Student number: 40462364

Name: Imama Jawad

Course: ECS805

Date: 28/03/2024

By submitting the work, I declare that:

1. I have read and understood the University regulations relating to academic offences, including collusion and plagiarism: http://www.qub.ac.uk/directorates/AcademicStudentAffairs/AcademicAffairs/GeneralRegulations/Procedures/P roceduresforDealingwithAcademicOffences/
2. The submission is my own original work and no part of it has been submitted for any other assignments, except as otherwise permitted.
3. All sources used, published or unpublished, have been acknowledged.
4. I give my consent for the work to be scanned using a plagiarism detection software.

Library Imports

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA, NMF, KernelPCA
from sklearn.manifold import TSNE
from sklearn.preprocessing import LabelEncoder, StandardScaler,
MinMaxScaler
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestClassifier,
BaggingClassifier, RandomForestRegressor
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression, BayesianRidge
from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report
from xgboost import XGBClassifier
from sklearn.datasets import load_iris
from sklearn.experimental import enable_iterative_imputer
from sklearn.impute import IterativeImputer
from mpl_toolkits.mplot3d import Axes3D
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout
from tensorflow.keras.optimizers import Adam
from collections import defaultdict

# Install additional libraries (if not already installed)
```

```
!pip install sweetviz ydata-profiling

# Comprehensive EDA Report
from ydata_profiling import ProfileReport
```

```
Requirement already satisfied: sweetviz in
/usr/local/lib/python3.10/dist-packages (2.3.1)
Requirement already satisfied: pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3
in /usr/local/lib/python3.10/dist-packages (from sweetviz) (2.2.3)
Requirement already satisfied: numpy>=1.16.0 in
/usr/local/lib/python3.10/dist-packages (from sweetviz) (1.26.4)
Requirement already satisfied: matplotlib>=3.1.3 in
/usr/local/lib/python3.10/dist-packages (from sweetviz) (3.7.5)
Requirement already satisfied: tqdm>=4.43.0 in
/usr/local/lib/python3.10/dist-packages (from sweetviz) (4.67.1)
Requirement already satisfied: scipy>=1.3.2 in
/usr/local/lib/python3.10/dist-packages (from sweetviz) (1.13.1)
Requirement already satisfied: jinja2>=2.11.1 in
/usr/local/lib/python3.10/dist-packages (from sweetviz) (3.1.4)
Requirement already satisfied: importlib-resources>=1.2.0 in
/usr/local/lib/python3.10/dist-packages (from sweetviz) (5.13.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2>=2.11.1-
>sweetviz) (3.0.2)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3-
>sweetviz) (1.3.1)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3-
>sweetviz) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3-
>sweetviz) (4.55.3)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3-
>sweetviz) (1.4.7)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3-
>sweetviz) (24.2)
Requirement already satisfied: pillow>=6.2.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3-
>sweetviz) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3-
>sweetviz) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3-
>sweetviz) (2.9.0.post0)
Requirement already satisfied: mkl_fft in
/usr/local/lib/python3.10/dist-packages (from numpy>=1.16.0->sweetviz)
```

```
(1.3.8)
Requirement already satisfied: mkl_random in
/usr/local/lib/python3.10/dist-packages (from numpy>=1.16.0->sweetviz)
(1.2.4)
Requirement already satisfied: mkl_umath in
/usr/local/lib/python3.10/dist-packages (from numpy>=1.16.0->sweetviz)
(0.1.1)
Requirement already satisfied: mkl in /usr/local/lib/python3.10/dist-
packages (from numpy>=1.16.0->sweetviz) (2025.0.1)
Requirement already satisfied: tbb4py in
/usr/local/lib/python3.10/dist-packages (from numpy>=1.16.0->sweetviz)
(2022.0.0)
Requirement already satisfied: mkl-service in
/usr/local/lib/python3.10/dist-packages (from numpy>=1.16.0->sweetviz)
(2.4.1)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas!=1.0.0,!=1.0.1,!
=1.0.2,>=0.25.3->sweetviz) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas!=1.0.0,!=1.0.1,!
=1.0.2,>=0.25.3->sweetviz) (2025.1)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7-
>matplotlib>=3.1.3->sweetviz) (1.17.0)
Requirement already satisfied: intel-openmp>=2024 in
/usr/local/lib/python3.10/dist-packages (from mkl->numpy>=1.16.0-
>sweetviz) (2024.2.0)
Requirement already satisfied: tbb==2022.* in
/usr/local/lib/python3.10/dist-packages (from mkl->numpy>=1.16.0-
>sweetviz) (2022.0.0)
Requirement already satisfied: tcmlib==1.* in
/usr/local/lib/python3.10/dist-packages (from tbb==2022.*->mkl-
>numpy>=1.16.0->sweetviz) (1.2.0)
Requirement already satisfied: intel-cmplr-lib-rt in
/usr/local/lib/python3.10/dist-packages (from mkl_umath-
>numpy>=1.16.0->sweetviz) (2024.2.0)
Requirement already satisfied: intel-cmplr-lib-ur==2024.2.0 in
/usr/local/lib/python3.10/dist-packages (from intel-openmp>=2024->mkl-
>numpy>=1.16.0->sweetviz) (2024.2.0)
Requirement already satisfied: ydata-profiling in
/usr/local/lib/python3.10/dist-packages (4.12.2)
Requirement already satisfied: scipy<1.16,>=1.4.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling)
(1.13.1)
Requirement already satisfied: pandas!=1.4.0,<3,>1.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling) (2.2.3)
Requirement already satisfied: matplotlib>=3.5 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling) (3.7.5)
Requirement already satisfied: pydantic>=2 in
```

```
/usr/local/lib/python3.10/dist-packages (from ydata-profiling)
(2.11.0a2)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling) (6.0.2)
Requirement already satisfied: jinja2<3.2,>=2.11.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling) (3.1.4)
Requirement already satisfied: visions<0.8.0,>=0.7.5 in
/usr/local/lib/python3.10/dist-packages (from
visions[type_image_path]<0.8.0,>=0.7.5->ydata-profiling) (0.7.6)
Requirement already satisfied: numpy<2.2,>=1.16.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling)
(1.26.4)
Requirement already satisfied: htmlmin==0.1.12 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling)
(0.1.12)
Requirement already satisfied: phik<0.13,>=0.11.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling)
(0.12.4)
Requirement already satisfied: requests<3,>=2.24.0 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling)
(2.32.3)
Requirement already satisfied: tqdm<5,>=4.48.2 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling)
(4.67.1)
Requirement already satisfied: seaborn<0.14,>=0.10.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling)
(0.12.2)
Requirement already satisfied: multimethod<2,>=1.4 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling) (1.12)
Requirement already satisfied: statsmodels<1,>=0.13.2 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling)
(0.14.4)
Requirement already satisfied: typeguard<5,>=3 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling) (4.4.1)
Requirement already satisfied: imagehash==4.3.1 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling) (4.3.1)
Requirement already satisfied: wordcloud>=1.9.3 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling) (1.9.4)
Requirement already satisfied: dacite>=1.8 in
/usr/local/lib/python3.10/dist-packages (from ydata-profiling) (1.9.2)
Requirement already satisfied: PyWavelets in
/usr/local/lib/python3.10/dist-packages (from imagehash==4.3.1->ydata-
profiling) (1.8.0)
Requirement already satisfied: pillow in
/usr/local/lib/python3.10/dist-packages (from imagehash==4.3.1->ydata-
profiling) (11.0.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2<3.2,>=2.11.1-
>ydata-profiling) (3.0.2)
```

```
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.5->ydata-
profiling) (1.3.1)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.5->ydata-
profiling) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.5->ydata-
profiling) (4.55.3)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.5->ydata-
profiling) (1.4.7)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.5->ydata-
profiling) (24.2)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.5->ydata-
profiling) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>=3.5->ydata-
profiling) (2.9.0.post0)
Requirement already satisfied: mkl_fft in
/usr/local/lib/python3.10/dist-packages (from numpy<2.2,>=1.16.0-
>ydata-profiling) (1.3.8)
Requirement already satisfied: mkl_random in
/usr/local/lib/python3.10/dist-packages (from numpy<2.2,>=1.16.0-
>ydata-profiling) (1.2.4)
Requirement already satisfied: mkl_umath in
/usr/local/lib/python3.10/dist-packages (from numpy<2.2,>=1.16.0-
>ydata-profiling) (0.1.1)
Requirement already satisfied: mkl in /usr/local/lib/python3.10/dist-
packages (from numpy<2.2,>=1.16.0->ydata-profiling) (2025.0.1)
Requirement already satisfied: tbb4py in
/usr/local/lib/python3.10/dist-packages (from numpy<2.2,>=1.16.0-
>ydata-profiling) (2022.0.0)
Requirement already satisfied: mkl-service in
/usr/local/lib/python3.10/dist-packages (from numpy<2.2,>=1.16.0-
>ydata-profiling) (2.4.1)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas!=1.4.0,<3,>1.1-
>ydata-profiling) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas!=1.4.0,<3,>1.1-
>ydata-profiling) (2025.1)
Requirement already satisfied: joblib>=0.14.1 in
/usr/local/lib/python3.10/dist-packages (from phik<0.13,>=0.11.1-
>ydata-profiling) (1.4.2)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-
```

```
profiling) (0.7.0)
Requirement already satisfied: pydantic-core==2.29.0 in
/usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-
profiling) (2.29.0)
Requirement already satisfied: typing-extensions>=4.12.2 in
/usr/local/lib/python3.10/dist-packages (from pydantic>=2->ydata-
profiling) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0-
>ydata-profiling) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0-
>ydata-profiling) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0-
>ydata-profiling) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests<3,>=2.24.0-
>ydata-profiling) (2025.1.31)
Requirement already satisfied: patsy>=0.5.6 in
/usr/local/lib/python3.10/dist-packages (from statsmodels<1,>=0.13.2-
>ydata-profiling) (1.0.1)
Requirement already satisfied: attrs>=19.3.0 in
/usr/local/lib/python3.10/dist-packages (from visions<0.8.0,>=0.7.5-
>visions[type_image_path]<0.8.0,>=0.7.5->ydata-profiling) (25.1.0)
Requirement already satisfied: networkx>=2.4 in
/usr/local/lib/python3.10/dist-packages (from visions<0.8.0,>=0.7.5-
>visions[type_image_path]<0.8.0,>=0.7.5->ydata-profiling) (3.4.2)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7-
>matplotlib>=3.5->ydata-profiling) (1.17.0)
Requirement already satisfied: intel-openmp>=2024 in
/usr/local/lib/python3.10/dist-packages (from mkl->numpy<2.2,>=1.16.0-
>ydata-profiling) (2024.2.0)
Requirement already satisfied: tbb==2022.* in
/usr/local/lib/python3.10/dist-packages (from mkl->numpy<2.2,>=1.16.0-
>ydata-profiling) (2022.0.0)
Requirement already satisfied: tcmlib==1.* in
/usr/local/lib/python3.10/dist-packages (from tbb==2022.*->mkl-
>numpy<2.2,>=1.16.0->ydata-profiling) (1.2.0)
Requirement already satisfied: intel-cmplr-lib-rt in
/usr/local/lib/python3.10/dist-packages (from mkl_umath-
>numpy<2.2,>=1.16.0->ydata-profiling) (2024.2.0)
Requirement already satisfied: intel-cmplr-lib-ur==2024.2.0 in
/usr/local/lib/python3.10/dist-packages (from intel-openmp>=2024->mkl-
>numpy<2.2,>=1.16.0->ydata-profiling) (2024.2.0)


import os
for dirname, _, filenames in os.walk('/kaggle/input'):
```

```
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
/kaggle/input/paccanatlas/
PanCanAtlas_9126RNASeqSamplesWithImmuneSubtypes_440Genes_SampleIdsOrde
red_SampleIdWithSubtypes_RR020718_RownamesGenesWithSignature.csv
/kaggle/input/paccanatlas/1-s2.0-S1074761318301213-mmc2.xlsx
/kaggle/input/ai-for-health2/Groups_7302_80Percent_from_Original.csv
/kaggle/input/ai-for-health2/EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNAS
eqV2.geneExp.tsv
```

Section 1- Data Preparation and Pre-processing

Objective: Prepare the RNA-seq and immune subtype data by cleaning, filtering, and normalizing the datasets.

```
# Immune Subtypes Dataset
file_pathcolumns =
"/kaggle/input/paccanatlas/PanCanAtlas_9126RNASeqSamplesWithImmuneSubt
ypes_440Genes_SampleIdsOrdered_SampleIdWithSubtypes_RR020718_RownamesG
enesWithSignature.csv"

df=pd.read_csv(file_pathcolumns,index_col=0)
df_T = df.T
df_T.head()
```

```
                      ACTL6A_S5  ADAM9_S2  ADAMTS1_S5  ADCY7_S3
AIMP2_S5  \
TCGA.02.0047.GBM.C4    745.567   4287.78     241.556  1067.640
406.736
TCGA.02.0055.GBM.C4   1154.310   9475.54    6098.950   556.132
537.088
TCGA.02.2483.GBM.C4   1498.680   2307.12     433.984   497.309
752.148
TCGA.02.2485.GBM.C4   1320.000   2685.71     911.905   316.667
785.552
TCGA.02.2486.GBM.C4   1404.270   2843.90     321.951   637.805
792.963

                      ALKBH7_S5  ALOX5AP_S3  AMPD3_S3  APITD1_S5
APOC1_S3  \
TCGA.02.0047.GBM.C4    518.148    1326.410   326.992    184.308
1370.66
TCGA.02.0055.GBM.C4    942.957    4211.350   361.598    319.535
3093.48
TCGA.02.2483.GBM.C4    656.042     566.543   196.728    311.443
3504.38
TCGA.02.2485.GBM.C4    953.809     307.143    80.000    260.462
```

```
2482.86
TCGA.02.2486.GBM.C4    815.244    5671.950   542.683    494.488
12512.80

                      ...  VTA1_S5  WDHD1_S5  WDR54_S5  WDR77_S5
WIPF1_S3  \
TCGA.02.0047.GBM.C4   ...  1230.52  107.5630   321.460   925.044
2154.950
TCGA.02.0055.GBM.C4   ...  1183.46  162.5790   585.844   931.184
1947.580
TCGA.02.2483.GBM.C4   ...  1514.30  267.2300   561.055  1370.770
1042.740
TCGA.02.2485.GBM.C4   ...  1577.62  309.5240   422.381   811.905
849.524
TCGA.02.2486.GBM.C4   ...  1484.76   45.7317   669.512   954.878
2454.880

                     WNT2B_S2  WNT8B_S2  WSB2_S5  ZWILCH_S5    ZYX_S2

TCGA.02.0047.GBM.C4   43.0253    0.0000  4291.47    260.211   3867.97

TCGA.02.0055.GBM.C4   44.8493    0.0000  2494.18    467.560  11481.40

TCGA.02.2483.GBM.C4   10.1319    0.8443  4128.76    682.242   5066.81

TCGA.02.2485.GBM.C4   18.5714    0.9524  3609.52    440.386   5090.95

TCGA.02.2486.GBM.C4    7.3171    0.6098  2747.56    166.689   7800.00


[5 rows x 440 columns]
```

```
print(df_T.shape[0],"Rows-",df_T.shape[1],"Columns" )
```

```
9126 Rows- 440 Columns
```

```
# Original Whole Tsv dataset unfiltered
file_geneexpressions =
"/kaggle/input/ai-for-health2/EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNA
SeqV2.geneExp.tsv"

df_geneExpressions=pd.read_csv(file_geneexpressions, sep="\
t",index_col=0)

df_geneExpressions_T=df_geneExpressions.T
df_geneExpressions_T.head()
```

```
gene_id                         ?|100130426  ?|100133144  ?|
100134869   ?|10357  \
TCGA-OR-A5J1-01A-11R-A29S-07          0.0       3.2661       3.9385
149.1350
```

```
TCGA-OR-A5J2-01A-11R-A29S-07          0.0       2.6815       8.9948
81.0777
TCGA-OR-A5J3-01A-11R-A29S-07          0.0       1.7301       6.5650
86.4879
TCGA-OR-A5J5-01A-11R-A29S-07          0.0       0.0000       1.5492
53.9117
TCGA-OR-A5J6-01A-31R-A29S-07          0.0       0.0000       4.4709
66.9063

gene_id                       ?|10431  ?|136542  ?|155060  ?|26823  ?|
280660  \
TCGA-OR-A5J1-01A-11R-A29S-07  2034.10       0.0   274.255   1.4409
0.0
TCGA-OR-A5J2-01A-11R-A29S-07  1304.93       0.0   199.302   0.0000
0.0
TCGA-OR-A5J3-01A-11R-A29S-07  1054.66       0.0   348.393   0.5925
0.0
TCGA-OR-A5J5-01A-11R-A29S-07  2350.89       0.0   439.194   0.7746
0.0
TCGA-OR-A5J6-01A-31R-A29S-07  1257.99       0.0   149.215   0.0000
0.0

gene_id                       ?|317712  ...  ZWILCH|55055  ZWINT|11130
\
TCGA-OR-A5J1-01A-11R-A29S-07       0.0  ...      183.9580     146.9740

TCGA-OR-A5J2-01A-11R-A29S-07       0.0  ...      264.1250     438.4640

TCGA-OR-A5J3-01A-11R-A29S-07       0.0  ...       90.6532     190.1940

TCGA-OR-A5J5-01A-11R-A29S-07       0.0  ...      260.2630     840.4340

TCGA-OR-A5J6-01A-31R-A29S-07       0.0  ...       64.2685      63.1508


gene_id                       ZXDA|7789  ZXDB|158586  ZXDC|79364  \
TCGA-OR-A5J1-01A-11R-A29S-07    20.6532      351.585    1050.910
TCGA-OR-A5J2-01A-11R-A29S-07    57.1735      279.023    2979.870
TCGA-OR-A5J3-01A-11R-A29S-07    39.6978      495.334     914.827
TCGA-OR-A5J5-01A-11R-A29S-07    17.8156      206.042     890.782
TCGA-OR-A5J6-01A-31R-A29S-07    12.8537      117.919     894.730

gene_id                       ZYG11A|440590  ZYG11B|79699  ZYX|7791  \
TCGA-OR-A5J1-01A-11R-A29S-07         0.4803       648.415  1841.020
TCGA-OR-A5J2-01A-11R-A29S-07        31.4052      1166.020  3059.990
TCGA-OR-A5J3-01A-11R-A29S-07         0.5925       806.399  2655.610
TCGA-OR-A5J5-01A-11R-A29S-07        11.6189       553.834  2367.930
TCGA-OR-A5J6-01A-31R-A29S-07         7.8240       795.812   708.071

gene_id                       ZZEF1|23140  ZZZ3|26009
```

```
TCGA-OR-A5J1-01A-11R-A29S-07     1157.540     596.062
TCGA-OR-A5J2-01A-11R-A29S-07     1895.990     801.637
TCGA-OR-A5J3-01A-11R-A29S-07     1482.450     437.269
TCGA-OR-A5J5-01A-11R-A29S-07     1140.200     512.781
TCGA-OR-A5J6-01A-31R-A29S-07      796.371     475.587

[5 rows x 20531 columns]
```

```
print(df_geneExpressions_T.shape[0],"Rows,",df_geneExpressions_T.shape
[1],"Columns")
```

```
11069 Rows, 20531 Columns
```

```
df_T.columns
```

```
Index(['ACTL6A_S5', 'ADAM9_S2', 'ADAMTS1_S5', 'ADCY7_S3', 'AIMP2_S5',
       'ALKBH7_S5', 'ALOX5AP_S3', 'AMPD3_S3', 'APITD1_S5', 'APOC1_S3',
       ...
       'VTA1_S5', 'WDHD1_S5', 'WDR54_S5', 'WDR77_S5', 'WIPF1_S3',
'WNT2B_S2',
       'WNT8B_S2', 'WSB2_S5', 'ZWILCH_S5', 'ZYX_S2'],
      dtype='object', length=440)
```

```
df_geneExpressions_T.columns
```

```
Index(['?|100130426', '?|100133144', '?|100134869', '?|10357', '?|
10431',
       '?|136542', '?|155060', '?|26823', '?|280660', '?|317712',
       ...
       'ZWILCH|55055', 'ZWINT|11130', 'ZXDA|7789', 'ZXDB|158586',
'ZXDC|79364',
       'ZYG11A|440590', 'ZYG11B|79699', 'ZYX|7791', 'ZZEF1|23140',
       'ZZZ3|26009'],
      dtype='object', name='gene_id', length=20531)
```

Extracting genotypes information relevant to immune subtype sampling

```
#Filter the columns found in df_T in the tsv to extract genes relevant
to immune subtype genes
#Extract prefixes before '|' from df_geneExpressions_T columns
prefixes_df1 = df_geneExpressions_T.columns.str.split('|').str[0]

# Extract prefixes before '_' from df_T columns
prefixes_df2 = df_T.columns.str.split('_').str[0]

# Find the common prefixes
common_prefixes = prefixes_df1.intersection(prefixes_df2)

# Filter df_geneExpressions_T to keep only matching columns
```

```
df_filtered = df_geneExpressions_T.loc[:,
prefixes_df1.isin(common_prefixes)]

