colors)

# Create separate ribbons for each "type" with matching colors
for type_name in graph_df['type'].unique():
    type_data = graph_df[graph_df['type'] == type_name]
    plt.fill_between(type_data["temperature"], type_data["lower_q"],␣
 ↪type_data["upper_q"], color=colors[type_name], alpha=0.3, label=type_name)
```



[ ]:

```
adeno_squamous = pd.read_excel("Adeno_Squamous.xlsx")
```

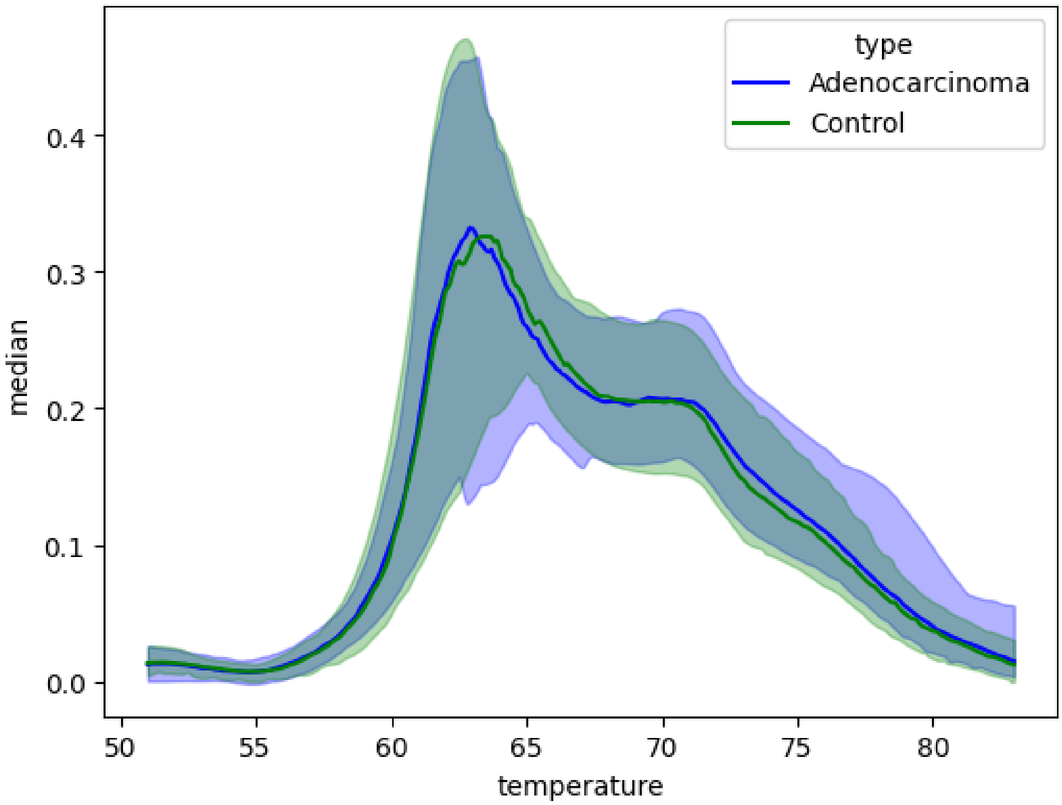
```
adeno_squamous['max_depth'] = np.where(pd.isna(adeno_squamous['max_depth']),␣
 ↪"None", adeno_squamous["max_depth"])
adeno_squamous['max_features'] = np.where(pd.
 ↪isna(adeno_squamous['max_features']), "None", adeno_squamous["max_features"])

adeno_squamous_results = adeno_squamous.groupby(['n_estimators', 'max_depth',␣
 ↪'max_features'], as_index = False).mean().sort_values("Weighted Accuracy",␣
 ↪ascending=False)
#adeno_squamous_results.head(5)
```

[ ]:

```
new_row = pd.DataFrame([["Adeno - Squamous", adeno_squamous_results.
 ↪iloc[0]["Weighted Accuracy"], adeno_squamous_results.iloc[0]["AUC"]]],␣
 ↪columns=["Pair", "Weighted Accuracy", "AUC"])

results_table = pd.concat([results_table, new_row], ignore_index=True)
```

[ ]:

```
#adeno, control
graph_df = median_df[(median_df["type"] == "Adenocarcinoma") |␣
 ↪(median_df["type"] == "Control")]
colors = {'Adenocarcinoma': 'blue', 'Control': 'green'}  # Specify colors for␣
 ↪each type

# Create a line plot using Seaborn with matching colors
sns.lineplot(data=graph_df, x='temperature', y='median', hue='type',␣
 ↪palette=colors)

# Create separate ribbons for each "type" with matching colors
for type_name in graph_df['type'].unique():
    type_data = graph_df[graph_df['type'] == type_name]
    plt.fill_between(type_data["temperature"], type_data["lower_q"],␣
 ↪type_data["upper_q"], color=colors[type_name], alpha=0.3, label=type_name)
```

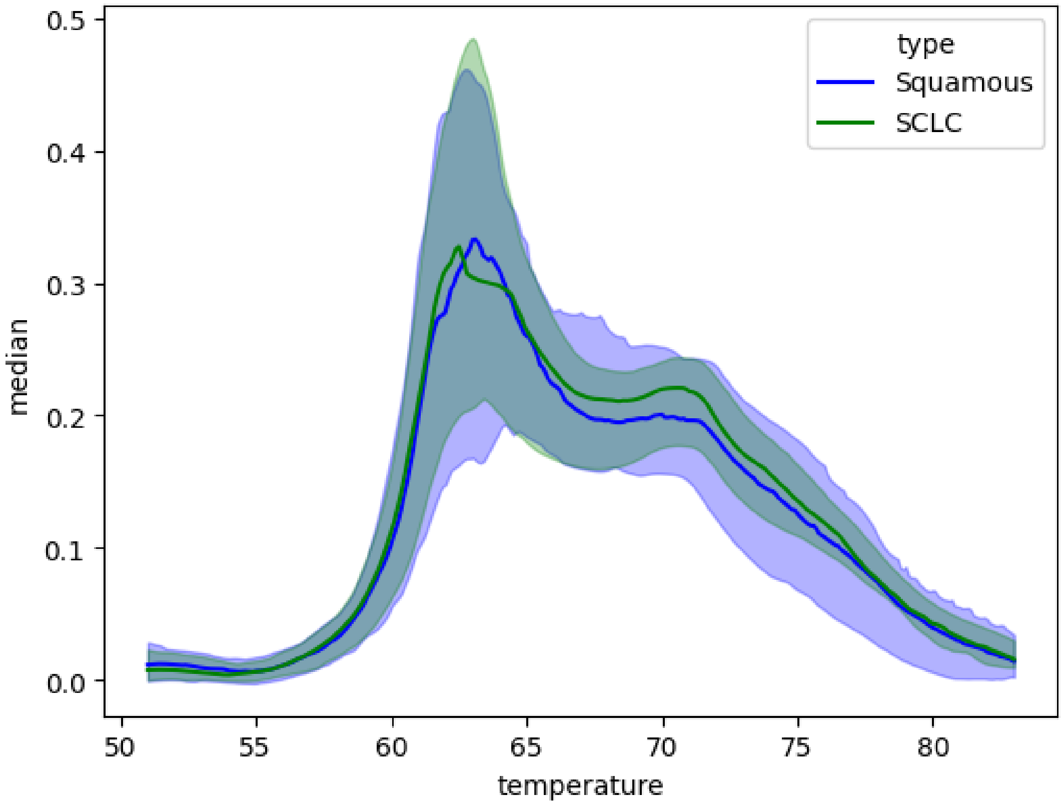


[ ]:
```
adeno_control = pd.read_excel("Adeno_Control.xlsx")