# Rename df_filtered columns to match naming convention from df_T
rename_mapping = {col: df_T.columns[prefixes_df2 == col.split('|')
[0]].values[0] for col in df_filtered.columns}
df_filtered = df_filtered.rename(columns=rename_mapping)

df_filtered.head()
```

```
gene_id                       ACTL6A_S5  ADAM9_S2  ADAMTS1_S5
ADCY7_S3  \
TCGA-OR-A5J1-01A-11R-A29S-07    829.011   542.267    377.0410
183.477
TCGA-OR-A5J2-01A-11R-A29S-07    935.310  1612.130   1453.5000
281.036
TCGA-OR-A5J3-01A-11R-A29S-07    756.036  1101.470    777.9590
117.316
TCGA-OR-A5J5-01A-11R-A29S-07    512.006   431.449     48.7994
333.075
TCGA-OR-A5J6-01A-31R-A29S-07    410.760  2296.900    310.7240
250.368

gene_id                       AIMP2_S5  ALKBH7_S5  ALOX5AP_S3
AMPD3_S3  \
TCGA-OR-A5J1-01A-11R-A29S-07   338.823    974.544     83.5735
163.785
TCGA-OR-A5J2-01A-11R-A29S-07  1001.020    812.508      9.2605
332.975
TCGA-OR-A5J3-01A-11R-A29S-07  1081.980   1422.010     53.3254
347.208
TCGA-OR-A5J5-01A-11R-A29S-07   198.474   1307.510     10.0697
115.414
TCGA-OR-A5J6-01A-31R-A29S-07   477.403   1442.970    326.3720
939.997

gene_id                       APITD1_S5   APOC1_S3  ...   VTA1_S5
WDHD1_S5  \
TCGA-OR-A5J1-01A-11R-A29S-07   321.7190   1875.120  ...  1124.880
64.3612
TCGA-OR-A5J2-01A-11R-A29S-07    61.1998    237.552  ...  1008.190
162.6630
TCGA-OR-A5J3-01A-11R-A29S-07   210.2740   4049.180  ...   537.994
183.6770
TCGA-OR-A5J5-01A-11R-A29S-07   208.9080    646.785  ...   455.461
192.8740
TCGA-OR-A5J6-01A-31R-A29S-07   225.7780  79054.200  ...   989.176
21.2366

gene_id                        WDR54_S5  WDR77_S5  WIPF1_S3  WNT2B_S2
```

```
\
TCGA-OR-A5J1-01A-11R-A29S-07   411.6230   951.489   571.086   11.5274

TCGA-OR-A5J2-01A-11R-A29S-07    64.0182   805.663   584.217   10.8710

TCGA-OR-A5J3-01A-11R-A29S-07  1183.8200   865.649   158.791    0.5925

TCGA-OR-A5J5-01A-11R-A29S-07   509.6820   574.748   343.919   17.0411

TCGA-OR-A5J6-01A-31R-A29S-07   483.9700   582.888   393.435   11.1771


gene_id                         WNT8B_S2  WSB2_S5  ZWILCH_S5    ZYX_S2
TCGA-OR-A5J1-01A-11R-A29S-07      0.0000  1731.51   183.9580  1841.020
TCGA-OR-A5J2-01A-11R-A29S-07      0.4026  5007.11   264.1250  3059.990
TCGA-OR-A5J3-01A-11R-A29S-07      3.5550  2508.67    90.6532  2655.610
TCGA-OR-A5J5-01A-11R-A29S-07      0.7746  1889.23   260.2630  2367.930
TCGA-OR-A5J6-01A-31R-A29S-07      0.0000  3739.87    64.2685   708.071

[5 rows x 440 columns]
```

```
print("Rows:", df_filtered.shape[0], "Columns:" ,df_filtered.shape[1])
```

```
Rows: 11069 Columns: 440
```

Extracting relevant subtypes for Samples C1 to C6

```
df_T.index
```

```
Index(['TCGA.02.0047.GBM.C4', 'TCGA.02.0055.GBM.C4',
'TCGA.02.2483.GBM.C4',
       'TCGA.02.2485.GBM.C4', 'TCGA.02.2486.GBM.C4',
'TCGA.04.1348.OV.C2',
       'TCGA.04.1357.OV.C2', 'TCGA.04.1362.OV.C4',
'TCGA.04.1364.OV.C1',
       'TCGA.04.1365.OV.C2',
       ...
       'TCGA.ZP.A9D2.LIHC.C2', 'TCGA.ZP.A9D4.LIHC.C4',
'TCGA.ZQ.A9CR.STAD.C1',
       'TCGA.ZR.A9CJ.ESCA.C1', 'TCGA.ZS.A9CD.LIHC.C3',
'TCGA.ZS.A9CE.LIHC.C4',
       'TCGA.ZS.A9CF.LIHC.C4', 'TCGA.ZS.A9CG.LIHC.C3',
'TCGA.ZU.A8S4.CHOL.C1',
       'TCGA.ZX.AA5X.CESC.C2'],
      dtype='object', length=9126)
```

```
df_filtered.index
```

```
Index(['TCGA-OR-A5J1-01A-11R-A29S-07', 'TCGA-OR-A5J2-01A-11R-A29S-07',
       'TCGA-OR-A5J3-01A-11R-A29S-07', 'TCGA-OR-A5J5-01A-11R-A29S-07',
       'TCGA-OR-A5J6-01A-31R-A29S-07', 'TCGA-OR-A5J7-01A-11R-A29S-07',
```

```
       'TCGA-OR-A5J8-01A-11R-A29S-07', 'TCGA-OR-A5J9-01A-11R-A29S-07',
       'TCGA-OR-A5JA-01A-11R-A29S-07', 'TCGA-OR-A5JB-01A-11R-A29S-07',
       ...
       'TCGA-CG-4449-01A-01R-1157-13', 'TCGA-CG-4462-01A-01R-1157-13',
       'TCGA-CG-4465-01A-01R-1157-13', 'TCGA-CG-4466-01A-01R-1157-13',
       'TCGA-CG-4469-01A-01R-1157-13', 'TCGA-CG-4472-01A-01R-1157-13',
       'TCGA-CG-4474-01A-02R-1157-13', 'TCGA-CG-4475-01A-01R-1157-13',
       'TCGA-CG-4476-01A-01R-1157-13', 'TCGA-CG-4477-01A-01R-1157-
13'],
      dtype='object', length=11069)

# Extract first three segments from the index
df_filtered_index_trimmed =
df_filtered.index.str.split('-').str[:3].str.join('-')
df_T_index_trimmed = df_T.index.str.split('.').str[:3].str.join('-')

# Subset df_filtered based on matching first three segments
df_filtered_subset =
df_filtered.loc[df_filtered_index_trimmed.isin(df_T_index_trimmed)]

df_filtered_subset.head()
```

```
gene_id                       ACTL6A_S5  ADAM9_S2  ADAMTS1_S5
ADCY7_S3  \
TCGA-OR-A5J1-01A-11R-A29S-07    829.011   542.267    377.0410
183.477
TCGA-OR-A5J2-01A-11R-A29S-07    935.310  1612.130   1453.5000
281.036
TCGA-OR-A5J3-01A-11R-A29S-07    756.036  1101.470    777.9590
117.316
TCGA-OR-A5J5-01A-11R-A29S-07    512.006   431.449     48.7994
333.075
TCGA-OR-A5J6-01A-31R-A29S-07    410.760  2296.900    310.7240
250.368

gene_id                       AIMP2_S5  ALKBH7_S5  ALOX5AP_S3
AMPD3_S3  \
TCGA-OR-A5J1-01A-11R-A29S-07   338.823    974.544     83.5735
163.785
TCGA-OR-A5J2-01A-11R-A29S-07  1001.020    812.508      9.2605
332.975
TCGA-OR-A5J3-01A-11R-A29S-07  1081.980   1422.010     53.3254
347.208
TCGA-OR-A5J5-01A-11R-A29S-07   198.474   1307.510     10.0697
115.414
TCGA-OR-A5J6-01A-31R-A29S-07   477.403   1442.970    326.3720
939.997

gene_id                       APITD1_S5   APOC1_S3  ...   VTA1_S5
WDHD1_S5  \
```

```
TCGA-OR-A5J1-01A-11R-A29S-07   321.7190   1875.120  ...  1124.880
64.3612
TCGA-OR-A5J2-01A-11R-A29S-07    61.1998    237.552  ...  1008.190
162.6630
TCGA-OR-A5J3-01A-11R-A29S-07   210.2740   4049.180  ...   537.994
183.6770
TCGA-OR-A5J5-01A-11R-A29S-07   208.9080    646.785  ...   455.461
192.8740
TCGA-OR-A5J6-01A-31R-A29S-07   225.7780  79054.200  ...   989.176
21.2366

gene_id                         WDR54_S5  WDR77_S5  WIPF1_S3  WNT2B_S2
\
TCGA-OR-A5J1-01A-11R-A29S-07   411.6230   951.489   571.086   11.5274

TCGA-OR-A5J2-01A-11R-A29S-07    64.0182   805.663   584.217   10.8710

TCGA-OR-A5J3-01A-11R-A29S-07  1183.8200   865.649   158.791    0.5925

TCGA-OR-A5J5-01A-11R-A29S-07   509.6820   574.748   343.919   17.0411

TCGA-OR-A5J6-01A-31R-A29S-07   483.9700   582.888   393.435   11.1771


gene_id                        WNT8B_S2  WSB2_S5  ZWILCH_S5    ZYX_S2
TCGA-OR-A5J1-01A-11R-A29S-07     0.0000  1731.51   183.9580  1841.020
TCGA-OR-A5J2-01A-11R-A29S-07     0.4026  5007.11   264.1250  3059.990
TCGA-OR-A5J3-01A-11R-A29S-07     3.5550  2508.67    90.6532  2655.610
TCGA-OR-A5J5-01A-11R-A29S-07     0.7746  1889.23   260.2630  2367.930
TCGA-OR-A5J6-01A-31R-A29S-07     0.0000  3739.87    64.2685   708.071

[5 rows x 440 columns]
```

```
print(df_filtered_subset.shape[0],df_filtered_subset.shape[1])
```

```
9879 440
```

```
print(df_T.shape[0],df_T.shape[1])
```

```
9126 440
```

```
df_filtered_subset.index
```

```
Index(['TCGA-OR-A5J1-01A-11R-A29S-07', 'TCGA-OR-A5J2-01A-11R-A29S-07',
       'TCGA-OR-A5J3-01A-11R-A29S-07', 'TCGA-OR-A5J5-01A-11R-A29S-07',
       'TCGA-OR-A5J6-01A-31R-A29S-07', 'TCGA-OR-A5J7-01A-11R-A29S-07',
       'TCGA-OR-A5J8-01A-11R-A29S-07', 'TCGA-OR-A5J9-01A-11R-A29S-07',
       'TCGA-OR-A5JA-01A-11R-A29S-07', 'TCGA-OR-A5JB-01A-11R-A29S-07',
       ...
       'TCGA-CG-4449-01A-01R-1157-13', 'TCGA-CG-4462-01A-01R-1157-13',
       'TCGA-CG-4465-01A-01R-1157-13', 'TCGA-CG-4466-01A-01R-1157-13',
```

```
       'TCGA-CG-4469-01A-01R-1157-13', 'TCGA-CG-4472-01A-01R-1157-13',
       'TCGA-CG-4474-01A-02R-1157-13', 'TCGA-CG-4475-01A-01R-1157-13',
       'TCGA-CG-4476-01A-01R-1157-13', 'TCGA-CG-4477-01A-01R-1157-
13'],
      dtype='object', length=9879)
```

Extracting Primary solid tumors based on 01 in 4th component in sample TCGA Barcode

The study is based only on primary solid tumors

```
# Extracting Primary solid tumors based on 01 in 4th component in
sample TCGA Barcode as the study is based only on primary solid tumors
# Assuming df_filtered_subset.index is a list of row names
df_filtered_subset_index = df_filtered_subset.index  # Replace with
your DataFrame index

# Sample valid row names

# Function to check if row name is valid
def is_valid_row_name(name):
    parts = name.split('-')
    if len(parts) > 3:  # Ensure the row name has enough parts
        number = parts[3][:2]  # Extract the 4th component and take
the first two characters
        return number.isdigit() and  int(number)==1
    return False

# Filter valid row names
valid_rows = [name for name in df_filtered_subset_index if
is_valid_row_name(name)]

# Check if row names in df_filtered_subset.index are in valid row
names
rows_in_valid_names = [row for row in df_filtered_subset_index if row
in valid_rows]

# Filter the DataFrame to include only the rows with valid row names
df_filtered_valid_rows = df_filtered_subset.loc[rows_in_valid_names]

# Show the filtered DataFrame
df_filtered_valid_rows.shape[0]
```

```
9130
```

Handling Duplicate TCGA Samples

General Rules for Selecting Representative Samples

When dealing with duplicate TCGA sample barcodes, we follow these selection rules:

1. **Primary tumor samples (01/TP) take priority** over other sample types.
2. **Metastatic samples (06/TM) are selected over additional metastatic (07/TAM)** when primary tumors are unavailable.
3. **If multiple primary tumor samples exist**, follow the BCR annotations to select the recommended sample.
4. **When BCR annotations are missing**, select the first available aliquot (`01R`) unless a specific second aliquot (`02R`) is chosen in BCR.

Specific Recommendations for Our Cases

1️⃣ TCGA-21-1076 (Both 01A Samples)

- Both are **Primary Tumor (01/TP)**
- **Selection:** `TCGA-21-1076-01A-01R-0692-07`
- **Rationale:** First aliquot (`01R`) is preferred when both are primary tumors.

2️⃣ TCGA-06-0156 (Both 01A Samples)

- Both are **Primary Tumor (01/TP)**
- **Selection:** `TCGA-06-0156-01A-02R-1849-01`
- **Rationale:** BCR annotation specifically selected second aliquot (`02R`).

3️⃣ TCGA-06-0211 (01A vs 01B)

- `01A`: **Primary Tumor (01/TP)**
- `01B`: **Additional New Primary (05/TAP)**
- **Selection:** `TCGA-06-0211-01A-01R-1849-01`
- **Rationale:** Primary tumor (`01A`) is preferred over "New Primary" (`05/TAP`).

4️⃣ TCGA-23-1023 (01A vs 01R)

- `01A`: **Primary Tumor (01/TP)**
- `01R`: **Human Tumor Original Cells (08/THOC)**
- **Selection:** `TCGA-23-1023-01A-02R-1564-13`
- **Rationale:** Primary tumor (`01A`) takes precedence over cell cultures.

```
#Checking for duplicates , handling as menitonned in paper
# Get unique TCGA sample barcodes
unique_sample_barcodes = df_filtered_valid_rows.index.unique()

# Extract first three parts of each unique TCGA sample barcode
```

```python
first_three_parts_set = {"-".join(sample_id.split("-")[:3]) for
sample_id in unique_sample_barcodes}

print("Number of unique first-three-part TCGA sample barcodes:",
len(first_three_parts_set))

# Dictionary to store full TCGA sample barcodes for each first-three-
part key
sample_barcode_dict = defaultdict(list)

# Extract first three parts and store full TCGA sample barcodes
for sample_id in unique_sample_barcodes:
    key = "-".join(sample_id.split("-")[:3])  # Extract first three
parts
    sample_barcode_dict[key].append(sample_id)

# Filter only the groups where the first three parts appear more than
once
duplicates = {key: val for key, val in sample_barcode_dict.items() if
len(val) > 1}

# Print results
print("Number of duplicated first-three-part TCGA sample barcodes:",
len(duplicates))
for key, sample_ids in duplicates.items():
    print(f"Duplicate Group ({key}):")
    print("\n".join(sample_ids))  # Print all full sample barcodes
within this duplicate group
    print("-" * 50)
```

```
Number of unique first-three-part TCGA sample barcodes: 9126
Number of duplicated first-three-part TCGA sample barcodes: 4
Duplicate Group (TCGA-21-1076):
TCGA-21-1076-01A-01R-0692-07
TCGA-21-1076-01A-02R-0692-07
--------------------------------------------------
Duplicate Group (TCGA-06-0156):
TCGA-06-0156-01A-02R-1849-01
TCGA-06-0156-01A-03R-1849-01
--------------------------------------------------
Duplicate Group (TCGA-06-0211):
TCGA-06-0211-01A-01R-1849-01
TCGA-06-0211-01B-01R-1849-01
--------------------------------------------------
Duplicate Group (TCGA-23-1023):
TCGA-23-1023-01A-02R-1564-13
TCGA-23-1023-01R-01R-1564-13
--------------------------------------------------
```

```
# List of TCGA sample IDs we want to remove
samples_to_remove = [
    "TCGA-21-1076-01A-02R",
    "TCGA-06-0156-01A-03R",
    "TCGA-06-0211-01B-01R",
    "TCGA-23-1023-01R-01R"
]

# Convert index to string to ensure proper matching
df_filtered_valid_rows.index =
df_filtered_valid_rows.index.astype(str)
# Use .str.contains() to remove indices that contain any of the
unwanted sample IDs
mask =
~df_filtered_valid_rows.index.str.contains("|".join(samples_to_remove)
, regex=True)

# Apply the mask to filter out the unwanted samples
df_filtered_valid_rows = df_filtered_valid_rows[mask]

# Display the number of samples after filtering
print(f"Final selected samples count: {len(df_filtered_valid_rows)}")

# Preview the final dataset
df_filtered_valid_rows.head()
print(df_filtered_valid_rows.shape[0])
```

```
Final selected samples count: 9126
9126
```

```
df_filtered_valid_rows = df_filtered_valid_rows.reset_index()
```

Joining Subtype Information

```
# Loading Table S1 from the research paper
file_pathcolumnsSample = "/kaggle/input/paccanatlas/1-s2.0-
S1074761318301213-mmc2.xlsx"

df_Sample=pd.read_excel(file_pathcolumnsSample)
df_Sample_T = df_Sample
df_Sample_T.head()
```

```
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/
format.py:1458: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
```

```
59: RuntimeWarning: invalid value encountered in greater
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
58: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in greater
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()

  TCGA Participant Barcode TCGA Study Immune Subtype          TCGA
Subtype  \
0             TCGA-01-0639         OV             NaN
NaN
1             TCGA-02-0007        GBM             NaN  GBM_LGG.Classic-
like
2             TCGA-02-0011        GBM             NaN
GBM_LGG.LGm6-GBM
3             TCGA-02-0023        GBM             NaN
GBM_LGG.
4             TCGA-02-0025        GBM             NaN
GBM_LGG.

   Leukocyte Fraction  Stromal Fraction  Intratumor Heterogeneity  \
0                 NaN               NaN                       NaN
1            0.047660               NaN                       NaN
2            0.062099               NaN                       NaN
3                 NaN               NaN                       NaN
4                 NaN               NaN                       NaN

   TIL Regional Fraction  Proliferation  Wound Healing  ...  T Cells
CD8  \
0                    NaN            NaN            NaN  ...
NaN
1                    NaN            NaN            NaN  ...
NaN
2                    NaN            NaN            NaN  ...
NaN
3                    NaN            NaN            NaN  ...
NaN
4                    NaN            NaN            NaN  ...
NaN

   T Cells Follicular Helper  T Cells gamma delta  T Cells Regulatory
Tregs  \
```

```
0                          NaN                    NaN
NaN
1                          NaN                    NaN
NaN
2                          NaN                    NaN
NaN
3                          NaN                    NaN
NaN
4                          NaN                    NaN
NaN

   Lymphocytes  Neutrophils.1  Eosinophils.1  Mast Cells  Dendritic
Cells  \
0          NaN            NaN            NaN         NaN
NaN
1          NaN            NaN            NaN         NaN
NaN
2          NaN            NaN            NaN         NaN
NaN
3          NaN            NaN            NaN         NaN
NaN
4          NaN            NaN            NaN         NaN
NaN

   Macrophages
0          NaN
1          NaN
2          NaN
3          NaN
4          NaN

[5 rows x 64 columns]
```

```
print(df_Sample_T.shape[0],"Rows-",df_Sample_T.shape[1],"Columns" )
```

```
11080 Rows- 64 Columns
```

```
# Viewing immune subtypes which are Not NA
df_Sample_NonNa = df_Sample_T[df_Sample_T["Immune Subtype"].notna()]
df_Sample_NonNa.head()
```

```
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/
format.py:1458: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in greater
```

```
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
58: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in greater
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()

   TCGA Participant Barcode TCGA Study Immune Subtype          TCGA
Subtype  \
9              TCGA-02-2485        GBM             C4
GBM_LGG.Classic-like
13             TCGA-05-4410       LUAD             C3
LUAD.3
14             TCGA-05-4424       LUAD             C2
LUAD.5
15             TCGA-05-4425       LUAD             C6
LUAD.4
17             TCGA-06-0156        GBM             C4
GBM_LGG.

    Leukocyte Fraction  Stromal Fraction  Intratumor Heterogeneity  \
9             0.049501              0.13                      0.03
13            0.428618              0.76                      0.39
14            0.326200              0.76                      0.45
15            0.228394              0.39                      0.08
17                 NaN               NaN                       NaN

    TIL Regional Fraction  Proliferation  Wound Healing  ...  T Cells
CD8  \
9                     NaN         0.3223          0.150  ...
0.015523
13               0.000000        -0.2923         -0.166  ...
0.109387
14               4.438672         0.3447          0.020  ...
0.055662
15               0.000000        -0.0918         -0.017  ...
0.052170
17                    NaN         0.4888          0.138  ...
0.096720

    T Cells Follicular Helper  T Cells gamma delta  T Cells Regulatory
Tregs  \
9                    0.198057             0.054341
```

```
0.000000
13                   0.013077              0.000000
0.013376
14                   0.026872              0.000000
0.008113
15                   0.014315              0.000000
0.050522
17                   0.022861              0.000000
0.000000

    Lymphocytes  Neutrophils.1  Eosinophils.1  Mast Cells  Dendritic
Cells  \
9      0.396934       0.000000       0.011272    0.094994
0.000000
13     0.726240       0.003574       0.000000    0.023075
0.000000
14     0.423194       0.012312       0.000000    0.033189
0.003214
15     0.325964       0.000000       0.000000    0.030454
0.033241
17     0.322847       0.079236       0.000000    0.082228
0.000142

    Macrophages
9      0.496801
13     0.247111
14     0.528092
15     0.610342
17     0.515548

[5 rows x 64 columns]
```

```
df_Sample_NonNa = df_Sample_NonNa.reset_index()
df_Sample_NonNa.set_index('TCGA Participant Barcode', inplace=True)
df_Sample_NonNa.drop(columns=['index', 'level_0'], inplace=True,
errors='ignore')
df_Sample_NonNa = df_Sample_NonNa.reset_index()

# Ensure row names are the index

# Extract the first three parts, replacing . with - for consistency
df_filtered_valid_rows['join_key'] =
df_filtered_valid_rows['index'].str.split('[.-]').str[:3].str.join('-'
)
df_Sample_NonNa['TCGA Participant Barcode'] = df_Sample_NonNa['TCGA
Participant Barcode'].str.split('[.-]').str[:3].str.join('-')

# Merge on the new join_key column
df_merged2 = df_Sample_NonNa.merge(df_filtered_valid_rows,
right_on='join_key', left_on='TCGA Participant Barcode',how='right')
```

```
# Drop the helper column if not needed
df_merged2.drop(columns=['join_key'], inplace=True)

df_merged2.head()
```

```
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/
format.py:1458: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in greater
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
58: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in greater
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
```

```
  TCGA Participant Barcode TCGA Study Immune Subtype           TCGA
Subtype  \
0             TCGA-OR-A5J1        ACC              C4
ACC.CIMP-high
1             TCGA-OR-A5J2        ACC              C3
ACC.CIMP-low
2             TCGA-OR-A5J3        ACC              C4  ACC.CIMP-
intermediate
3             TCGA-OR-A5J5        ACC              C4  ACC.CIMP-
intermediate
4             TCGA-OR-A5J6        ACC              C4
ACC.CIMP-low

   Leukocyte Fraction  Stromal Fraction  Intratumor Heterogeneity  \
0            0.046374              0.10                      0.02
1            0.057859              0.11                      0.16
2            0.048460              0.07                      0.11
3            0.016759              0.07                      0.15
4            0.189302              0.31                      0.06

   TIL Regional Fraction  Proliferation  Wound Healing  ...   VTA1_S5
```

```
\
0                    NaN        -1.2768         -0.0515  ...  1124.880

1                    NaN        -0.5351         -0.2730  ...  1008.190

2                    NaN        -1.0503         -0.1125  ...   537.994

3                    NaN         0.0606          0.0060  ...   455.461

4                    NaN        -2.2332         -0.4140  ...   989.176


   WDHD1_S5   WDR54_S5  WDR77_S5  WIPF1_S3  WNT2B_S2  WNT8B_S2
WSB2_S5  \
0   64.3612   411.6230   951.489   571.086   11.5274    0.0000
1731.51
1  162.6630    64.0182   805.663   584.217   10.8710    0.4026
5007.11
2  183.6770  1183.8200   865.649   158.791    0.5925    3.5550
2508.67
3  192.8740   509.6820   574.748   343.919   17.0411    0.7746
1889.23
4   21.2366   483.9700   582.888   393.435   11.1771    0.0000
3739.87

   ZWILCH_S5    ZYX_S2
0   183.9580  1841.020
1   264.1250  3059.990
2    90.6532  2655.610
3   260.2630  2367.930
4    64.2685   708.071

[5 rows x 505 columns]
```

```
# Set 'index_y' as the new index

# Drop unnecessary columns
df_merged2.drop(columns=['index', 'level_0'], inplace=True,
errors='ignore')

df_merged2
```

```
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/
format.py:1458: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in greater
```

```
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
58: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in greater
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
```

	TCGA Participant Barcode	TCGA Study	Immune Subtype	\
0	TCGA-OR-A5J1	ACC	C4	
1	TCGA-OR-A5J2	ACC	C3	
2	TCGA-OR-A5J3	ACC	C4	
3	TCGA-OR-A5J5	ACC	C4	
4	TCGA-OR-A5J6	ACC	C4	
...	...	...	...	
9121	TCGA-CG-4472	STAD	C3	
9122	TCGA-CG-4474	STAD	C2	
9123	TCGA-CG-4475	STAD	C4	
9124	TCGA-CG-4476	STAD	C2	
9125	TCGA-CG-4477	STAD	C2	

	TCGA Subtype	Leukocyte Fraction	Stromal Fraction	\
0	ACC.CIMP-high	0.046374	0.10	
1	ACC.CIMP-low	0.057859	0.11	
2	ACC.CIMP-intermediate	0.048460	0.07	
3	ACC.CIMP-intermediate	0.016759	0.07	
4	ACC.CIMP-low	0.189302	0.31	
...	...	...	...	
9121	NaN	0.208206	NaN	
9122	GI.GS	0.292010	0.60	
9123	GI.CIN	0.174603	0.57	
9124	GI.CIN	0.280260	0.72	
9125	GI.CIN	0.301511	0.51	

	Intratumor Heterogeneity	TIL Regional Fraction		Proliferation	\
0	0.02	NaN	-1.2768		
1	0.16	NaN	-0.5351		
2	0.11	NaN	-1.0503		
3	0.15	NaN	0.0606		

```
4                          0.06                    NaN        -2.2332

...                         ...                    ...           ...

9121                        NaN               9.149475        -0.2401

9122                       0.32              13.349560         0.5126

9123                       0.13              19.300401         0.5432

9124                       0.25               7.341158         0.2374

9125                       0.14              31.114986         0.8605


      Wound Healing  ...      VTA1_S5    WDHD1_S5     WDR54_S5
WDR77_S5  \
0           -0.0515  ...  1124.880000   64.361200   411.623000
951.489000
1           -0.2730  ...  1008.190000  162.663000    64.018200
805.663000
2           -0.1125  ...   537.994000  183.677000  1183.820000
865.649000
3            0.0060  ...   455.461000  192.874000   509.682000
574.748000
4           -0.4140  ...   989.176000   21.236600   483.970000
582.888000
...             ...  ...          ...         ...          ...
...
9121        -0.1060  ...  1002.318114  203.007108   156.681161
642.255901
9122         0.1460  ...  1596.401835  419.267633    71.266353
1004.123256
9123         0.1570  ...  1071.598647  437.232142   369.559696
421.023012
9124         0.0310  ...  1128.296079  346.089113   165.670587
1047.209793
9125         0.2880  ...  1828.291508  411.704146   329.701859
1035.886535

         WIPF1_S3    WNT2B_S2   WNT8B_S2      WSB2_S5   ZWILCH_S5
ZYX_S2
0      571.086000   11.527400   0.000000  1731.510000  183.958000
1841.020000
1      584.217000   10.871000   0.402600  5007.110000  264.125000
3059.990000
2      158.791000    0.592500   3.555000  2508.670000   90.653200
2655.610000
3      343.919000   17.041100   0.774600  1889.230000  260.263000
2367.930000
```

```
4      393.435000   11.177100   0.000000  3739.870000   64.268500
708.071000
...           ...         ...        ...          ...         ...
...
9121  1964.794656   63.843908   2.351098  1716.985841  292.541116
3359.241568
9122  1782.515239  319.359797   1.061085  2914.005907  745.532095
4264.469081
9123  1400.004751   21.267203   0.283374  1535.171082  556.284407
3103.609391
9124  1945.712810  325.416409   0.574527  1652.487434  516.997610
3302.569055
9125  2546.771364   77.610289  15.219672  1446.443138  923.544322
2497.814797

[9126 rows x 504 columns]
```

```
df_merged2.shape[0]
```

```
9126
```

```
column_names_Table_S1 = df_merged2.columns[:64].tolist()
print(column_names_Table_S1)
```

```
['TCGA Participant Barcode', 'TCGA Study', 'Immune Subtype', 'TCGA
Subtype', 'Leukocyte Fraction', 'Stromal Fraction', 'Intratumor
Heterogeneity', 'TIL Regional Fraction', 'Proliferation', 'Wound
Healing', 'Macrophage Regulation', 'Lymphocyte Infiltration Signature
Score', 'IFN-gamma Response', 'TGF-beta Response', 'SNV Neoantigens',
'Indel Neoantigens', 'Silent Mutation Rate', 'Nonsilent Mutation
Rate', 'Number of Segments', 'Fraction Altered', 'Aneuploidy Score',
'Homologous Recombination Defects', 'BCR Evenness', 'BCR Shannon',
'BCR Richness', 'TCR Shannon', 'TCR Richness', 'TCR Evenness', 'CTA
Score', 'Th1 Cells', 'Th2 Cells', 'Th17 Cells', 'OS', 'OS Time',
'PFI', 'PFI Time', 'B Cells Memory', 'B Cells Naive', 'Dendritic Cells
Activated', 'Dendritic Cells Resting', 'Eosinophils', 'Macrophages
M0', 'Macrophages M1', 'Macrophages M2', 'Mast Cells Activated', 'Mast
Cells Resting', 'Monocytes', 'Neutrophils', 'NK Cells Activated', 'NK
Cells Resting', 'Plasma Cells', 'T Cells CD4 Memory Activated', 'T
Cells CD4 Memory Resting', 'T Cells CD4 Naive', 'T Cells CD8', 'T
Cells Follicular Helper', 'T Cells gamma delta', 'T Cells Regulatory
Tregs', 'Lymphocytes', 'Neutrophils.1', 'Eosinophils.1', 'Mast Cells',
'Dendritic Cells', 'Macrophages']
```

```
column_names_PanCan = df_merged2.columns[64:].tolist()
print(column_names_PanCan)
```

```
['ACTL6A_S5', 'ADAM9_S2', 'ADAMTS1_S5', 'ADCY7_S3', 'AIMP2_S5',
'ALKBH7_S5', 'ALOX5AP_S3', 'AMPD3_S3', 'APITD1_S5', 'APOC1_S3',
'APOE_S3', 'APOO_S5', 'ARHGAP1_S2', 'ARHGAP15_S3', 'ARHGDIA_S2',
'ARRB2_S3', 'B2M_S3', 'BCCIP_S5', 'BRCA2_S5', 'BRIP1_S5', 'BSG_S2',
```

```
'BTK_S3', 'C11orf24_S5', 'C12orf24_S5', 'C13orf1_S5', 'C13orf18_S3',
'C13orf27_S5', 'C16orf61_S5', 'C19orf48_S5', 'C1orf54_S3', 'C3AR1_S3',
'C3orf26_S5', 'CASP1_S3', 'CBX1_S5', 'CCBL2_S5', 'CCL2_S3', 'CCL5_S3',
'CCRL2_S3', 'CCT5_S5', 'CD14_S4', 'CD163_S3', 'CD19_S4', 'CD247_S4',
'CD33_S3', 'CD37_S4', 'CD3D_S4', 'CD3E_S4', 'CD3G_S4', 'CD44_S2',
'CD48_S3', 'CD52_S3', 'CD53_S3', 'CD59_S2', 'CD79A_S4', 'CD79B_S4',
'CD84_S3', 'CD86_S3', 'CD8A_S3', 'CD8B_S4', 'CD97_S3', 'CDCA4_S5',
'CDH6_S2', 'CDK2_S5', 'CECR1_S3', 'CELF2_S3', 'CENPJ_S5', 'CENPN_S5',
'CENPO_S5', 'CENPW_S5', 'CEP78_S5', 'CHEK1_S5', 'CKLF_S5',
'COL16A1_S2', 'COL1A2_S2', 'COL3A1_S2', 'COL6A1_S2', 'COL6A3_S2',
'COL8A1_S2', 'COPS6_S5', 'COQ2_S5', 'CORO1A_S3', 'CORO1C_S5',
'COTL1_S5', 'COX17_S5', 'CPEB4_S5', 'CPVL_S3', 'CSF1R_S3',
'CTNNA1_S2', 'CTNNB1_S2', 'CTSC_S3', 'CTSL1_S3', 'CTSS_S3',
'CXCL10_S1', 'CXCR4_S3', 'CYBB_S3', 'CYTIP_S3', 'DAPK1_S3',
'DBNDD1_S5', 'DCBLD2_S5', 'DCK_S5', 'DCLRE1B_S5', 'DDX58_S1',
'DDX60_S1', 'DHFR_S5', 'DLEU1_S5', 'DLEU2_S5', 'DOCK2_S3', 'DSP_S2',
'DUT_S5', 'DVL1_S2', 'DVL3_S2', 'DYNLT1_S5', 'EBNA1BP2_S5',
'EFNA5_S2', 'EIF2AK1_S5', 'EIF4EBP1_S5', 'EIF4G1_S5', 'EMP2_S5',
'ENO1_S5', 'EPHA2_S2', 'EPHB2_S2', 'EPHB3_S2', 'EPHB4_S2',
'ERLIN1_S5', 'EVI2A_S3', 'EVI2B_S3', 'EXOSC8_S5', 'EZR_S5', 'F3_S5',
'FAM167A_S5', 'FAM89B_S5', 'FARSA_S5', 'FARSB_S5', 'FBXO41_S5',
'FCER1G_S3', 'FCGR1A_S3', 'FCGR2A_S3', 'FCGR2B_S3', 'FCGR3A_S3',
'FGD1_S2', 'FGL2_S3', 'FGR_S3', 'FLI1_S3', 'FLNC_S5', 'FMNL1_S3',
'FN1_S2', 'FNBP1_S3', 'FPR3_S3', 'FYB_S3', 'GGH_S5', 'GIMAP4_S3',
'GLRX3_S5', 'GNG11_S5', 'GNPTAB_S3', 'GPLD1_S5', 'GPNMB_S3',
'GSTCD_S5', 'H2AFZ_S5', 'HAS2_S5', 'HAUS1_S5', 'HCK_S3', 'HCLS1_S3',
'HDC_S4', 'HERC5_S1', 'HERC6_S1', 'HLA-DMA_S3', 'HLA-DRB1_S3',
'HMGN2_S5', 'HMHA1_S3', 'HN1L_S5', 'HNRNPA2B1_S5', 'HNRNPR_S5',
'HSPB11_S5', 'HSPG2_S2', 'HYLS1_S5', 'ICAM1_S2', 'ID2_S5', 'ID3_S5',
'IFI16_S3', 'IFI27_S1', 'IFI30_S3', 'IFI44_S1', 'IFI44L_S1',
'IFI6_S1', 'IFIH1_S1', 'IFIT1_S1', 'IFIT2_S1', 'IFIT3_S1', 'IFRD2_S5',
'IGF2R_S2', 'IGFBP2_S2', 'IGFBP3_S2', 'IGFBP4_S2', 'IGFBP5_S2',
'IGJ_S4', 'IKZF1_S3', 'IL10RA_S3', 'IL18_S3', 'IL7R_S3', 'IMP4_S5',
'IPO4_S5', 'ISG15_S1', 'ITGA3_S2', 'ITGA4_S2', 'ITGA5_S2', 'ITGA6_S5',
'ITGB2_S2', 'ITGB3_S2', 'ITGB5_S2', 'ITGB8_S2', 'JUNB_S5', 'JUP_S2',
'KIAA0090_S5', 'KRR1_S5', 'LAIR1_S3', 'LAMA4_S2', 'LAMB1_S2',
'LAPTM5_S3', 'LCK_S4', 'LCP1_S3', 'LCP2_S3', 'LHFPL2_S3', 'LILRB4_S3',
'LMNB2_S5', 'LOXL2_S5', 'LRMP_S3', 'LRP1_S2', 'LRRC17_S2',
'LRRC40_S5', 'LSM3_S5', 'LSM4_S5', 'LST1_S3', 'LTB_S4', 'LY86_S3',
'LYAR_S5', 'LYN_S3', 'LYZ_S3', 'MAGOHB_S5', 'MAP3K8_S5', 'MAPRE1_S5',
'MARCKSL1_S2', 'MARVELD2_S5', 'MCM3_S5', 'MCM7_S5', 'MCTS1_S5',
'MERTK_S3', 'MET_S5', 'MFSD11_S5', 'MKKS_S5', 'MLF1IP_S5', 'MMP11_S2',
'MMP1_S2', 'MMP14_S2', 'MMP17_S2', 'MMP19_S2', 'MMP2_S2', 'MMP3_S2',
'MNAT1_S5', 'MNDA_S3', 'MPP1_S3', 'MRPL12_S5', 'MRPL37_S5',
'MRPS16_S5', 'MRPS28_S5', 'MRTO4_S5', 'MS4A1_S4', 'MSN_S3', 'MT1F_S5',
'MT3_S5', 'MTA1_S2', 'MTHFD1_S5', 'MX1_S1', 'MYBL1_S5', 'MYBL2_S5',
'MYCBP_S5', 'MYL6_S5', 'MYO1F_S3', 'NCEH1_S5', 'NCF2_S3',
'NCKAP1L_S3', 'NCLN_S5', 'NEO1_S2', 'NID1_S2', 'NLN_S5', 'NME1_S5',
'NOP16_S5', 'NOTCH2_S2', 'NPC2_S3', 'NPL_S3', 'NRIP3_S5', 'NUDT1_S5',
```

```
'NUDT15_S5', 'NUP107_S5', 'NUP35_S5', 'NUP85_S5', 'NUP93_S5',
'NUPL1_S5', 'NUTF2_S5', 'OAS1_S1', 'OAS2_S1', 'OAS3_S1', 'OASL_S1',
'OSBPL3_S3', 'PA2G4_S5', 'PAICS_S5', 'PAK1_S2', 'PAK2_S2', 'PDAP1_S5',
'PDIA4_S5', 'PDLIM7_S5', 'PFKP_S5', 'PFN1_S5', 'PGM2_S5', 'PHF19_S5',
'PIK3CG_S3', 'PITPNC1_S5', 'PLAT_S2', 'PLAUR_S5', 'PLCG2_S3',
'PLEK_S3', 'PLG_S5', 'PLK4_S5', 'PLOD2_S5', 'PNN_S5', 'PNO1_S5',
'POLE2_S5', 'POLR3K_S5', 'PPIH_S5', 'PSMA7_S5', 'PSMC3_S5',
'PSMD12_S5', 'PSMD14_S5', 'PSMD2_S5', 'PTPLB_S5', 'PTPRC_S3',
'PXN_S2', 'RAB3B_S5', 'RAC1_S2', 'RASSF2_S3', 'RASSF4_S3', 'RBM14_S5',
'RFC3_S5', 'RGS8_S5', 'RHOA_S2', 'RHOB_S2', 'RHOC_S5', 'RHOG_S2',
'RHOQ_S2', 'RMND5B_S5', 'RNASE6_S3', 'RND3_S2', 'RNF138_S5',
'RNF41_S5', 'RPN1_S5', 'RPP40_S5', 'RSAD2_S1', 'RTP4_S1', 'RUNX3_S3',
'RUVBL1_S5', 'SAMD9_S1', 'SAMHD1_S3', 'SAMSN1_S3', 'SAR1A_S5',
'SAR1B_S5', 'SDC1_S5', 'SELL_S3', 'SELPLG_S3', 'SEMA3F_S2',
'SERPINE1_S2', 'SH2B3_S3', 'SH3BP5L_S5', 'SKA1_S5', 'SKA2_S5',
'SLC16A1_S5', 'SLC1A3_S3', 'SLC25A40_S5', 'SLC25A5_S5', 'SLC7A7_S3',
'SMC2_S5', 'SMO_S2', 'SMS_S5', 'SMURF2_S5', 'SNRPA1_S5', 'SNRPA_S5',
'SNRPB_S5', 'SNRPC_S5', 'SNRPD1_S5', 'SNRPE_S5', 'SNX17_S5',
'SP140_S3', 'SPAG17_S5', 'SPARC_S2', 'SRM_S5', 'STAT1_S1',
'STK17A_S5', 'STRA13_S5', 'SYK_S3', 'TAGLN_S5', 'TAP1_S1',
'TBXAS1_S3', 'TCEB1_S5', 'TCF7L2_S2', 'THBS1_S2', 'THBS2_S2',
'TIMP1_S2', 'TLR2_S3', 'TMEM130_S5', 'TNC_S2', 'TNFAIP3_S3',
'TNFRSF12A_S5', 'TNFRSF1A_S2', 'TNFRSF1B_S3', 'TNFSF13B_S3',
'TOMM40_S5', 'TPI1_S5', 'TPM1_S5', 'TPM2_S5', 'TPM3_S5', 'TPRKB_S5',
'TRA2B_S5', 'TUBA4A_S5', 'TUBG1_S5', 'UAP1_S5', 'UBE2J1_S5',
'UMPS_S5', 'UQCR10_S5', 'USPL1_S5', 'VCAN_S2', 'VDAC1_S5', 'VSIG4_S3',
'VTA1_S5', 'WDHD1_S5', 'WDR54_S5', 'WDR77_S5', 'WIPF1_S3', 'WNT2B_S2',
'WNT8B_S2', 'WSB2_S5', 'ZWILCH_S5', 'ZYX_S2']
```

More Preprocessing

```
# Remove display limits
pd.set_option('display.max_rows', None)  # Show all rows
pd.set_option('display.max_columns', None)  # Show all columns
pd.set_option('display.width', None)  # Adjust width to fit content
pd.set_option('display.max_colwidth', None)  # Show full column
content
#print(df_final.isna().sum())

df_merged2["Immune Subtype"].value_counts()
```

```
Immune Subtype
C2    2591
C1    2416
C3    2397
C4    1157
C5     385
```

```
C6     180
Name: count, dtype: int64
```

Balanced Representation of Immune Subtypes

I selected patient samples randomly while ensuring that each immune subtype was represented according to the required distribution: C1 (385 samples), C2 (414 samples), C3 (383 samples), C4 (462 samples), C5 (231 samples), and C6 (134 samples). This resulted in a total of 2009 samples in my dataset.