adeno_control['max_depth'] = np.where(pd.isna(adeno_control['max_depth']),␣
  ↪"None", adeno_control["max_depth"])
adeno_control['max_features'] = np.where(pd.
  ↪isna(adeno_control['max_features']), "None", adeno_control["max_features"])

adeno_control_results = adeno_control.groupby(['n_estimators', 'max_depth',␣
  ↪'max_features'], as_index = False).mean().sort_values("Weighted Accuracy",␣
  ↪ascending=False)
```

[ ]:
```
new_row = pd.DataFrame([["Adeno - Control", adeno_control_results.
  ↪iloc[0]["Weighted Accuracy"], adeno_control_results.iloc[0]["AUC"]]],␣
  ↪columns=["Pair", "Weighted Accuracy", "AUC"])

results_table = pd.concat([results_table, new_row], ignore_index=True)
```

[ ]:
```
# sclc, squamous

graph_df = median_df[(median_df["type"] == "Squamous") | (median_df["type"] ==␣
  ↪"SCLC")]
```

```
colors = {'Squamous': 'blue', 'SCLC': 'green'}  # Specify colors for each type

# Create a line plot using Seaborn with matching colors
sns.lineplot(data=graph_df, x='temperature', y='median', hue='type',␣
 ↪palette=colors)

# Create separate ribbons for each "type" with matching colors
for type_name in graph_df['type'].unique():
    type_data = graph_df[graph_df['type'] == type_name]
    plt.fill_between(type_data["temperature"], type_data["lower_q"],␣
 ↪type_data["upper_q"], color=colors[type_name], alpha=0.3, label=type_name)
```



[ ]:
```
squamous_sclc = pd.read_excel("Squamous_SCLC.xlsx")

squamous_sclc['max_depth'] = np.where(pd.isna(squamous_sclc['max_depth']),␣
 ↪"None", squamous_sclc["max_depth"])
squamous_sclc['max_features'] = np.where(pd.
 ↪isna(squamous_sclc['max_features']), "None", squamous_sclc["max_features"])
```

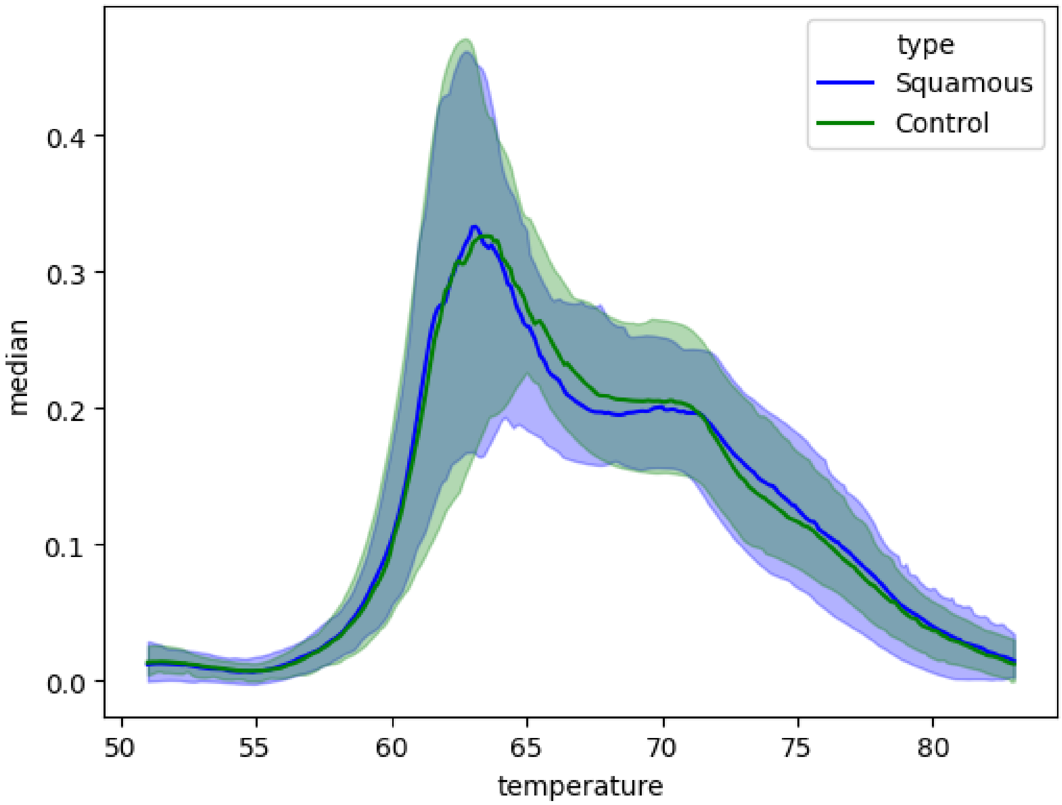
```python
squamous_sclc_results = squamous_sclc.groupby(['n_estimators', 'max_depth', 'max_features'], as_index = False).mean().sort_values("Weighted Accuracy", ascending=False)
```

[ ]:

```python
new_row = pd.DataFrame([["Squamous - SCLC", squamous_sclc_results.iloc[0]["Weighted Accuracy"], squamous_sclc_results.iloc[0]["AUC"]]], columns=["Pair", "Weighted Accuracy", "AUC"])

results_table = pd.concat([results_table, new_row], ignore_index=True)
```

[ ]:

```python
# squamous, control

graph_df = median_df[(median_df["type"] == "Squamous") | (median_df["type"] == "Control")]
colors = {'Squamous': 'blue', 'Control': 'green'}  # Specify colors for each type

# Create a line plot using Seaborn with matching colors
sns.lineplot(data=graph_df, x='temperature', y='median', hue='type', palette=colors)

# Create separate ribbons for each "type" with matching colors
for type_name in graph_df['type'].unique():
    type_data = graph_df[graph_df['type'] == type_name]
    plt.fill_between(type_data["temperature"], type_data["lower_q"], type_data["upper_q"], color=colors[type_name], alpha=0.3, label=type_name)
```



[ ]:

```
squamous_control = pd.read_excel("Control_Squamous.xlsx")

squamous_control['max_depth'] = np.where(pd.
 ↪isna(squamous_control['max_depth']), "None", squamous_control["max_depth"])
squamous_control['max_features'] = np.where(pd.
 ↪isna(squamous_control['max_features']), "None",␣
 ↪squamous_control["max_features"])