```
# Define the target sample sizes
target_samples = {
    "C1": 385,
    "C2": 414,
    "C3": 383,
    "C4": 462,
    "C5": 231,
    "C6": 134
}

# Perform stratified sampling
df_balanced = pd.concat([
    df_merged2[df_merged2["Immune Subtype"] ==
subtype].sample(n=target_samples[subtype], random_state=42)
    for subtype in target_samples
])

# Verify the final distribution
print(df_balanced["Immune Subtype"].value_counts())
```

```
Immune Subtype
C4    462
C2    414
C1    385
C3    383
C5    231
C6    134
Name: count, dtype: int64
```

```
df_balanced.shape[0]
```

```
2009
```

```
# This Python 3 environment comes with many helpful analytics
libraries installed
# It is defined by the kaggle/python Docker image:
https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load
```

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
from sklearn.decomposition import NMF
from sklearn.preprocessing import MinMaxScaler
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix
df_balanced.head()
```

```
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/
format.py:1458: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in greater
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
```

```
     TCGA Participant Barcode TCGA Study Immune Subtype  TCGA Subtype
\
1264              TCGA-WA-A7H4       HNSC              C1           NaN

548               TCGA-B5-A3F9       UCEC              C1  UCEC.CN_HIGH

6812              TCGA-49-6767       LUAD              C1        LUAD.3

7863              TCGA-AA-A03J       COAD              C1        GI.CIN

3228              TCGA-77-7337       LUSC              C1           NaN


      Leukocyte Fraction  Stromal Fraction  Intratumor
Heterogeneity  \
1264            0.379482              0.82                      0.40

548             0.219421              0.74                      0.08

6812            0.348569              0.57                      0.26

7863            0.241930              0.55                      0.05

3228            0.297990              0.48                      0.22
```

```
      TIL Regional Fraction  Proliferation  Wound Healing  \
1264                    NaN         0.5746         0.1880
548                8.503943         0.1406         0.1875
6812              14.644431         0.6171         0.2375
7863               8.225735         0.2587         0.1150
3228              15.458627         0.9131         0.3410

      Macrophage Regulation  Lymphocyte Infiltration Signature
Score  \
1264                -0.1361                                  0.3799

548                 -0.1861                                 -0.1434

6812                 0.3779                                  0.6282

7863                -0.3616                                  0.2081

3228                 0.2877                                  0.2484


      IFN-gamma Response  TGF-beta Response  SNV Neoantigens  \
1264              0.0053             0.7164             27.0
548              -0.1845            -0.1684             27.0
6812              0.2599             0.2542            166.0
7863              0.1503             0.1076             25.0
3228             -0.1595             0.8806            176.0

      Indel Neoantigens  Silent Mutation Rate  Nonsilent Mutation Rate
\
1264                0.0              0.511520                 1.591394

548                 3.0              0.494144                 1.095711

6812               60.0              4.161290                11.800230

7863                NaN              0.835700                 2.328021

3228              127.0              2.884412                 8.474817


      Number of Segments  Fraction Altered  Aneuploidy Score  \
1264                72.0          0.197422              10.0
548                123.0          0.271444              12.0
6812               172.0          0.581840              21.0
7863               119.0          0.507269              18.0
3228               169.0          0.578212              13.0

      Homologous Recombination Defects  BCR Evenness  BCR Shannon  \
1264                              34.0      0.922180     2.026238
548                                4.0           NaN          NaN
```

```
6812                              30.0           NaN          NaN
7863                              10.0           NaN          NaN
3228                              54.0      0.877917     3.549464

      BCR Richness  TCR Shannon  TCR Richness  TCR Evenness  CTA Score
\
1264           9.0     0.636514           2.0      0.918296   5.581650

548            NaN     1.549826           5.0      0.962961   3.171152

6812           NaN     2.679544          17.0      0.945761   2.304206

7863           NaN          NaN           NaN           NaN   0.542844

3228          57.0     2.871476          18.0      0.993463   2.134452


        Th1 Cells    Th2 Cells   Th17 Cells   OS  OS Time  PFI  PFI
Time  \
1264 -1107.504119   335.699054  -633.233703  0.0    443.0  0.0
443.0
548   -706.967641  -589.271926 -1741.589978  0.0   4155.0  0.0
4155.0
6812    68.317547   282.227079 -2484.773402  0.0    677.0  0.0
677.0
7863  -800.007476  -581.504381   395.084039  0.0   1246.0  0.0
1246.0
3228  -159.628570  1137.349162 -1602.002392  1.0   3253.0  0.0
3253.0

      B Cells Memory  B Cells Naive  Dendritic Cells Activated  \
1264        0.000000       0.075893                   0.000000
548         0.039246       0.000000                   0.002850
6812        0.010711       0.014774                   0.000000
7863        0.015177       0.008338                   0.000000
3228        0.000000       0.049060                   0.044275

      Dendritic Cells Resting  Eosinophils  Macrophages M0
Macrophages M1  \
1264                 0.000000     0.000000        0.285709
0.026984
548                  0.000000     0.003538        0.076611
0.007250
6812                 0.000000     0.000000        0.252027
0.067449
7863                 0.018624     0.000000        0.034319
0.078631
3228                 0.002638     0.000000        0.128922
0.111607
```

```
      Macrophages M2  Mast Cells Activated  Mast Cells Resting
Monocytes  \
1264        0.153365              0.000000            0.026059
0.023109
548         0.341072              0.000000            0.000000
0.000000
6812        0.164293              0.000000            0.033246
0.000000
7863        0.196650              0.042318            0.000000
0.033560
3228        0.218733              0.037657            0.000000
0.011673

      Neutrophils  NK Cells Activated  NK Cells Resting  Plasma Cells
\
1264     0.003019            0.087552          0.008897      0.111249

548      0.001251            0.064078          0.000000      0.035881

6812     0.000000            0.111910          0.000000      0.000000

7863     0.000000            0.093794          0.000000      0.026578

3228     0.015157            0.000000          0.059039      0.095555


      T Cells CD4 Memory Activated  T Cells CD4 Memory Resting  \
1264                      0.000000                    0.106067
548                       0.000000                    0.070163
6812                      0.027125                    0.043710
7863                      0.000000                    0.000000
3228                      0.014349                    0.107035

      T Cells CD4 Naive  T Cells CD8  T Cells Follicular Helper  \
1264           0.017419     0.027415                   0.036103
548            0.000000     0.169194                   0.133036
6812           0.000000     0.167443                   0.060282
7863           0.000000     0.277904                   0.089188
3228           0.000000     0.087035                   0.016079

      T Cells gamma delta  T Cells Regulatory Tregs  Lymphocytes  \
1264                  0.0                  0.011161     0.481756
548                   0.0                  0.055830     0.567428
6812                  0.0                  0.047029     0.482986
7863                  0.0                  0.084921     0.595899
3228                  0.0                  0.001184     0.429337

      Neutrophils.1  Eosinophils.1  Mast Cells  Dendritic Cells
Macrophages  \
1264       0.003019       0.000000    0.026059         0.000000
```

```
0.489166
548        0.001251       0.003538    0.000000          0.002850
0.424933
6812       0.000000       0.000000    0.033246          0.000000
0.483769
7863       0.000000       0.000000    0.042318          0.018624
0.343160
3228       0.015157       0.000000    0.037657          0.046913
0.470935

         ACTL6A_S5      ADAM9_S2   ADAMTS1_S5     ADCY7_S3     AIMP2_S5
\
1264   1414.260000   5574.520000   942.308000   946.314000   563.534000

548    1354.770000   1138.580000  1640.530000  1227.540000   482.911000

6812   1637.870000   5176.330000   161.538000  1004.730000   488.988000

7863    930.343078   1329.678264   216.256995   370.030352  1349.628189

3228   3289.310000  14204.300000  2384.660000   573.328000  1188.010000


         ALKBH7_S5  ALOX5AP_S3    AMPD3_S3   APITD1_S5     APOC1_S3  \
1264    521.635000    237.1790  587.340000  216.987000   138.622000
548     872.698000    185.8710  353.850000  340.800000  1306.070000
6812    299.408000    482.2480  276.331000  274.065000  1394.080000
7863   1956.948165    131.8582  149.806773  125.083134   400.202368
3228    317.341000    743.7580  308.820000  368.092000   126.118000

           APOE_S3     APOO_S5   ARHGAP1_S2  ARHGAP15_S3    ARHGDIA_S2
\
1264   1125.000000  110.577000  3838.940000    44.070500   8888.620000

548    4475.310000  620.729000  1648.980000    12.921500   6709.740000

6812   5431.360000  372.781000  1626.630000    71.597600   5318.340000

7863   2667.522991  252.673868  3276.934604    12.524268  15147.247057

3228    620.366000  315.296000  4495.270000    28.632300   8677.970000


          ARRB2_S3        B2M_S3     BCCIP_S5    BRCA2_S5
BRIP1_S5  \
1264    523.237000  48758.800000   951.234000  224.359000  169.071000

548     521.333000  66318.100000  1651.590000   70.571300   51.189000

6812    695.858000  53988.100000  1667.410000  109.468000   92.307700
```

```
7863  1306.517461  41306.987107   769.615704  146.312249   36.943463

3228   564.124000  33636.100000  1540.930000  147.593000  209.970000


           BSG_S2      BTK_S3  C11orf24_S5  C12orf24_S5  C13orf1_S5  \
1264  16931.90000   86.538500  1333.330000     66.50640  310.096000
548   21927.30000   60.631700   754.417000     93.92940  273.836000
6812   8614.20000  156.805000  1263.310000     68.04730  381.657000
7863  19473.23631   28.087146  1035.470316     53.88891  492.691892
3228  13157.20000   85.896900  1387.980000     84.19260  330.976000

      C13orf18_S3  C13orf27_S5  C16orf61_S5  C19orf48_S5
C1orf54_S3  \
1264   127.404000   193.910000   358.758000   976.763000  155.449000

548    199.786000   180.404000   548.716000  1952.140000  203.265000

6812    38.461500   166.864000   548.941000   863.314000  137.278000

7863  2546.069144   117.261586   479.162611  4435.195669   66.573924

3228    44.652700   250.533000   565.549000  1145.290000   81.806600


       C3AR1_S3   C3orf26_S5    CASP1_S3      CBX1_S5    CCBL2_S5  \
1264  133.814000   787.660000  790.865000  1612.180000  716.090000
548   154.561000   635.141000  120.766000  2532.120000  626.022000
6812  370.414000   468.047000  199.408000  1936.090000  795.284000
7863  130.132203   375.363702  463.236328   620.153334  424.498066
3228  369.493000  1113.930000  672.177000  2124.930000  621.464000

          CCL2_S3      CCL5_S3    CCRL2_S3       CCT5_S5
CD14_S4  \
1264   351.763000  1123.400000   36.057700   7229.970000  1347.76000

548    620.232000   500.957000   27.333900   6531.320000  1638.55000

6812   642.604000  2890.530000   71.005900  10427.200000  1537.87000

7863   122.172779   607.562311  370.657909   8111.215987  1392.85134

3228  1044.060000   555.944000   62.377500  12827.600000  1825.99000


         CD163_S3   CD19_S4    CD247_S4    CD33_S3   CD37_S4
CD3D_S4  \
1264   320.513000  9.615400  101.763000  31.250000  169.8720
92.948700
548    583.456000  5.466800   56.158800  67.092400  190.8410
93.432400
```

```
6812   834.319000  1.775100  133.728000  71.005900  450.8880
200.592000
7863   164.663887  8.810904   71.947677  30.009939  246.5297
109.859681
3228  1267.660000  3.408600   63.400100  46.357100  182.0200
79.420500

         CD3E_S4    CD3G_S4       CD44_S2     CD48_S3     CD52_S3
CD53_S3  \
1264  187.500000  14.423100  18930.300000   94.551300  186.699000
338.14100
548   194.320000   4.472800   2058.990000  157.543000  171.458000
465.17400
6812  602.367000  56.804700   9745.560000  253.846000  629.586000
863.90500
7863  189.354061  17.990105   3261.290314   96.453995  268.239955
315.95604
3228  256.327000  41.244100  23292.700000  132.595000  219.514000
878.05700

           CD59_S2    CD79A_S4   CD79B_S4     CD84_S3     CD86_S3
CD8A_S3  \
1264  14369.400000  133.814000  56.089700   20.833300  233.173000
33.65380
548    6228.660000   16.400400  18.388300   28.824900  104.863000
197.79800
6812   6684.020000   37.869800  22.485200   62.721900  304.734000
652.07100
7863   2817.520364   64.726949  33.807438   25.931826   77.030344
276.25988
3228  13698.500000  174.862000  31.359200  117.597000  397.103000
208.60700

         CD8B_S4      CD97_S3     CDCA4_S5     CDH6_S2      CDK2_S5  \
1264   24.038500   761.218000  1024.840000    84.13460   818.910000
548   229.605000  1288.170000   685.834000    69.57730   876.177000
6812   97.041400  1813.020000   786.982000    11.83430  1046.150000
7863   77.724818  6066.851717   480.374708    23.29196   787.869981
3228  102.940000  1295.950000  1040.310000  2406.48000   923.392000

         CECR1_S3    CELF2_S3    CENPJ_S5     CENPN_S5    CENPO_S5  \
1264   422.276000  397.436000  226.763000   702.941000  202.780000
548   1813.480000  328.007000  616.256000   295.157000   74.810500
6812   971.598000  270.414000   89.349100   584.787000  145.675000
7863   447.176864  700.887602  196.492764   630.422866  190.410746
3228   659.225000  306.775000  107.712000  1274.420000  436.735000

         CENPW_S5    CEP78_S5    CHEK1_S5    CKLF_S5   COL16A1_S2  \
1264   568.109000  243.590000  537.660000  213.14100  5395.030000
548    454.240000  284.273000  269.364000  227.61700   182.392000
```

```
6812   353.846000  223.077000  502.959000  897.04100   710.059000
7863   931.541342   89.391743  175.446418  465.70463   681.668235
3228  1029.400000  493.566000  662.633000  235.19400   862.378000

          COL1A2_S2      COL3A1_S2     COL6A1_S2     COL6A3_S2
COL8A1_S2  \
1264   93153.000000  103386.000000  20055.300000  26830.900000
208.33300
548    10381.900000   22155.900000   9287.580000   1272.270000
9.93960
6812   21293.500000   21016.000000  18462.100000   8066.860000
395.85800
7863   30497.210498   23821.008655   8179.105901   8199.892726
237.28691
3228  105622.000000  155202.000000   6743.590000  34333.200000
896.12300

         COPS6_S5     COQ2_S5    CORO1A_S3    CORO1C_S5
COTL1_S5  \
1264  1607.370000  274.038000   713.141000  3767.630000  2705.130000

548   2249.340000  250.478000   567.552000  1157.970000  3072.830000

6812  8586.390000  269.822000  1411.830000  5149.700000  9766.860000

7863  2323.238838  333.605287  3106.212752  3463.398929  2085.454571

3228  1875.760000  253.941000   495.271000  8162.930000  3322.710000


        COX17_S5    CPEB4_S5    CPVL_S3     CSF1R_S3  CTNNA1_S2
CTNNB1_S2  \
1264  538.462000  811.699000  203.52600   724.359000  8176.2800
8176.970000
548   778.769000  460.701000  230.10200   653.033000  7995.4300
6567.580000
6812  839.645000  245.562000  393.49100   988.757000  6928.4000
13272.800000
7863  591.715669  303.291589  222.88383   671.456392  8261.3316
9974.124921
3228  399.489000  894.759000  299.95700  1158.930000  7950.5800
12085.200000

          CTSC_S3     CTSL1_S3       CTSS_S3    CXCL10_S1
CXCR4_S3  \
1264   5598.560000  1512.820000    835.737000   187.500000
298.077000
548    3442.090000  2101.730000   1766.270000   308.625000
2493.850000
6812   4711.830000  1643.790000  10181.700000  2600.000000
```

```
601.775000
7863   2538.514223   658.990557   3005.643104   304.784386
270.766523
3228  11806.700000  2636.220000   2027.100000   568.215000
1160.630000

          CYBB_S3   CYTIP_S3    DAPK1_S3   DBNDD1_S5     DCBLD2_S5
DCK_S5  \
1264   471.154000  129.80800  905.449000  197.115000   3229.170000
548.87800
548    468.156000  163.50700  462.689000  824.988000   2031.160000
357.82600
6812  1659.760000  208.87600  169.822000  862.130000  24820.100000
872.18900
7863   350.122639   53.78859  304.985752  337.246512    503.510219
352.77352
3228  1642.610000  252.91900  320.409000  167.703000   2947.080000
997.01800

      DCLRE1B_S5    DDX58_S1    DDX60_S1     DHFR_S5    DLEU1_S5
DLEU2_S5  \
1264  323.718000  770.833000  713.141000  225.561000   80.769200
69.487200
548   238.551000  387.645000  578.983000  422.026000  201.471000
49.285600
6812  264.497000  452.663000  763.905000  227.876000   92.112400
107.444000
7863  142.552052  396.642255  705.860341  388.993479  134.812507
25.400413
3228  276.438000  222.923000  731.487000  175.202000  170.703000
102.940000

        DOCK2_S3        DSP_S2       DUT_S5      DVL1_S2      DVL3_S2
\
1264  203.526000  44610.600000  1073.000000  2220.350000  2595.350000

548   100.390000   4294.410000  1422.170000  1487.960000  2026.190000

6812  321.302000   8887.570000   624.645000  1872.780000  2149.110000

7863  146.765031   8165.998316  1715.641893  3209.336764  2974.540775

3228  315.296000  52847.400000   577.541000   804.431000  4245.420000


        DYNLT1_S5  EBNA1BP2_S5    EFNA5_S2   EIF2AK1_S5
EIF4EBP1_S5  \
1264  1262.020000   2073.72000  518.429000  3613.550000  1009.620000

548   2670.780000   1220.58000   64.607500  5030.100000  1500.880000
```

```
6812  1210.650000   1886.98000   18.343200  4418.110000  1133.140000

7863  1160.197487   1246.37362    9.236864  5112.901411  1457.794599

3228  1212.100000   1114.96000  297.912000  5569.550000  3609.710000


         EIF4G1_S5      EMP2_S5        ENO1_S5     EPHA2_S2
EPHB2_S2  \
1264  12786.100000  2140.220000   45858.200000  1617.790000
504.808000
548    6257.490000  1126.160000   28507.300000   631.663000
450.265000
6812  13284.600000  2677.510000   47065.700000  1069.230000
2076.330000
7863  10527.587171  1362.348884   23591.512322  3201.655891
4132.074258
3228  16915.900000  3410.310000  143228.000000  2470.900000
2208.440000

        EPHB3_S2      EPHB4_S2   ERLIN1_S5    EVI2A_S3    EVI2B_S3
EXOSC8_S5  \
1264  1241.99000   2685.100000   955.12800   88.141000  157.051000
427.72400
548    351.86200   1916.860000  1254.88000   81.504900  127.724000
481.60400
6812   550.29600  13614.800000  2472.19000  200.592000  373.964000
585.28400
7863  5667.72571   5248.529283  1297.53771   22.786344   81.030267
471.38021
3228  2232.64000   1957.900000  2386.71000  232.467000  320.068000
306.17100

            EZR_S5        F3_S5  FAM167A_S5    FAM89B_S5     FARSA_S5
\
1264  11897.400000  8418.270000   90.544900  1355.770000  1358.970000

548    5497.600000   474.120000   20.873200  1135.600000  1468.080000

6812  17770.400000  1340.240000   69.822500  2065.680000   919.527000

7863  10746.871452   510.121969   39.215757  1077.730029  3382.821044

3228   9246.530000  1477.970000   29.314000   884.874000  1493.310000


         FARSB_S5   FBXO41_S5  FCER1G_S3  FCGR1A_S3    FCGR2A_S3
FCGR2B_S3  \
1264  1342.150000  324.519000  226.76300   27.24360   520.833000
```

```
120.994000
548   1338.370000   51.189000  541.21200  121.24300   434.361000
99.396200
6812  1501.780000  334.911000  937.27800  286.95300  1111.240000
200.592000
7863  1201.770979  734.838076  215.44325   30.38253   477.218772
35.273426
3228  1174.270000   85.215200  552.87600   85.53220  2778.360000
283.255000

        FCGR3A_S3     FGD1_S2   FGL2_S3      FGR_S3     FLI1_S3
FLNC_S5  \
1264   465.545000  204.327000  303.6860  165.064000  245.192000
5137.820000
548    828.964000  830.952000  142.6330   87.965600  137.664000
91.444500
6812  2390.530000  876.923000  492.3080  343.195000  191.124000
11415.400000
7863   420.841219  253.368838  250.4135  157.269486  106.790429
994.628822
3228  1908.480000  415.850000  512.3140  203.494000  254.282000
157.478000

         FMNL1_S3        FN1_S2     FNBP1_S3     FPR3_S3       FYB_S3
\
1264   574.103000  52455.900000   741.987000  255.609000   406.250000

548    138.658000   2513.230000  1015.830000   91.444500   315.086000

6812  1997.630000  54245.000000  1497.040000  469.822000   505.917000

7863   425.568343   7291.669478   639.341658  182.866036    84.920137

3228   345.292000  98427.600000   537.537000  563.443000  2971.280000


           GGH_S5   GIMAP4_S3     GLRX3_S5    GNG11_S5    GNPTAB_S3
GPLD1_S5  \
1264  1058.490000  187.500000  1020.030000  451.122000  1433.490000
4.807700
548    801.630000  278.806000   952.215000  611.783000  1293.140000
8.945700
6812   420.710000  485.207000  2571.010000  155.030000   735.503000
8.875700
7863   725.162161  182.790949  1104.235578  118.798676   664.235627
5.973108
3228  1462.970000  341.542000  2100.380000  263.826000  2109.590000
9.203200

          GPNMB_S3    GSTCD_S5     H2AFZ_S5     HAS2_S5    HAUS1_S5  \
1264   9549.680000  254.808000  5116.190000  198.638000  183.494000
```

```
548    2307.480000  318.565000  6503.990000   77.032000  531.273000
6812   6445.560000  238.462000  4285.210000   25.390500  150.296000
7863    656.968787  174.494972  4937.989857   53.483964  190.910932
3228  38693.500000  366.425000  5925.860000  753.104000  280.869000

          HCK_S3    HCLS1_S3     HDC_S4    HERC5_S1    HERC6_S1   HLA-
DMA_S3  \
1264  326.122000   532.85300  66.506400  175.481000  601.763000
532.853000
548   266.879000   777.77500   1.490900   97.905200  428.397000
2299.030000
6812  675.740000   770.41400   1.775100   63.905300  179.882000
1369.230000
7863  343.856397   342.87048  51.860779   36.200517  333.053243
1201.901095
3228  450.959000  1998.13000  20.110800  247.806000  414.146000
1186.200000

       HLA-DRB1_S3     HMGN2_S5     HMHA1_S3      HN1L_S5
HNRNPA2B1_S5  \
1264   2939.590000  5672.280000   722.756000  4440.640000
20217.900000
548    5338.540000  9221.480000   369.754000  4585.140000
22632.000000
6812  12539.900000  5878.700000   881.065000  2697.630000
12819.500000
7863   4294.439114  3542.639361  1596.510033  2639.516134
9956.139522
3228   4501.430000  6724.500000   230.081000  5817.460000
10226.500000

        HNRNPR_S5   HSPB11_S5      HSPG2_S2    HYLS1_S5
ICAM1_S2  \
1264  3596.150000  517.628000  12338.200000  217.949000  1789.260000

548   3508.680000  528.788000   4327.120000  124.742000   685.834000

6812  2786.390000  257.988000   5197.800000  156.805000  7046.750000

7863  1741.845332  139.147844  11231.842641   46.389359  1215.882523

3228  3903.880000  208.607000  13568.800000  298.935000  3448.830000


           ID2_S5       ID3_S5     IFI16_S3     IFI27_S1     IFI30_S3
\
1264   681.090000   822.917000  6826.120000  1596.150000  2146.550000

548   3584.230000  4047.910000  1736.450000  4601.050000  2636.480000
```

```
6812   600.592000   244.379000  2541.420000  5004.730000  5024.040000

7863  1846.198195  1449.654233   455.437316  8435.411981  4218.586672

3228   686.493000   932.595000  7451.210000  3560.630000  4289.150000


        IFI44_S1   IFI44L_S1      IFI6_S1    IFIH1_S1    IFIT1_S1
IFIT2_S1  \
1264  956.731000   293.26900  2016.830000  541.667000  358.173000
173.077000
548   437.840000   476.60500  5028.450000  457.719000  296.698000
255.945000
6812  846.746000  2188.17000   748.521000  908.876000  218.935000
714.793000
7863  126.583741   358.27975  3904.530045  365.052522  280.111308
218.129019
3228  233.830000    89.30550   989.859000  632.297000  137.367000
158.500000

         IFIT3_S1     IFRD2_S5     IGF2R_S2    IGFBP2_S2     IGFBP3_S2
\
1264   522.436000  1132.650000  7806.090000  1262.820000   2625.000000

548    367.269000   967.299000  1889.020000  7163.480000   3840.670000

6812  1108.280000  1590.020000  5543.200000   101.183000   2001.180000

7863   746.590341  2946.979229  5759.380637   769.358703   2308.976639

3228   466.979000   821.133000  5435.710000  4615.590000  25963.400000


         IGFBP4_S2     IGFBP5_S2       IGJ_S4    IKZF1_S3   IL10RA_S3
\
1264   7622.600000   4780.450000  6902.240000  229.167000  295.673000

548   25464.800000   2210.570000   294.213000  124.245000  228.611000

6812   5388.760000   5168.640000   202.367000  296.450000  513.610000

7863   6273.287026   3090.207707   879.747499  108.191389  310.704089

3228   9842.690000  20129.900000  3414.400000  206.562000  290.754000


          IL18_S3     IL7R_S3      IMP4_S5      IPO4_S5
ISG15_S1  \
1264   874.199000  229.968000   741.186000  1429.780000   820.513000

548     60.134700   31.806800   695.276000  1465.530000  1994.880000
```

```
6812  1228.400000   37.869800   848.521000  1385.810000  1202.960000

7863   264.227307   30.011095  1463.603138  2068.202687  3030.829389

3228   539.923000  277.801000   795.228000  1700.700000   467.320000


          ITGA3_S2    ITGA4_S2     ITGA5_S2      ITGA6_S5     ITGB2_S2
\
1264  15915.100000  282.340000  3802.080000  25782.900000   801.939000

548     918.421000   97.895300   583.456000   2266.230000  1002.910000

6812  15562.700000  183.462000  2215.380000   2964.500000  3699.830000

7863   3940.566087  123.141682  1132.763751   4288.080636  1357.033063

3228   6089.480000  358.493000  6613.380000  13324.200000  1438.310000


        ITGB3_S2     ITGB5_S2     ITGB8_S2      JUNB_S5         JUP_S2
\
1264  217.949000  5173.880000  1781.250000  3051.280000  20550.500000

548    13.915500  2538.580000   906.493000  3195.090000   5095.050000

6812   91.716000  2510.060000  1984.020000  1322.490000  12150.900000

7863   50.598276  2787.829954   214.333668  5502.742214  19274.491393

3228  191.905000  2923.220000   975.884000  2970.600000  33214.500000


      KIAA0090_S5      KRR1_S5   LAIR1_S3     LAMA4_S2     LAMB1_S2  \
1264  2780.450000  1008.450000  183.49400  2374.200000  5812.500000
548   1177.350000   633.680000  313.09800   421.440000  7204.730000
6812  1236.090000   564.284000  499.40800   483.432000  4458.580000
7863   720.656034   352.075264  208.89802   783.651454  5772.950729
3228  4090.670000  1403.150000  293.82200  2424.540000  5426.160000

        LAPTM5_S3      LCK_S4      LCP1_S3     LCP2_S3    LHFPL2_S3  \
1264  1458.330000  125.801000  1877.400000  274.038000  1033.650000
548   2701.590000  130.706000  2575.350000  169.967000   733.544000
6812  6992.900000  328.994000  3215.380000  467.450000   971.598000
7863  1804.768877  147.246087   817.064039  158.550938   716.245256
3228  3877.970000  159.523000  3375.880000  522.539000  1385.600000

       LILRB4_S3     LMNB2_S5     LOXL2_S5     LRMP_S3
LRP1_S2  \
1264   97.756400  3442.310000  1612.100000   52.083300   5866.990000
```

```
548   244.018000  2038.620000   444.301000   29.818900   2684.190000

6812  468.639000  2414.790000  1230.070000   21.301800   3021.890000

7863  202.107483  4983.516023  1275.832329   23.441881  11460.799714

3228  326.204000  3216.020000  2383.960000  326.885000  13292.900000


      LRRC17_S2   LRRC40_S5     LSM3_S5      LSM4_S5    LST1_S3
LTB_S4  \
1264  420.67300  415.064000  673.878000  1681.090000  126.60300
374.199000
548   177.42200  448.774000  945.258000  3186.140000   89.45660
162.016000
6812   79.88170  540.237000  910.651000  1820.710000  170.41400
1493.490000
7863   27.08007   48.738581  469.580308  2784.870978  104.39482
218.043389
3228  117.25600  318.705000  682.403000  1516.150000   75.67110
22.837700

         LY86_S3      LYAR_S5       LYN_S3        LYZ_S3
MAGOHB_S5  \
1264   25.641000   749.199000   883.814000  13019.200000  275.641000

548    79.020000   386.651000   773.799000    834.431000  372.239000

6812  103.550000  1045.560000  2642.600000   6668.050000  364.497000

7863   38.308787   678.810049   562.685169   3068.651264  192.543608

3228   51.129100   860.332000  1834.510000   6712.910000  458.798000


       MAP3K8_S5    MAPRE1_S5   MARCKSL1_S2  MARVELD2_S5      MCM3_S5
\
1264  154.647000  3822.120000   1948.720000   252.404000  4629.810000

548   268.370000  2547.030000  17631.900000   424.919000  3315.860000

6812  117.160000  3569.820000   7843.200000   400.592000  2574.560000

7863   80.427806  4168.603566   5833.127538   507.566001  3776.425366

3228  164.977000  7763.100000   1351.850000   402.897000  1764.290000


           MCM7_S5     MCTS1_S5    MERTK_S3        MET_S5   MFSD11_S5
\
```

```
1264   2780.830000  1647.440000  106.570000   3673.080000  856.763000

548    2835.220000  1015.330000  442.313000   1659.920000  312.591000

6812  12199.600000  1617.160000  321.894000  40805.900000  667.325000

7863   7193.775818   891.683567   71.470187   2531.035123  445.909178

3228   3067.410000   774.435000  337.452000   4123.390000  753.769000


          MKKS_S5   MLF1IP_S5      MMP11_S2       MMP1_S2
MMP14_S2  \
1264  1009.620000  614.728000   8549.680000  33935.100000
19246.000000
548   1106.780000  681.395000  15438.200000     13.915500
7597.350000
6812   592.308000  610.467000   7279.290000   3834.910000
29959.800000
7863   433.255505  269.988979   6572.061363   1598.728465
7607.790173
3228   662.292000  379.163000   1861.780000  23431.800000
12179.000000

       MMP17_S2    MMP19_S2       MMP2_S2      MMP3_S2    MNAT1_S5  \
1264  199.51900  592.949000  11664.300000  5686.700000  314.103000
548    36.27960  793.181000  10317.300000   102.378000  483.562000
6812   90.53250  338.462000   4640.240000    13.017800  378.106000
7863   26.74886  308.075381   6768.362318   896.832598   90.714612
3228   66.46780  371.879000  14209.500000  1543.080000  366.425000

         MNDA_S3     MPP1_S3    MRPL12_S5    MRPL37_S5    MRPS16_S5  \
1264   84.134600  214.744000  1384.620000  2020.830000  1345.720000
548    57.152800  176.925000  1809.510000  1491.940000  2112.120000
6812  324.852000  820.710000  1376.920000  1696.450000  1979.640000
7863   36.692463  862.539391  3222.565362  3021.732273  2019.185734
3228  133.617000  245.079000  1209.710000  1497.060000  2782.770000

       MRPS28_S5     MRTO4_S5   MS4A1_S4        MSN_S3     MT1F_S5  \
1264  588.878000  1703.530000  43.269200  11398.200000  167.468000
548   418.667000  1512.310000   1.987900   5326.140000  697.264000
6812  261.840000  1215.380000  11.834300  25971.600000  153.254000
7863  158.477752   976.047526   4.576277   2585.419706  459.608169
3228  461.283000  1683.510000  11.248400  17127.900000  172.135000

          MT3_S5      MTA1_S2    MTHFD1_S5       MX1_S1    MYBL1_S5  \
1264    3.205100  1064.900000  1481.530000  1167.470000  238.782000
548   147.106000  1970.030000  1389.490000  1137.590000   64.110500
6812    0.000000  1739.050000  2720.590000  1135.500000  431.361000
7863   91.901208  1555.890562  1733.931097  2361.615361   34.276176
```

```
3228    0.340900   998.722000  2440.870000   916.915000   78.738800

         MYBL2_S5     MYCBP_S5       MYL6_S5   MYO1F_S3
NCEH1_S5  \
1264  1282.850000   720.353000  15859.000000  259.61500   411.859000

548   1157.970000  1166.910000  18998.100000  301.66700   694.282000

6812  1129.590000   715.414000  18542.000000  450.29600  1969.230000

7863  4624.135012   617.681299  14625.742112  438.45008   451.314543

3228  3666.640000   489.817000  13088.400000  201.44900  1193.350000


          NCF2_S3  NCKAP1L_S3      NCLN_S5      NEO1_S2
NID1_S2  \
1264   375.801000  237.179000  2168.270000  1062.500000  1839.500000

548     82.001800  190.344000  2321.400000  2178.760000  1702.100000

6812   771.006000  445.562000  1911.240000   804.142000  1350.880000

7863   152.308393  256.994692  6476.962655  3104.024952  1136.524032

3228  1061.100000  351.087000  1691.690000   782.957000  5154.120000


          NLN_S5      NME1_S5    NOP16_S5    NOTCH2_S2      NPC2_S3  \
1264  435.096000  1711.670000  379.006000  2720.350000  2388.620000
548   422.434000  3232.230000  589.916000  2957.530000  6377.260000
6812  600.592000  2551.910000  322.485000  3546.150000  7479.290000
7863  374.185195  2190.579905  734.285336   712.373458  1552.987229
3228  511.973000  1949.150000  258.372000  3836.730000  6966.510000

          NPL_S3    NRIP3_S5    NUDT1_S5   NUDT15_S5    NUP107_S5
NUP35_S5  \
1264  213.141000   96.955100  445.513000  854.199000  1161.860000
250.801000
548   130.706000   16.400400  327.510000  688.771000  1000.420000
360.808000
6812  120.118000  134.320000  288.757000  879.846000  1000.590000
251.479000
7863   88.620145   14.386538  402.260555  600.915288   711.490968
148.455339
3228  691.947000  247.124000  491.521000  854.865000  1263.570000
408.010000

        NUP85_S5     NUP93_S5     NUPL1_S5     NUTF2_S5
OAS1_S1  \
1264  905.449000  1128.210000  1248.400000  1809.290000   560.096000
```

```
548   726.089000  1247.420000   756.405000  3353.130000   193.326000

6812  890.533000  1489.940000  1099.410000  3146.750000  1168.050000

7863  965.871197  1438.451723  1158.237576  3557.105678  1869.724499

3228  727.397000  1467.410000  1026.670000  2287.860000   477.887000


          OAS2_S1      OAS3_S1     OASL_S1    OSBPL3_S3
PA2G4_S5  \
1264  1220.350000  1335.740000  160.256000  1112.980000  2838.140000

548    385.657000   634.148000  114.306000   428.894000  4439.530000

6812  1243.200000  1981.070000  281.657000   815.385000  5613.610000

7863   935.073373  2756.552264  533.469981   596.266612  5580.006673

3228  1499.450000  1513.420000   82.147400   468.002000  4908.050000


         PAICS_S5      PAK1_S2      PAK2_S2      PDAP1_S5
PDIA4_S5  \
1264  3400.640000  1366.190000  3371.790000   2505.610000
5277.240000
548   4425.120000  2163.850000  2059.490000   3060.410000
12363.900000
6812  2313.610000  2145.560000  4628.990000  10816.600000
16620.700000
7863  3331.453327  1544.604524  1565.465184   3126.018611
6665.578363
3228  5149.380000  2339.330000  8009.890000   1954.840000
6801.870000

        PDLIM7_S5      PFKP_S5       PFN1_S5      PGM2_S5    PHF19_S5
\
1264  3774.840000  1290.870000  18421.500000  1930.290000   897.43600

548   1167.900000   834.431000  14217.600000   537.733000   265.88500

6812  1638.460000  3407.100000  12803.000000  1115.980000  1609.47000

7863  1790.815205  3396.808005  21598.814199   543.057721  1775.78342

3228   838.858000  4072.260000   8874.990000  2447.720000   550.49000


      PIK3CG_S3  PITPNC1_S5      PLAT_S2     PLAUR_S5    PLCG2_S3
PLEK_S3  \
```

```
1264   64.90380  311.699000   434.295000   758.013000  356.570000
229.167000
548     7.95170  519.345000   550.158000   603.832000  254.951000
299.679000
6812   47.33730  175.148000  2979.880000  2352.660000  207.101000
688.757000
7863   23.68202  214.514288   659.910176  2324.934366  114.980755
161.426919
3228   99.19050  411.760000  3454.960000  1480.700000  536.515000
561.057000

         PLG_S5     PLK4_S5     PLOD2_S5       PNN_S5      PNO1_S5
POLE2_S5  \
1264   0.000000  506.410000  1907.850000  2253.210000   594.551000
126.60300
548   23.855100  258.927000  2288.100000  2935.170000   393.112000
89.45660
6812   0.000000  369.231000  1333.730000  1604.730000   665.681000
140.82800
7863   3.938344  153.655922   439.280843  1478.458546   599.694356
67.87253
3228   0.000000  387.218000  7612.440000  1761.910000  1266.980000
168.72600

       POLR3K_S5    PPIH_S5      PSMA7_S5     PSMC3_S5    PSMD12_S5  \
1264  274.038000  492.78800   4498.400000  2745.990000  1370.990000
548   389.633000  743.98000   2975.920000  2598.220000   723.107000
6812  243.787000  580.47300   4809.470000  1921.890000  1149.110000
7863  574.739875  652.58254  12850.164704  1866.885196   501.157671
3228  508.223000  320.75000   6267.750000  2897.320000  2220.710000

        PSMD14_S5      PSMD2_S5    PTPLB_S5     PTPRC_S3
PXN_S2  \
1264  1602.560000   6587.340000  308.494000   516.026000  6759.01000

548   1436.770000   3072.830000   43.237300   339.438000  1946.43000

6812  1317.160000   5385.800000  243.195000   875.148000  7575.14000

7863   978.831589   4231.503959   75.527356   208.733265  7988.21454

3228  3146.480000  14749.000000  605.709000  1036.560000  4754.74000


       RAB3B_S5       RAC1_S2   RASSF2_S3    RASSF4_S3     RBM14_S5  \
1264   5.609000   6165.060000  365.385000   736.114000  1018.350000
548    1.987900   6536.290000  648.063000   825.982000  1355.860000
6812  26.627200   6388.170000  122.485000  1518.930000  1006.050000
7863   7.277026   7220.930029  122.210979   772.989068  2152.712306
3228  16.020500  11040.800000  252.578000   215.083000  1341.370000
```

```
         RFC3_S5  RGS8_S5       RHOA_S2      RHOB_S2     RHOC_S5
RHOG_S2  \
1264  365.385000   0.0000   9258.810000   2931.09000  3935.10000
1266.830000
548   442.810000   0.0000   9805.930000  11439.50000  3642.37000
898.541000
6812  404.734000   0.0000  13426.000000   1077.51000  6773.37000
1168.050000
7863  546.375419      NaN   9807.319352   3607.80426  5682.28049
1714.439988
3228  546.740000   0.3409  10884.400000   2874.14000  3976.82000
885.215000

          RHOQ_S2   RMND5B_S5   RNASE6_S3      RND3_S2   RNF138_S5  \
1264  1629.020000  351.763000  143.429000  1867.790000  368.590000
548   1649.260000  622.717000  124.742000  2087.320000  763.363000
6812  1291.900000  531.953000  175.148000  1360.360000  431.361000
7863   360.002602  937.189697  142.656267   438.776252  284.101016
3228  1348.960000  273.711000  116.574000  5216.870000  601.278000

         RNF41_S5       RPN1_S5    RPP40_S5    RSAD2_S1     RTP4_S1  \
1264   684.295000   6653.040000  213.141000  165.865000   92.147400
548    879.159000  11263.100000  134.682000   90.947500  193.326000
6812   828.402000   7985.800000  184.615000  308.876000  133.136000
7863  1102.745864   7524.043806  165.900124  630.334114  123.607153
3228  1178.700000  15401.800000  252.578000  294.844000  259.395000

         RUNX3_S3   RUVBL1_S5     SAMD9_S1    SAMHD1_S3   SAMSN1_S3  \
1264   657.853000  1214.70000  2354.970000   495.633000  111.378000
548    193.822000  1955.01000   236.066000  1195.850000  123.748000
6812   204.142000  1519.73000   305.325000   756.562000  193.491000
7863  1165.381881  1381.89423   380.608313   628.068511   74.518924
3228  1033.150000  2640.05000  1423.430000  1060.230000  184.406000

         SAR1A_S5   SAR1B_S5       SDC1_S5     SELL_S3   SELPLG_S3  \
1264  2307.690000  526.44200  16976.800000  138.622000  278.045000
548   2155.900000  437.84000   2425.760000  120.269000  216.187000
6812  1915.380000  364.49700  10459.200000  159.172000  585.799000
7863  1033.663388  291.35525   7002.780348   51.673906  422.963295
3228  3508.140000  580.82700  22700.000000  144.184000  383.468000

        SEMA3F_S2  SERPINE1_S2     SH2B3_S3   SH3BP5L_S5
SKA1_S5  \
1264  1383.810000   5875.00000   687.500000  1310.900000   93.750000

548    972.095000    117.28700   254.951000   594.389000  108.839000

6812   575.148000   1481.66000  1301.780000   553.254000  128.994000
```

```
7863   536.348425    237.63618   504.611923   983.091103  157.042894

3228  1699.530000   6642.01000   707.627000   593.779000  261.099000


         SKA2_S5   SLC16A1_S5    SLC1A3_S3  SLC25A40_S5    SLC25A5_S5
\
1264  786.859000  3054.490000   477.564000   132.212000  11109.000000

548   842.880000  1665.380000   179.410000    59.140700   4494.690000

6812  793.491000  2778.700000   168.047000   346.154000   8710.650000

7863  216.737482  1117.373004    26.435782   120.723276  18811.796336

3228  831.018000  6389.430000  1772.130000   227.695000  10477.400000


      SLC7A7_S3     SMC2_S5       SMO_S2       SMS_S5    SMURF2_S5  \
1264  176.28200  1104.17000   475.160000   888.622000   483.974000
548   183.88300   605.82000  1728.000000  2884.480000   338.444000
6812  632.54400  1209.47000   931.953000  3064.500000  1871.010000
7863  363.96797   315.73208   326.766597  1884.414469   243.065824
3228  353.81300  2964.47000   509.587000  2040.730000   571.283000

       SNRPA1_S5     SNRPA_S5     SNRPB_S5     SNRPC_S5
SNRPD1_S5  \
1264  424.680000  1115.380000  3641.030000  1903.040000   712.340000

548   595.880000  2843.230000  4689.010000  1951.640000  1430.310000

6812  614.201000  1108.280000  4876.920000  2188.760000   981.065000

7863  537.599672  2810.940332  4348.216621  2118.107643   686.240596

3228  855.901000  1157.900000  2653.260000  2204.690000  1225.050000


         SNRPE_S5     SNX17_S5   SP140_S3   SPAG17_S5      SPARC_S2  \
1264   600.962000  2597.760000  53.685900   41.666700  43480.000000
548   1544.620000  2713.520000  26.340000  289.740000  10374.000000
6812  1222.490000  2168.640000  97.041400    1.775100  12467.500000
7863   992.908169  2212.668745  21.068411    1.794282  12455.566615
3228   971.794000  2252.070000  33.063500   68.172100  53367.500000

           SRM_S5     STAT1_S1    STK17A_S5   STRA13_S5
SYK_S3  \
1264  1790.060000  3429.490000  1818.110000  697.917000   607.372000

548   1435.280000  4222.850000   463.186000  896.553000  1712.600000
```

```
6812  2275.150000  8585.800000  2156.800000  507.692000   547.337000

7863  2229.710794  3373.295658   204.670983  944.119588  1408.615337

3228  2054.370000  7263.060000  1204.600000  260.418000  1351.850000


        TAGLN_S5      TAP1_S1    TBXAS1_S3     TCEB1_S5   TCF7L2_S2  \
1264  7978.37000   3900.64000   156.250000  1051.280000  436.699000
548    899.53500   4301.37000   378.202000   883.632000  704.222000
6812  1727.81000   7538.46000   170.414000  1136.090000  855.621000
7863  5412.55443  11272.08586  1303.232132   570.101003  977.233934
3228  5734.30000   3268.85000   170.089000  1231.530000  709.331000

          THBS1_S2      THBS2_S2     TIMP1_S2      TLR2_S3  TMEM130_S5
\
1264  16249.200000   6189.100000  4869.390000   669.071000     9.61540

548    1140.570000    136.670000  4652.730000   343.911000     5.96380

6812   1164.500000   2704.730000  6509.470000   218.935000    27.81070

7863   2851.471087   1790.479229  8027.404873   107.980266    14.09625

3228   8540.260000  14896.000000  7015.590000  1279.930000    54.53770


             TNC_S2   TNFAIP3_S3  TNFRSF12A_S5  TNFRSF1A_S2
TNFRSF1B_S3  \
1264  51591.300000  1375.800000   3986.380000  3036.860000
710.737000
548    1885.550000  1716.570000   1070.000000  2660.840000
1382.600000
6812  18590.500000   947.929000   2894.080000  4164.500000
1972.780000
7863   2139.767469  1004.281681   2584.787646  3277.586534
2469.421757
3228  36879.800000  1954.840000   1190.290000  5124.500000
652.067000

      TNFSF13B_S3    TOMM40_S5       TPI1_S5      TPM1_S5      TPM2_S5
\
1264    26.442300  1080.930000   6519.230000  7943.420000  10874.20000

548     18.885300  1375.150000  11693.000000  1983.870000   1485.48000

6812    99.408300  1551.480000  15510.100000  3089.760000   4177.51000

7863    28.323496  2911.791666  13301.708123  4638.258269   1984.92103

3228    72.262500  1441.840000  21580.200000  2700.970000   1286.41000
```

```
          TPM3_S5    TPRKB_S5     TRA2B_S5    TUBA4A_S5     TUBG1_S5
\
1264   9447.920000  412.660000  3585.740000  7373.400000  1071.310000

548    7467.630000  722.113000  3100.170000   119.275000   984.022000

6812   8412.430000  335.503000  1897.630000  1947.930000  1028.400000

7863   8429.630182  211.853076  2050.314713  5324.263102  1245.895273

3228  10409.500000  435.620000  3686.750000  3058.880000  1813.040000


         UAP1_S5   UBE2J1_S5      UMPS_S5    UQCR10_S5    USPL1_S5  \
1264  1302.880000  1580.93000   970.353000  1076.920000  534.455000
548   1015.830000  2069.93000   620.729000  1863.180000  506.921000
6812   999.408000  2207.10000   849.112000  1656.800000  478.106000
7863   576.843949  1253.38716   781.709909  1179.679495  286.464799
3228  1534.210000  2274.56000  1168.130000   743.076000  164.295000

           VCAN_S2     VDAC1_S5   VSIG4_S3     VTA1_S5    WDHD1_S5
WDR54_S5  \
1264   4942.310000  4340.540000   81.73080  1551.28000  425.481000
296.47400
548    7417.440000  4943.470000  348.88100  1079.44000  187.362000
766.84100
6812   2764.500000  4909.470000  621.89400  1128.40000  508.284000
151.47900
7863   1249.379279  7151.557949  160.41858   716.27527  101.055152
140.38918
3228  14117.100000  5529.780000  479.93200  1951.09000  611.163000
215.08300

         WDR77_S5     WIPF1_S3    WNT2B_S2  WNT8B_S2      WSB2_S5
ZWILCH_S5  \
1264   906.250000   989.583000   28.044900  6.410300  1991.190000
483.061000
548   4555.820000   750.441000   16.897300  1.490900  1495.420000
452.988000
6812  1057.400000  1007.100000    5.325400  1.775100  2242.600000
325.953000
7863  1725.383932   327.268858   31.568168  2.028401  2127.285705
266.422502
3228   842.948000  1371.620000  249.169000  0.000000  2586.450000
498.921000

           ZYX_S2
1264  5231.570000
```

```
548   3509.180000
6812  7905.920000
7863  6938.987139
3228  3666.980000

# Keep only the required columns
df_final = df_balanced.set_index("TCGA Participant Barcode")
[column_names_PanCan + ["Immune Subtype"]]

# Display the first few rows
df_final.head()
```

```
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/
format.py:1458: RuntimeWarning: invalid value encountered in greater
  has_large_values = (abs_vals > 1e6).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in less
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
/usr/local/lib/python3.10/dist-packages/pandas/io/formats/format.py:14
59: RuntimeWarning: invalid value encountered in greater
  has_small_values = ((abs_vals < 10 ** (-self.digits)) & (abs_vals >
0)).any()
```