squamous_control_results = squamous_control.groupby(['n_estimators',␣
 ↪'max_depth', 'max_features'], as_index = False).mean().sort_values("Weighted␣
 ↪Accuracy", ascending=False)
```

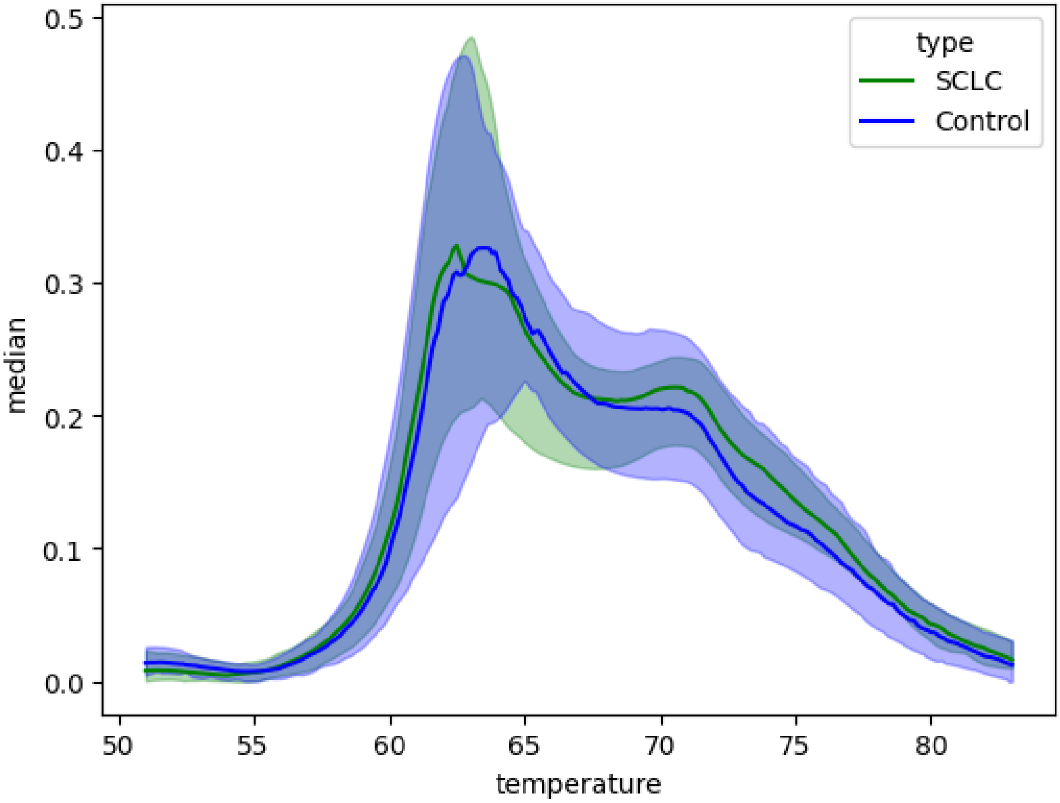
[ ]:

```
new_row = pd.DataFrame([["Squamous - Control", squamous_control_results.
 ↪iloc[0]["Weighted Accuracy"], squamous_control_results.iloc[0]["AUC"]]],␣
 ↪columns=["Pair", "Weighted Accuracy", "AUC"])

results_table = pd.concat([results_table, new_row], ignore_index=True)
```

[ ]:

```
# control, sclc
```

```
graph_df = median_df[(median_df["type"] == "Control") | (median_df["type"] ==␣
 ↪"SCLC")]
colors = {'Control': 'blue', 'SCLC': 'green'}  # Specify colors for each type

# Create a line plot using Seaborn with matching colors
sns.lineplot(data=graph_df, x='temperature', y='median', hue='type',␣
 ↪palette=colors)

# Create separate ribbons for each "type" with matching colors
for type_name in graph_df['type'].unique():
    type_data = graph_df[graph_df['type'] == type_name]
    plt.fill_between(type_data["temperature"], type_data["lower_q"],␣
 ↪type_data["upper_q"], color=colors[type_name], alpha=0.3, label=type_name)
```



```
[ ]: control_sclc = pd.read_excel("Control_SCLC.xlsx")

     control_sclc['max_depth'] = np.where(pd.isna(control_sclc['max_depth']),␣
      ↪"None", control_sclc["max_depth"])
     control_sclc['max_features'] = np.where(pd.isna(control_sclc['max_features']),␣
      ↪"None", control_sclc["max_features"])
```

```
control_sclc_results = control_sclc.groupby(['n_estimators', 'max_depth',␣
 ↪'max_features']).mean().sort_values("Weighted Accuracy", ascending=False)
```

[ ]:
```
new_row = pd.DataFrame([["Control - SCLC", control_sclc_results.
 ↪iloc[0]["Weighted Accuracy"], control_sclc_results.iloc[0]["AUC"]]],␣
 ↪columns=["Pair", "Weighted Accuracy", "AUC"])

results_table = pd.concat([results_table, new_row], ignore_index=True)
results_table
```

[ ]:
```
                 Pair  Weighted Accuracy       AUC
0        Adeno - SCLC           0.530405  0.593018
1    Adeno - Squamous           0.513628  0.524180
2     Adeno - Control           0.575928  0.619652
3     Squamous - SCLC           0.541017  0.598070
4  Squamous - Control           0.588179  0.622351
5      Control - SCLC           0.677083  0.799829
```