TCGA Participant Barcode	ACTL6A_S5	ADAM9_S2	ADAMTS1_S5	ADCY7_S3
TCGA-WA-A7H4	1414.260000	5574.520000	942.308000	946.314000
TCGA-B5-A3F9	1354.770000	1138.580000	1640.530000	1227.540000
TCGA-49-6767	1637.870000	5176.330000	161.538000	1004.730000
TCGA-AA-A03J	930.343078	1329.678264	216.256995	370.030352
TCGA-77-7337	3289.310000	14204.300000	2384.660000	573.328000

TCGA Participant Barcode	AIMP2_S5	ALKBH7_S5	ALOX5AP_S3	AMPD3_S3
TCGA-WA-A7H4	563.534000	521.635000	237.1790	587.340000
TCGA-B5-A3F9	482.911000	872.698000	185.8710	353.850000
TCGA-49-6767	488.988000	299.408000	482.2480	276.331000
TCGA-AA-A03J	1349.628189	1956.948165	131.8582	

```
149.806773
TCGA-77-7337              1188.010000   317.341000    743.7580
308.820000

                           APITD1_S5     APOC1_S3      APOE_S3
APOO_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              216.987000   138.622000  1125.000000
110.577000
TCGA-B5-A3F9              340.800000  1306.070000  4475.310000
620.729000
TCGA-49-6767              274.065000  1394.080000  5431.360000
372.781000
TCGA-AA-A03J              125.083134   400.202368  2667.522991
252.673868
TCGA-77-7337              368.092000   126.118000   620.366000
315.296000

                           ARHGAP1_S2  ARHGAP15_S3    ARHGDIA_S2
ARRB2_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              3838.940000    44.070500   8888.620000
523.237000
TCGA-B5-A3F9              1648.980000    12.921500   6709.740000
521.333000
TCGA-49-6767              1626.630000    71.597600   5318.340000
695.858000
TCGA-AA-A03J              3276.934604    12.524268  15147.247057
1306.517461
TCGA-77-7337              4495.270000    28.632300   8677.970000
564.124000

                                 B2M_S3     BCCIP_S5    BRCA2_S5
BRIP1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              48758.800000   951.234000  224.359000
169.071000
TCGA-B5-A3F9              66318.100000  1651.590000   70.571300
51.189000
TCGA-49-6767              53988.100000  1667.410000  109.468000
92.307700
TCGA-AA-A03J              41306.987107   769.615704  146.312249
36.943463
TCGA-77-7337              33636.100000  1540.930000  147.593000
209.970000

                              BSG_S2      BTK_S3  C11orf24_S5
```

```
C12orf24_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              16931.90000   86.538500  1333.330000
66.50640
TCGA-B5-A3F9              21927.30000   60.631700   754.417000
93.92940
TCGA-49-6767               8614.20000  156.805000  1263.310000
68.04730
TCGA-AA-A03J              19473.23631   28.087146  1035.470316
53.88891
TCGA-77-7337              13157.20000   85.896900  1387.980000
84.19260

                          C13orf1_S5  C13orf18_S3  C13orf27_S5
C16orf61_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              310.096000   127.404000   193.910000
358.758000
TCGA-B5-A3F9              273.836000   199.786000   180.404000
548.716000
TCGA-49-6767              381.657000    38.461500   166.864000
548.941000
TCGA-AA-A03J              492.691892  2546.069144   117.261586
479.162611
TCGA-77-7337              330.976000    44.652700   250.533000
565.549000

                          C19orf48_S5  C1orf54_S3    C3AR1_S3
C3orf26_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               976.763000  155.449000  133.814000
787.660000
TCGA-B5-A3F9              1952.140000  203.265000  154.561000
635.141000
TCGA-49-6767               863.314000  137.278000  370.414000
468.047000
TCGA-AA-A03J              4435.195669   66.573924  130.132203
375.363702
TCGA-77-7337              1145.290000   81.806600  369.493000
1113.930000

                            CASP1_S3      CBX1_S5    CCBL2_S5
CCL2_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              790.865000  1612.180000  716.090000
351.763000
```

```
TCGA-B5-A3F9               120.766000  2532.120000  626.022000
620.232000
TCGA-49-6767               199.408000  1936.090000  795.284000
642.604000
TCGA-AA-A03J               463.236328   620.153334  424.498066
122.172779
TCGA-77-7337               672.177000  2124.930000  621.464000
1044.060000

                                CCL5_S3    CCRL2_S3       CCT5_S5
CD14_S4  \
TCGA Participant Barcode

TCGA-WA-A7H4               1123.400000   36.057700   7229.970000
1347.76000
TCGA-B5-A3F9                500.957000   27.333900   6531.320000
1638.55000
TCGA-49-6767               2890.530000   71.005900  10427.200000
1537.87000
TCGA-AA-A03J                607.562311  370.657909   8111.215987
1392.85134
TCGA-77-7337                555.944000   62.377500  12827.600000
1825.99000

                              CD163_S3   CD19_S4    CD247_S4    CD33_S3
\
TCGA Participant Barcode

TCGA-WA-A7H4                320.513000  9.615400  101.763000  31.250000

TCGA-B5-A3F9                583.456000  5.466800   56.158800  67.092400

TCGA-49-6767                834.319000  1.775100  133.728000  71.005900

TCGA-AA-A03J                164.663887  8.810904   71.947677  30.009939

TCGA-77-7337               1267.660000  3.408600   63.400100  46.357100


                             CD37_S4     CD3D_S4     CD3E_S4    CD3G_S4
\
TCGA Participant Barcode

TCGA-WA-A7H4               169.8720   92.948700  187.500000  14.423100

TCGA-B5-A3F9               190.8410   93.432400  194.320000   4.472800

TCGA-49-6767               450.8880  200.592000  602.367000  56.804700

TCGA-AA-A03J               246.5297  109.859681  189.354061  17.990105
```

```
TCGA-77-7337              182.0200   79.420500  256.327000  41.244100


                                CD44_S2     CD48_S3     CD52_S3
CD53_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              18930.300000   94.551300  186.699000
338.14100
TCGA-B5-A3F9               2058.990000  157.543000  171.458000
465.17400
TCGA-49-6767               9745.560000  253.846000  629.586000
863.90500
TCGA-AA-A03J               3261.290314   96.453995  268.239955
315.95604
TCGA-77-7337              23292.700000  132.595000  219.514000
878.05700

                                CD59_S2    CD79A_S4   CD79B_S4
CD84_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              14369.400000  133.814000  56.089700
20.833300
TCGA-B5-A3F9               6228.660000   16.400400  18.388300
28.824900
TCGA-49-6767               6684.020000   37.869800  22.485200
62.721900
TCGA-AA-A03J               2817.520364   64.726949  33.807438
25.931826
TCGA-77-7337              13698.500000  174.862000  31.359200
117.597000

                              CD86_S3    CD8A_S3     CD8B_S4
CD97_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              233.173000   33.65380   24.038500
761.218000
TCGA-B5-A3F9              104.863000  197.79800  229.605000
1288.170000
TCGA-49-6767              304.734000  652.07100   97.041400
1813.020000
TCGA-AA-A03J               77.030344  276.25988   77.724818
6066.851717
TCGA-77-7337              397.103000  208.60700  102.940000
1295.950000

                              CDCA4_S5     CDH6_S2      CDK2_S5
CECR1_S3  \
```

```
TCGA Participant Barcode

TCGA-WA-A7H4               1024.840000    84.13460   818.910000
422.276000
TCGA-B5-A3F9                685.834000    69.57730   876.177000
1813.480000
TCGA-49-6767                786.982000    11.83430  1046.150000
971.598000
TCGA-AA-A03J                480.374708    23.29196   787.869981
447.176864
TCGA-77-7337               1040.310000  2406.48000   923.392000
659.225000

                             CELF2_S3    CENPJ_S5     CENPN_S5
CENPO_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               397.436000  226.763000   702.941000
202.780000
TCGA-B5-A3F9               328.007000  616.256000   295.157000
74.810500
TCGA-49-6767               270.414000   89.349100   584.787000
145.675000
TCGA-AA-A03J               700.887602  196.492764   630.422866
190.410746
TCGA-77-7337               306.775000  107.712000  1274.420000
436.735000

                               CENPW_S5    CEP78_S5    CHEK1_S5
CKLF_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                568.109000  243.590000  537.660000
213.14100
TCGA-B5-A3F9                454.240000  284.273000  269.364000
227.61700
TCGA-49-6767                353.846000  223.077000  502.959000
897.04100
TCGA-AA-A03J                931.541342   89.391743  175.446418
465.70463
TCGA-77-7337               1029.400000  493.566000  662.633000
235.19400

                            COL16A1_S2      COL1A2_S2      COL3A1_S2  \
TCGA Participant Barcode
TCGA-WA-A7H4               5395.030000   93153.000000  103386.000000
TCGA-B5-A3F9                182.392000   10381.900000   22155.900000
TCGA-49-6767                710.059000   21293.500000   21016.000000
TCGA-AA-A03J                681.668235   30497.210498   23821.008655
TCGA-77-7337                862.378000  105622.000000  155202.000000
```

```
                                COL6A1_S2     COL6A3_S2  COL8A1_S2
COPS6_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               20055.300000  26830.900000  208.33300
1607.370000
TCGA-B5-A3F9                9287.580000   1272.270000    9.93960
2249.340000
TCGA-49-6767               18462.100000   8066.860000  395.85800
8586.390000
TCGA-AA-A03J                8179.105901   8199.892726  237.28691
2323.238838
TCGA-77-7337                6743.590000  34333.200000  896.12300
1875.760000

                              COQ2_S5    CORO1A_S3    CORO1C_S5
COTL1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               274.038000   713.141000  3767.630000
2705.130000
TCGA-B5-A3F9               250.478000   567.552000  1157.970000
3072.830000
TCGA-49-6767               269.822000  1411.830000  5149.700000
9766.860000
TCGA-AA-A03J               333.605287  3106.212752  3463.398929
2085.454571
TCGA-77-7337               253.941000   495.271000  8162.930000
3322.710000

                             COX17_S5    CPEB4_S5    CPVL_S3
CSF1R_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               538.462000  811.699000  203.52600
724.359000
TCGA-B5-A3F9               778.769000  460.701000  230.10200
653.033000
TCGA-49-6767               839.645000  245.562000  393.49100
988.757000
TCGA-AA-A03J               591.715669  303.291589  222.88383
671.456392
TCGA-77-7337               399.489000  894.759000  299.95700
1158.930000

                           CTNNA1_S2     CTNNB1_S2       CTSC_S3
CTSL1_S3  \
TCGA Participant Barcode
```

```
TCGA-WA-A7H4              8176.2800   8176.970000   5598.560000
1512.820000
TCGA-B5-A3F9              7995.4300   6567.580000   3442.090000
2101.730000
TCGA-49-6767              6928.4000  13272.800000   4711.830000
1643.790000
TCGA-AA-A03J              8261.3316   9974.124921   2538.514223
658.990557
TCGA-77-7337              7950.5800  12085.200000  11806.700000
2636.220000

                                CTSS_S3    CXCL10_S1     CXCR4_S3
CYBB_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                 835.737000   187.500000   298.077000
471.154000
TCGA-B5-A3F9                1766.270000   308.625000  2493.850000
468.156000
TCGA-49-6767               10181.700000  2600.000000   601.775000
1659.760000
TCGA-AA-A03J                3005.643104   304.784386   270.766523
350.122639
TCGA-77-7337                2027.100000   568.215000  1160.630000
1642.610000

                             CYTIP_S3    DAPK1_S3   DBNDD1_S5
DCBLD2_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              129.80800  905.449000  197.115000
3229.170000
TCGA-B5-A3F9              163.50700  462.689000  824.988000
2031.160000
TCGA-49-6767              208.87600  169.822000  862.130000
24820.100000
TCGA-AA-A03J               53.78859  304.985752  337.246512
503.510219
TCGA-77-7337              252.91900  320.409000  167.703000
2947.080000

                               DCK_S5  DCLRE1B_S5    DDX58_S1
DDX60_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4              548.87800  323.718000  770.833000
713.141000
TCGA-B5-A3F9              357.82600  238.551000  387.645000
578.983000
TCGA-49-6767              872.18900  264.497000  452.663000
```

```
763.905000
TCGA-AA-A03J              352.77352  142.552052  396.642255
705.860341
TCGA-77-7337              997.01800  276.438000  222.923000
731.487000

                             DHFR_S5    DLEU1_S5    DLEU2_S5
DOCK2_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              225.561000   80.769200   69.487200
203.526000
TCGA-B5-A3F9              422.026000  201.471000   49.285600
100.390000
TCGA-49-6767              227.876000   92.112400  107.444000
321.302000
TCGA-AA-A03J              388.993479  134.812507   25.400413
146.765031
TCGA-77-7337              175.202000  170.703000  102.940000
315.296000

                                DSP_S2       DUT_S5      DVL1_S2
DVL3_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4              44610.600000  1073.000000  2220.350000
2595.350000
TCGA-B5-A3F9               4294.410000  1422.170000  1487.960000
2026.190000
TCGA-49-6767               8887.570000   624.645000  1872.780000
2149.110000
TCGA-AA-A03J               8165.998316  1715.641893  3209.336764
2974.540775
TCGA-77-7337              52847.400000   577.541000   804.431000
4245.420000

                             DYNLT1_S5  EBNA1BP2_S5    EFNA5_S2
EIF2AK1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              1262.020000   2073.72000  518.429000
3613.550000
TCGA-B5-A3F9              2670.780000   1220.58000   64.607500
5030.100000
TCGA-49-6767              1210.650000   1886.98000   18.343200
4418.110000
TCGA-AA-A03J              1160.197487   1246.37362    9.236864
5112.901411
TCGA-77-7337              1212.100000   1114.96000  297.912000
5569.550000
```

```
                            EIF4EBP1_S5     EIF4G1_S5      EMP2_S5  \
TCGA Participant Barcode
TCGA-WA-A7H4                1009.620000  12786.100000  2140.220000
TCGA-B5-A3F9                1500.880000   6257.490000  1126.160000
TCGA-49-6767                1133.140000  13284.600000  2677.510000
TCGA-AA-A03J                1457.794599  10527.587171  1362.348884
TCGA-77-7337                3609.710000  16915.900000  3410.310000

                                  ENO1_S5     EPHA2_S2     EPHB2_S2
EPHB3_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                 45858.200000  1617.790000   504.808000
1241.99000
TCGA-B5-A3F9                 28507.300000   631.663000   450.265000
351.86200
TCGA-49-6767                 47065.700000  1069.230000  2076.330000
550.29600
TCGA-AA-A03J                 23591.512322  3201.655891  4132.074258
5667.72571
TCGA-77-7337                143228.000000  2470.900000  2208.440000
2232.64000

                               EPHB4_S2   ERLIN1_S5    EVI2A_S3
EVI2B_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                 2685.100000   955.12800   88.141000
157.051000
TCGA-B5-A3F9                 1916.860000  1254.88000   81.504900
127.724000
TCGA-49-6767                13614.800000  2472.19000  200.592000
373.964000
TCGA-AA-A03J                 5248.529283  1297.53771   22.786344
81.030267
TCGA-77-7337                 1957.900000  2386.71000  232.467000
320.068000

                            EXOSC8_S5        EZR_S5        F3_S5
FAM167A_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                427.72400  11897.400000  8418.270000
90.544900
TCGA-B5-A3F9                481.60400   5497.600000   474.120000
20.873200
TCGA-49-6767                585.28400  17770.400000  1340.240000
69.822500
TCGA-AA-A03J                471.38021  10746.871452   510.121969
```

```
39.215757
TCGA-77-7337              306.17100   9246.530000  1477.970000
29.314000

                            FAM89B_S5     FARSA_S5     FARSB_S5
FBXO41_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              1355.770000  1358.970000  1342.150000
324.519000
TCGA-B5-A3F9              1135.600000  1468.080000  1338.370000
51.189000
TCGA-49-6767              2065.680000   919.527000  1501.780000
334.911000
TCGA-AA-A03J              1077.730029  3382.821044  1201.770979
734.838076
TCGA-77-7337               884.874000  1493.310000  1174.270000
85.215200

                          FCER1G_S3  FCGR1A_S3    FCGR2A_S3
FCGR2B_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              226.76300   27.24360   520.833000
120.994000
TCGA-B5-A3F9              541.21200  121.24300   434.361000
99.396200
TCGA-49-6767              937.27800  286.95300  1111.240000
200.592000
TCGA-AA-A03J              215.44325   30.38253   477.218772
35.273426
TCGA-77-7337              552.87600   85.53220  2778.360000
283.255000

                            FCGR3A_S3     FGD1_S2   FGL2_S3
FGR_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               465.545000  204.327000  303.6860
165.064000
TCGA-B5-A3F9               828.964000  830.952000  142.6330
87.965600
TCGA-49-6767              2390.530000  876.923000  492.3080
343.195000
TCGA-AA-A03J               420.841219  253.368838  250.4135
157.269486
TCGA-77-7337              1908.480000  415.850000  512.3140
203.494000

                             FLI1_S3       FLNC_S5     FMNL1_S3
```

```
FN1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               245.192000   5137.820000   574.103000
52455.900000
TCGA-B5-A3F9               137.664000     91.444500   138.658000
2513.230000
TCGA-49-6767               191.124000  11415.400000  1997.630000
54245.000000
TCGA-AA-A03J               106.790429    994.628822   425.568343
7291.669478
TCGA-77-7337               254.282000    157.478000   345.292000
98427.600000

                              FNBP1_S3     FPR3_S3       FYB_S3
GGH_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                741.987000  255.609000   406.250000
1058.490000
TCGA-B5-A3F9               1015.830000   91.444500   315.086000
801.630000
TCGA-49-6767               1497.040000  469.822000   505.917000
420.710000
TCGA-AA-A03J                639.341658  182.866036    84.920137
725.162161
TCGA-77-7337                537.537000  563.443000  2971.280000
1462.970000

                             GIMAP4_S3     GLRX3_S5    GNG11_S5
GNPTAB_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               187.500000  1020.030000  451.122000
1433.490000
TCGA-B5-A3F9               278.806000   952.215000  611.783000
1293.140000
TCGA-49-6767               485.207000  2571.010000  155.030000
735.503000
TCGA-AA-A03J               182.790949  1104.235578  118.798676
664.235627
TCGA-77-7337               341.542000  2100.380000  263.826000
2109.590000

                           GPLD1_S5      GPNMB_S3    GSTCD_S5
H2AFZ_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               4.807700   9549.680000  254.808000
5116.190000
```

```
TCGA-B5-A3F9               8.945700   2307.480000  318.565000
6503.990000
TCGA-49-6767               8.875700   6445.560000  238.462000
4285.210000
TCGA-AA-A03J               5.973108    656.968787  174.494972
4937.989857
TCGA-77-7337               9.203200  38693.500000  366.425000
5925.860000

                              HAS2_S5    HAUS1_S5      HCK_S3
HCLS1_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              198.638000  183.494000  326.122000
532.85300
TCGA-B5-A3F9               77.032000  531.273000  266.879000
777.77500
TCGA-49-6767               25.390500  150.296000  675.740000
770.41400
TCGA-AA-A03J               53.483964  190.910932  343.856397
342.87048
TCGA-77-7337              753.104000  280.869000  450.959000
1998.13000

                              HDC_S4    HERC5_S1    HERC6_S1   HLA-
DMA_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              66.506400  175.481000  601.763000
532.853000
TCGA-B5-A3F9               1.490900   97.905200  428.397000
2299.030000
TCGA-49-6767               1.775100   63.905300  179.882000
1369.230000
TCGA-AA-A03J              51.860779   36.200517  333.053243
1201.901095
TCGA-77-7337              20.110800  247.806000  414.146000
1186.200000

                           HLA-DRB1_S3     HMGN2_S5     HMHA1_S3
HN1L_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               2939.590000  5672.280000   722.756000
4440.640000
TCGA-B5-A3F9               5338.540000  9221.480000   369.754000
4585.140000
TCGA-49-6767              12539.900000  5878.700000   881.065000
2697.630000
TCGA-AA-A03J               4294.439114  3542.639361  1596.510033
```

```
2639.516134
TCGA-77-7337               4501.430000  6724.500000   230.081000
5817.460000

                          HNRNPA2B1_S5    HNRNPR_S5   HSPB11_S5
HSPG2_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4              20217.900000  3596.150000  517.628000
12338.200000
TCGA-B5-A3F9              22632.000000  3508.680000  528.788000
4327.120000
TCGA-49-6767              12819.500000  2786.390000  257.988000
5197.800000
TCGA-AA-A03J               9956.139522  1741.845332  139.147844
11231.842641
TCGA-77-7337              10226.500000  3903.880000  208.607000
13568.800000

                            HYLS1_S5     ICAM1_S2       ID2_S5
ID3_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              217.949000  1789.260000   681.090000
822.917000
TCGA-B5-A3F9              124.742000   685.834000  3584.230000
4047.910000
TCGA-49-6767              156.805000  7046.750000   600.592000
244.379000
TCGA-AA-A03J               46.389359  1215.882523  1846.198195
1449.654233
TCGA-77-7337              298.935000  3448.830000   686.493000
932.595000

                             IFI16_S3     IFI27_S1     IFI30_S3
IFI44_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4              6826.120000  1596.150000  2146.550000
956.731000
TCGA-B5-A3F9              1736.450000  4601.050000  2636.480000
437.840000
TCGA-49-6767              2541.420000  5004.730000  5024.040000
846.746000
TCGA-AA-A03J               455.437316  8435.411981  4218.586672
126.583741
TCGA-77-7337              7451.210000  3560.630000  4289.150000
233.830000

                           IFI44L_S1      IFI6_S1    IFIH1_S1
```

```
IFIT1_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4               293.26900  2016.830000  541.667000
358.173000
TCGA-B5-A3F9               476.60500  5028.450000  457.719000
296.698000
TCGA-49-6767              2188.17000   748.521000  908.876000
218.935000
TCGA-AA-A03J               358.27975  3904.530045  365.052522
280.111308
TCGA-77-7337                89.30550   989.859000  632.297000
137.367000

                            IFIT2_S1     IFIT3_S1     IFRD2_S5
IGF2R_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4              173.077000   522.436000  1132.650000
7806.090000
TCGA-B5-A3F9              255.945000   367.269000   967.299000
1889.020000
TCGA-49-6767              714.793000  1108.280000  1590.020000
5543.200000
TCGA-AA-A03J              218.129019   746.590341  2946.979229
5759.380637
TCGA-77-7337              158.500000   466.979000   821.133000
5435.710000

                            IGFBP2_S2     IGFBP3_S2     IGFBP4_S2  \
TCGA Participant Barcode
TCGA-WA-A7H4              1262.820000   2625.000000   7622.600000
TCGA-B5-A3F9              7163.480000   3840.670000  25464.800000
TCGA-49-6767               101.183000   2001.180000   5388.760000
TCGA-AA-A03J               769.358703   2308.976639   6273.287026
TCGA-77-7337              4615.590000  25963.400000   9842.690000

                             IGFBP5_S2       IGJ_S4    IKZF1_S3
IL10RA_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               4780.450000  6902.240000  229.167000
295.673000
TCGA-B5-A3F9               2210.570000   294.213000  124.245000
228.611000
TCGA-49-6767               5168.640000   202.367000  296.450000
513.610000
TCGA-AA-A03J               3090.207707   879.747499  108.191389
310.704089
TCGA-77-7337              20129.900000  3414.400000  206.562000
```

```
290.754000

                                IL18_S3     IL7R_S3      IMP4_S5
IPO4_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                874.199000  229.968000   741.186000
1429.780000
TCGA-B5-A3F9                 60.134700   31.806800   695.276000
1465.530000
TCGA-49-6767               1228.400000   37.869800   848.521000
1385.810000
TCGA-AA-A03J                264.227307   30.011095  1463.603138
2068.202687
TCGA-77-7337                539.923000  277.801000   795.228000
1700.700000

                              ISG15_S1      ITGA3_S2    ITGA4_S2
ITGA5_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                820.513000  15915.100000  282.340000
3802.080000
TCGA-B5-A3F9               1994.880000    918.421000   97.895300
583.456000
TCGA-49-6767               1202.960000  15562.700000  183.462000
2215.380000
TCGA-AA-A03J               3030.829389   3940.566087  123.141682
1132.763751
TCGA-77-7337                467.320000   6089.480000  358.493000
6613.380000

                                ITGA6_S5     ITGB2_S2    ITGB3_S2
ITGB5_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               25782.900000   801.939000  217.949000
5173.880000
TCGA-B5-A3F9                2266.230000  1002.910000   13.915500
2538.580000
TCGA-49-6767                2964.500000  3699.830000   91.716000
2510.060000
TCGA-AA-A03J                4288.080636  1357.033063   50.598276
2787.829954
TCGA-77-7337               13324.200000  1438.310000  191.905000
2923.220000

                              ITGB8_S2      JUNB_S5        JUP_S2
KIAA0090_S5  \
TCGA Participant Barcode
```

```
TCGA-WA-A7H4              1781.250000  3051.280000  20550.500000
2780.450000
TCGA-B5-A3F9               906.493000  3195.090000   5095.050000
1177.350000
TCGA-49-6767              1984.020000  1322.490000  12150.900000
1236.090000
TCGA-AA-A03J               214.333668  5502.742214  19274.491393
720.656034
TCGA-77-7337               975.884000  2970.600000  33214.500000
4090.670000

                              KRR1_S5   LAIR1_S3     LAMA4_S2
LAMB1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4              1008.450000  183.49400  2374.200000
5812.500000
TCGA-B5-A3F9               633.680000  313.09800   421.440000
7204.730000
TCGA-49-6767               564.284000  499.40800   483.432000
4458.580000
TCGA-AA-A03J               352.075264  208.89802   783.651454
5772.950729
TCGA-77-7337              1403.150000  293.82200  2424.540000
5426.160000

                            LAPTM5_S3      LCK_S4      LCP1_S3
LCP2_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              1458.330000  125.801000  1877.400000
274.038000
TCGA-B5-A3F9              2701.590000  130.706000  2575.350000
169.967000
TCGA-49-6767              6992.900000  328.994000  3215.380000
467.450000
TCGA-AA-A03J              1804.768877  147.246087   817.064039
158.550938
TCGA-77-7337              3877.970000  159.523000  3375.880000
522.539000

                            LHFPL2_S3   LILRB4_S3     LMNB2_S5
LOXL2_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              1033.650000   97.756400  3442.310000
1612.100000
TCGA-B5-A3F9               733.544000  244.018000  2038.620000
444.301000
```

```
TCGA-49-6767                971.598000  468.639000  2414.790000
1230.070000
TCGA-AA-A03J                716.245256  202.107483  4983.516023
1275.832329
TCGA-77-7337               1385.600000  326.204000  3216.020000
2383.960000

                              LRMP_S3       LRP1_S2  LRRC17_S2
LRRC40_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                52.083300   5866.990000  420.67300
415.064000
TCGA-B5-A3F9                29.818900   2684.190000  177.42200
448.774000
TCGA-49-6767                21.301800   3021.890000   79.88170
540.237000
TCGA-AA-A03J                23.441881  11460.799714   27.08007
48.738581
TCGA-77-7337               326.885000  13292.900000  117.25600
318.705000

                              LSM3_S5      LSM4_S5    LST1_S3
LTB_S4  \
TCGA Participant Barcode

TCGA-WA-A7H4               673.878000  1681.090000  126.60300
374.199000
TCGA-B5-A3F9               945.258000  3186.140000   89.45660
162.016000
TCGA-49-6767               910.651000  1820.710000  170.41400
1493.490000
TCGA-AA-A03J               469.580308  2784.870978  104.39482
218.043389
TCGA-77-7337               682.403000  1516.150000   75.67110
22.837700

                              LY86_S3      LYAR_S5       LYN_S3
LYZ_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                25.641000   749.199000   883.814000
13019.200000
TCGA-B5-A3F9                79.020000   386.651000   773.799000
834.431000
TCGA-49-6767               103.550000  1045.560000  2642.600000
6668.050000
TCGA-AA-A03J                38.308787   678.810049   562.685169
3068.651264
TCGA-77-7337                51.129100   860.332000  1834.510000
```

```
6712.910000

                            MAGOHB_S5   MAP3K8_S5    MAPRE1_S5
MARCKSL1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               275.641000  154.647000  3822.120000
1948.720000
TCGA-B5-A3F9               372.239000  268.370000  2547.030000
17631.900000
TCGA-49-6767               364.497000  117.160000  3569.820000
7843.200000
TCGA-AA-A03J               192.543608   80.427806  4168.603566
5833.127538
TCGA-77-7337               458.798000  164.977000  7763.100000
1351.850000

                           MARVELD2_S5      MCM3_S5       MCM7_S5
MCTS1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                252.404000  4629.810000   2780.830000
1647.440000
TCGA-B5-A3F9                424.919000  3315.860000   2835.220000
1015.330000
TCGA-49-6767                400.592000  2574.560000  12199.600000
1617.160000
TCGA-AA-A03J                507.566001  3776.425366   7193.775818
891.683567
TCGA-77-7337                402.897000  1764.290000   3067.410000
774.435000

                             MERTK_S3        MET_S5   MFSD11_S5
MKKS_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               106.570000   3673.080000  856.763000
1009.620000
TCGA-B5-A3F9               442.313000   1659.920000  312.591000
1106.780000
TCGA-49-6767               321.894000  40805.900000  667.325000
592.308000
TCGA-AA-A03J                71.470187   2531.035123  445.909178
433.255505
TCGA-77-7337               337.452000   4123.390000  753.769000
662.292000

                            MLF1IP_S5      MMP11_S2       MMP1_S2  \
TCGA Participant Barcode
TCGA-WA-A7H4               614.728000   8549.680000  33935.100000
```

```
TCGA-B5-A3F9               681.395000  15438.200000     13.915500
TCGA-49-6767               610.467000   7279.290000   3834.910000
TCGA-AA-A03J               269.988979   6572.061363   1598.728465
TCGA-77-7337               379.163000   1861.780000  23431.800000

                                MMP14_S2   MMP17_S2    MMP19_S2
MMP2_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               19246.000000  199.51900  592.949000
11664.300000
TCGA-B5-A3F9                7597.350000   36.27960  793.181000
10317.300000
TCGA-49-6767               29959.800000   90.53250  338.462000
4640.240000
TCGA-AA-A03J                7607.790173   26.74886  308.075381
6768.362318
TCGA-77-7337               12179.000000   66.46780  371.879000
14209.500000

                                MMP3_S2    MNAT1_S5     MNDA_S3
MPP1_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               5686.700000  314.103000   84.134600
214.744000
TCGA-B5-A3F9                102.378000  483.562000   57.152800
176.925000
TCGA-49-6767                 13.017800  378.106000  324.852000
820.710000
TCGA-AA-A03J                896.832598   90.714612   36.692463
862.539391
TCGA-77-7337               1543.080000  366.425000  133.617000
245.079000

                              MRPL12_S5    MRPL37_S5    MRPS16_S5
MRPS28_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               1384.620000  2020.830000  1345.720000
588.878000
TCGA-B5-A3F9               1809.510000  1491.940000  2112.120000
418.667000
TCGA-49-6767               1376.920000  1696.450000  1979.640000
261.840000
TCGA-AA-A03J               3222.565362  3021.732273  2019.185734
158.477752
TCGA-77-7337               1209.710000  1497.060000  2782.770000
461.283000
```

	MRTO4_S5	MS4A1_S4	MSN_S3	MT1F_S5
TCGA Participant Barcode				
TCGA-WA-A7H4	1703.530000	43.269200	11398.200000	167.468000
TCGA-B5-A3F9	1512.310000	1.987900	5326.140000	697.264000
TCGA-49-6767	1215.380000	11.834300	25971.600000	153.254000
TCGA-AA-A03J	976.047526	4.576277	2585.419706	459.608169
TCGA-77-7337	1683.510000	11.248400	17127.900000	172.135000

	MT3_S5	MTA1_S2	MTHFD1_S5	MX1_S1
TCGA Participant Barcode				
TCGA-WA-A7H4	3.205100	1064.900000	1481.530000	1167.470000
TCGA-B5-A3F9	147.106000	1970.030000	1389.490000	1137.590000
TCGA-49-6767	0.000000	1739.050000	2720.590000	1135.500000
TCGA-AA-A03J	91.901208	1555.890562	1733.931097	2361.615361
TCGA-77-7337	0.340900	998.722000	2440.870000	916.915000

	MYBL1_S5	MYBL2_S5	MYCBP_S5	MYL6_S5
TCGA Participant Barcode				
TCGA-WA-A7H4	238.782000	1282.850000	720.353000	15859.000000
TCGA-B5-A3F9	64.110500	1157.970000	1166.910000	18998.100000
TCGA-49-6767	431.361000	1129.590000	715.414000	18542.000000
TCGA-AA-A03J	34.276176	4624.135012	617.681299	14625.742112
TCGA-77-7337	78.738800	3666.640000	489.817000	13088.400000

	MYO1F_S3	NCEH1_S5	NCF2_S3	NCKAP1L_S3
TCGA Participant Barcode				
TCGA-WA-A7H4	259.61500	411.859000	375.801000	

```
237.179000
TCGA-B5-A3F9               301.66700   694.282000    82.001800
190.344000
TCGA-49-6767               450.29600  1969.230000   771.006000
445.562000
TCGA-AA-A03J               438.45008   451.314543   152.308393
256.994692
TCGA-77-7337               201.44900  1193.350000  1061.100000
351.087000

                                NCLN_S5      NEO1_S2      NID1_S2
NLN_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               2168.270000  1062.500000  1839.500000
435.096000
TCGA-B5-A3F9               2321.400000  2178.760000  1702.100000
422.434000
TCGA-49-6767               1911.240000   804.142000  1350.880000
600.592000
TCGA-AA-A03J               6476.962655  3104.024952  1136.524032
374.185195
TCGA-77-7337               1691.690000   782.957000  5154.120000
511.973000

                                NME1_S5    NOP16_S5    NOTCH2_S2
NPC2_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               1711.670000  379.006000  2720.350000
2388.620000
TCGA-B5-A3F9               3232.230000  589.916000  2957.530000
6377.260000
TCGA-49-6767               2551.910000  322.485000  3546.150000
7479.290000
TCGA-AA-A03J               2190.579905  734.285336   712.373458
1552.987229
TCGA-77-7337               1949.150000  258.372000  3836.730000
6966.510000

                                NPL_S3    NRIP3_S5    NUDT1_S5
NUDT15_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               213.141000   96.955100  445.513000
854.199000
TCGA-B5-A3F9               130.706000   16.400400  327.510000
688.771000
TCGA-49-6767               120.118000  134.320000  288.757000
879.846000
```

```
TCGA-AA-A03J                88.620145   14.386538  402.260555
600.915288
TCGA-77-7337               691.947000  247.124000  491.521000
854.865000

                             NUP107_S5    NUP35_S5    NUP85_S5
NUP93_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               1161.860000  250.801000  905.449000
1128.210000
TCGA-B5-A3F9               1000.420000  360.808000  726.089000
1247.420000
TCGA-49-6767               1000.590000  251.479000  890.533000
1489.940000
TCGA-AA-A03J                711.490968  148.455339  965.871197
1438.451723
TCGA-77-7337               1263.570000  408.010000  727.397000
1467.410000

                               NUPL1_S5     NUTF2_S5      OAS1_S1
OAS2_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4               1248.400000  1809.290000   560.096000
1220.350000
TCGA-B5-A3F9                756.405000  3353.130000   193.326000
385.657000
TCGA-49-6767               1099.410000  3146.750000  1168.050000
1243.200000
TCGA-AA-A03J               1158.237576  3557.105678  1869.724499
935.073373
TCGA-77-7337               1026.670000  2287.860000   477.887000
1499.450000

                              OAS3_S1     OASL_S1    OSBPL3_S3
PA2G4_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               1335.740000  160.256000  1112.980000
2838.140000
TCGA-B5-A3F9                634.148000  114.306000   428.894000
4439.530000
TCGA-49-6767               1981.070000  281.657000   815.385000
5613.610000
TCGA-AA-A03J               2756.552264  533.469981   596.266612
5580.006673
TCGA-77-7337               1513.420000   82.147400   468.002000
4908.050000
```

	PAICS_S5	PAK1_S2	PAK2_S2	PDAP1_S5	\
TCGA Participant Barcode					
TCGA-WA-A7H4	3400.640000	1366.190000	3371.790000	2505.610000	
TCGA-B5-A3F9	4425.120000	2163.850000	2059.490000	3060.410000	
TCGA-49-6767	2313.610000	2145.560000	4628.990000	10816.600000	
TCGA-AA-A03J	3331.453327	1544.604524	1565.465184	3126.018611	
TCGA-77-7337	5149.380000	2339.330000	8009.890000	1954.840000	

	PDIA4_S5	PDLIM7_S5	PFKP_S5	\
TCGA Participant Barcode				
TCGA-WA-A7H4	5277.240000	3774.840000	1290.870000	
TCGA-B5-A3F9	12363.900000	1167.900000	834.431000	
TCGA-49-6767	16620.700000	1638.460000	3407.100000	
TCGA-AA-A03J	6665.578363	1790.815205	3396.808005	
TCGA-77-7337	6801.870000	838.858000	4072.260000	

	PFN1_S5	PGM2_S5	PHF19_S5	PIK3CG_S3	\
TCGA Participant Barcode					
TCGA-WA-A7H4	18421.500000	1930.290000	897.43600	64.90380	
TCGA-B5-A3F9	14217.600000	537.733000	265.88500	7.95170	
TCGA-49-6767	12803.000000	1115.980000	1609.47000	47.33730	
TCGA-AA-A03J	21598.814199	543.057721	1775.78342	23.68202	
TCGA-77-7337	8874.990000	2447.720000	550.49000	99.19050	

	PITPNC1_S5	PLAT_S2	PLAUR_S5	PLCG2_S3	\
TCGA Participant Barcode					
TCGA-WA-A7H4	311.699000	434.295000	758.013000	356.570000	
TCGA-B5-A3F9	519.345000	550.158000	603.832000	254.951000	
TCGA-49-6767	175.148000	2979.880000	2352.660000	207.101000	
TCGA-AA-A03J	214.514288	659.910176	2324.934366	114.980755	

```
TCGA-77-7337              411.760000  3454.960000  1480.700000
536.515000

                             PLEK_S3     PLG_S5     PLK4_S5
PLOD2_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              229.167000   0.000000  506.410000
1907.850000
TCGA-B5-A3F9              299.679000  23.855100  258.927000
2288.100000
TCGA-49-6767              688.757000   0.000000  369.231000
1333.730000
TCGA-AA-A03J              161.426919   3.938344  153.655922
439.280843
TCGA-77-7337              561.057000   0.000000  387.218000
7612.440000

                               PNN_S5      PNO1_S5   POLE2_S5
POLR3K_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              2253.210000   594.551000  126.60300
274.038000
TCGA-B5-A3F9              2935.170000   393.112000   89.45660
389.633000
TCGA-49-6767              1604.730000   665.681000  140.82800
243.787000
TCGA-AA-A03J              1478.458546   599.694356   67.87253
574.739875
TCGA-77-7337              1761.910000  1266.980000  168.72600
508.223000

                             PPIH_S5      PSMA7_S5     PSMC3_S5
PSMD12_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              492.78800   4498.400000  2745.990000
1370.990000
TCGA-B5-A3F9              743.98000   2975.920000  2598.220000
723.107000
TCGA-49-6767              580.47300   4809.470000  1921.890000
1149.110000
TCGA-AA-A03J              652.58254  12850.164704  1866.885196
501.157671
TCGA-77-7337              320.75000   6267.750000  2897.320000
2220.710000

                             PSMD14_S5      PSMD2_S5    PTPLB_S5
PTPRC_S3  \
```

```
TCGA Participant Barcode

TCGA-WA-A7H4               1602.560000   6587.340000  308.494000
516.026000
TCGA-B5-A3F9               1436.770000   3072.830000   43.237300
339.438000
TCGA-49-6767               1317.160000   5385.800000  243.195000
875.148000
TCGA-AA-A03J                978.831589   4231.503959   75.527356
208.733265
TCGA-77-7337               3146.480000  14749.000000  605.709000
1036.560000

                                PXN_S2   RAB3B_S5       RAC1_S2
RASSF2_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               6759.01000   5.609000   6165.060000
365.385000
TCGA-B5-A3F9               1946.43000   1.987900   6536.290000
648.063000
TCGA-49-6767               7575.14000  26.627200   6388.170000
122.485000
TCGA-AA-A03J               7988.21454   7.277026   7220.930029
122.210979
TCGA-77-7337               4754.74000  16.020500  11040.800000
252.578000

                              RASSF4_S3     RBM14_S5     RFC3_S5
RGS8_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                736.114000  1018.350000  365.385000
0.0000
TCGA-B5-A3F9                825.982000  1355.860000  442.810000
0.0000
TCGA-49-6767               1518.930000  1006.050000  404.734000
0.0000
TCGA-AA-A03J                772.989068  2152.712306  546.375419
NaN
TCGA-77-7337                215.083000  1341.370000  546.740000
0.3409

                                 RHOA_S2      RHOB_S2     RHOC_S5
RHOG_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                9258.810000   2931.09000  3935.10000
1266.830000
TCGA-B5-A3F9                9805.930000  11439.50000  3642.37000
```

```
898.541000
TCGA-49-6767              13426.000000   1077.51000  6773.37000
1168.050000
TCGA-AA-A03J               9807.319352   3607.80426  5682.28049
1714.439988
TCGA-77-7337              10884.400000   2874.14000  3976.82000
885.215000

                              RHOQ_S2   RMND5B_S5   RNASE6_S3
RND3_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4              1629.020000  351.763000  143.429000
1867.790000
TCGA-B5-A3F9              1649.260000  622.717000  124.742000
2087.320000
TCGA-49-6767              1291.900000  531.953000  175.148000
1360.360000
TCGA-AA-A03J               360.002602  937.189697  142.656267
438.776252
TCGA-77-7337              1348.960000  273.711000  116.574000
5216.870000

                           RNF138_S5     RNF41_S5       RPN1_S5
RPP40_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              368.590000   684.295000   6653.040000
213.141000
TCGA-B5-A3F9              763.363000   879.159000  11263.100000
134.682000
TCGA-49-6767              431.361000   828.402000   7985.800000
184.615000
TCGA-AA-A03J              284.101016  1102.745864   7524.043806
165.900124
TCGA-77-7337              601.278000  1178.700000  15401.800000
252.578000

                            RSAD2_S1     RTP4_S1     RUNX3_S3
RUVBL1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              165.865000   92.147400   657.853000
1214.70000
TCGA-B5-A3F9               90.947500  193.326000   193.822000
1955.01000
TCGA-49-6767              308.876000  133.136000   204.142000
1519.73000
TCGA-AA-A03J              630.334114  123.607153  1165.381881
1381.89423
```

```
TCGA-77-7337              294.844000  259.395000  1033.150000
2640.05000

                             SAMD9_S1    SAMHD1_S3   SAMSN1_S3
SAR1A_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              2354.970000   495.633000  111.378000
2307.690000
TCGA-B5-A3F9               236.066000  1195.850000  123.748000
2155.900000
TCGA-49-6767               305.325000   756.562000  193.491000
1915.380000
TCGA-AA-A03J               380.608313   628.068511   74.518924
1033.663388
TCGA-77-7337              1423.430000  1060.230000  184.406000
3508.140000

                            SAR1B_S5       SDC1_S5     SELL_S3
SELPLG_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              526.44200  16976.800000  138.622000
278.045000
TCGA-B5-A3F9              437.84000   2425.760000  120.269000
216.187000
TCGA-49-6767              364.49700  10459.200000  159.172000
585.799000
TCGA-AA-A03J              291.35525   7002.780348   51.673906
422.963295
TCGA-77-7337              580.82700  22700.000000  144.184000
383.468000

                             SEMA3F_S2  SERPINE1_S2     SH2B3_S3
SH3BP5L_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              1383.810000   5875.00000   687.500000
1310.900000
TCGA-B5-A3F9               972.095000    117.28700   254.951000
594.389000
TCGA-49-6767               575.148000   1481.66000  1301.780000
553.254000
TCGA-AA-A03J               536.348425    237.63618   504.611923
983.091103
TCGA-77-7337              1699.530000   6642.01000   707.627000
593.779000

                             SKA1_S5     SKA2_S5   SLC16A1_S5
SLC1A3_S3  \
```

```
TCGA Participant Barcode

TCGA-WA-A7H4                93.750000  786.859000  3054.490000
477.564000
TCGA-B5-A3F9               108.839000  842.880000  1665.380000
179.410000
TCGA-49-6767               128.994000  793.491000  2778.700000
168.047000
TCGA-AA-A03J               157.042894  216.737482  1117.373004
26.435782
TCGA-77-7337               261.099000  831.018000  6389.430000
1772.130000

                           SLC25A40_S5    SLC25A5_S5  SLC7A7_S3
SMC2_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                132.212000  11109.000000  176.28200
1104.17000
TCGA-B5-A3F9                 59.140700   4494.690000  183.88300
605.82000
TCGA-49-6767                346.154000   8710.650000  632.54400
1209.47000
TCGA-AA-A03J                120.723276  18811.796336  363.96797
315.73208
TCGA-77-7337                227.695000  10477.400000  353.81300
2964.47000

                                SMO_S2       SMS_S5    SMURF2_S5
SNRPA1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                475.160000   888.622000   483.974000
424.680000
TCGA-B5-A3F9               1728.000000  2884.480000   338.444000
595.880000
TCGA-49-6767                931.953000  3064.500000  1871.010000
614.201000
TCGA-AA-A03J                326.766597  1884.414469   243.065824
537.599672
TCGA-77-7337                509.587000  2040.730000   571.283000
855.901000

                              SNRPA_S5     SNRPB_S5     SNRPC_S5
SNRPD1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               1115.380000  3641.030000  1903.040000
712.340000
TCGA-B5-A3F9               2843.230000  4689.010000  1951.640000
```

```
1430.310000
TCGA-49-6767              1108.280000  4876.920000  2188.760000
981.065000
TCGA-AA-A03J              2810.940332  4348.216621  2118.107643
686.240596
TCGA-77-7337              1157.900000  2653.260000  2204.690000
1225.050000

                             SNRPE_S5     SNX17_S5   SP140_S3
SPAG17_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               600.962000  2597.760000  53.685900
41.666700
TCGA-B5-A3F9              1544.620000  2713.520000  26.340000
289.740000
TCGA-49-6767              1222.490000  2168.640000  97.041400
1.775100
TCGA-AA-A03J               992.908169  2212.668745  21.068411
1.794282
TCGA-77-7337               971.794000  2252.070000  33.063500
68.172100

                            SPARC_S2       SRM_S5     STAT1_S1
STK17A_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              43480.000000  1790.060000  3429.490000
1818.110000
TCGA-B5-A3F9              10374.000000  1435.280000  4222.850000
463.186000
TCGA-49-6767              12467.500000  2275.150000  8585.800000
2156.800000
TCGA-AA-A03J              12455.566615  2229.710794  3373.295658
204.670983
TCGA-77-7337              53367.500000  2054.370000  7263.060000
1204.600000

                           STRA13_S5       SYK_S3    TAGLN_S5
TAP1_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4              697.917000   607.372000  7978.37000
3900.64000
TCGA-B5-A3F9              896.553000  1712.600000   899.53500
4301.37000
TCGA-49-6767              507.692000   547.337000  1727.81000
7538.46000
TCGA-AA-A03J              944.119588  1408.615337  5412.55443
11272.08586
```

```
TCGA-77-7337              260.418000  1351.850000  5734.30000
3268.85000

                            TBXAS1_S3     TCEB1_S5   TCF7L2_S2
THBS1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               156.250000  1051.280000  436.699000
16249.200000
TCGA-B5-A3F9               378.202000   883.632000  704.222000
1140.570000
TCGA-49-6767               170.414000  1136.090000  855.621000
1164.500000
TCGA-AA-A03J              1303.232132   570.101003  977.233934
2851.471087
TCGA-77-7337               170.089000  1231.530000  709.331000
8540.260000

                              THBS2_S2     TIMP1_S2      TLR2_S3
TMEM130_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               6189.100000  4869.390000   669.071000
9.61540
TCGA-B5-A3F9                136.670000  4652.730000   343.911000
5.96380
TCGA-49-6767               2704.730000  6509.470000   218.935000
27.81070
TCGA-AA-A03J               1790.479229  8027.404873   107.980266
14.09625
TCGA-77-7337              14896.000000  7015.590000  1279.930000
54.53770

                                TNC_S2   TNFAIP3_S3  TNFRSF12A_S5  \
TCGA Participant Barcode
TCGA-WA-A7H4              51591.300000  1375.800000   3986.380000
TCGA-B5-A3F9               1885.550000  1716.570000   1070.000000
TCGA-49-6767              18590.500000   947.929000   2894.080000
TCGA-AA-A03J               2139.767469  1004.281681   2584.787646
TCGA-77-7337              36879.800000  1954.840000   1190.290000

                          TNFRSF1A_S2  TNFRSF1B_S3  TNFSF13B_S3
TOMM40_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              3036.860000   710.737000    26.442300
1080.930000
TCGA-B5-A3F9              2660.840000  1382.600000    18.885300
1375.150000
TCGA-49-6767              4164.500000  1972.780000    99.408300
```

```
1551.480000
TCGA-AA-A03J              3277.586534  2469.421757    28.323496
2911.791666
TCGA-77-7337              5124.500000   652.067000    72.262500
1441.840000

                                  TPI1_S5      TPM1_S5      TPM2_S5  \
TCGA Participant Barcode
TCGA-WA-A7H4               6519.230000  7943.420000  10874.20000
TCGA-B5-A3F9              11693.000000  1983.870000   1485.48000
TCGA-49-6767              15510.100000  3089.760000   4177.51000
TCGA-AA-A03J              13301.708123  4638.258269   1984.92103
TCGA-77-7337              21580.200000  2700.970000   1286.41000

                                  TPM3_S5    TPRKB_S5     TRA2B_S5
TUBA4A_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               9447.920000  412.660000  3585.740000
7373.400000
TCGA-B5-A3F9               7467.630000  722.113000  3100.170000
119.275000
TCGA-49-6767               8412.430000  335.503000  1897.630000
1947.930000
TCGA-AA-A03J               8429.630182  211.853076  2050.314713
5324.263102
TCGA-77-7337              10409.500000  435.620000  3686.750000
3058.880000

                                 TUBG1_S5      UAP1_S5   UBE2J1_S5
UMPS_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              1071.310000  1302.880000  1580.93000
970.353000
TCGA-B5-A3F9               984.022000  1015.830000  2069.93000
620.729000
TCGA-49-6767              1028.400000   999.408000  2207.10000
849.112000
TCGA-AA-A03J              1245.895273   576.843949  1253.38716
781.709909
TCGA-77-7337              1813.040000  1534.210000  2274.56000
1168.130000

                                UQCR10_S5    USPL1_S5       VCAN_S2
VDAC1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              1076.920000  534.455000   4942.310000
4340.540000
```

```
TCGA-B5-A3F9               1863.180000  506.921000   7417.440000
4943.470000
TCGA-49-6767               1656.800000  478.106000   2764.500000
4909.470000
TCGA-AA-A03J               1179.679495  286.464799   1249.379279
7151.557949
TCGA-77-7337                743.076000  164.295000  14117.100000
5529.780000

                            VSIG4_S3     VTA1_S5    WDHD1_S5   WDR54_S5
\
TCGA Participant Barcode

TCGA-WA-A7H4                81.73080  1551.28000  425.481000  296.47400

TCGA-B5-A3F9               348.88100  1079.44000  187.362000  766.84100

TCGA-49-6767               621.89400  1128.40000  508.284000  151.47900

TCGA-AA-A03J               160.41858   716.27527  101.055152  140.38918

TCGA-77-7337               479.93200  1951.09000  611.163000  215.08300


                              WDR77_S5     WIPF1_S3    WNT2B_S2
WNT8B_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                906.250000   989.583000   28.044900
6.410300
TCGA-B5-A3F9               4555.820000   750.441000   16.897300
1.490900
TCGA-49-6767               1057.400000  1007.100000    5.325400
1.775100
TCGA-AA-A03J               1725.383932   327.268858   31.568168
2.028401
TCGA-77-7337                842.948000  1371.620000  249.169000
0.000000

                               WSB2_S5   ZWILCH_S5       ZYX_S2 Immune
Subtype
TCGA Participant Barcode

TCGA-WA-A7H4               1991.190000  483.061000  5231.570000
C1
TCGA-B5-A3F9               1495.420000  452.988000  3509.180000
C1
TCGA-49-6767               2242.600000  325.953000  7905.920000
C1
TCGA-AA-A03J               2127.285705  266.422502  6938.987139
C1
```

```
TCGA-77-7337              2586.450000  498.921000  3666.980000
C1
```

Imputing Missing Information

```
df_final.isnull().sum().sum()

gene_cols = df_final.columns[:-1]
df_final[gene_cols] =
df_final[gene_cols].fillna(df_final[gene_cols].mean())

df_final.isnull().sum().sum()

0
```

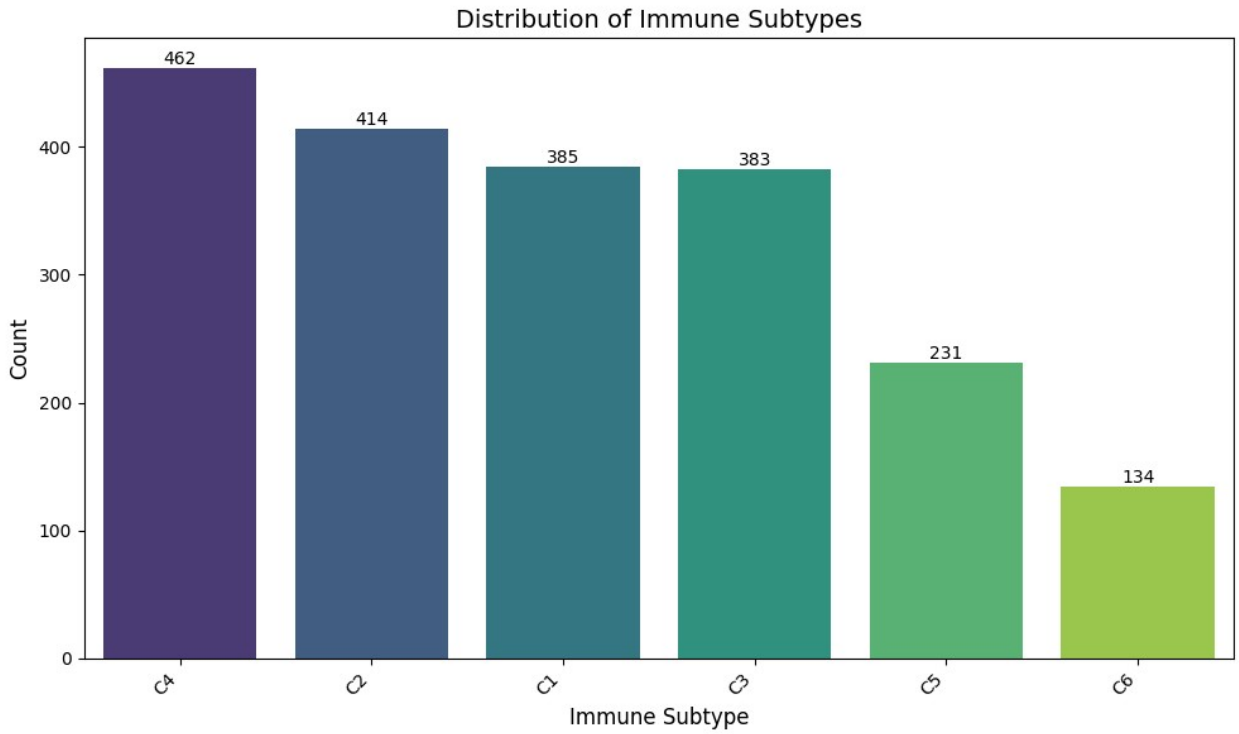
Outputting the final csv

```
#Final 2009 Rows CSV * (440 Input Gene features + 1 Output Immune
Subtype feature)
df_final.to_csv("Final_csv_2009_samples.csv")
```

Visualizing the balanced dataset

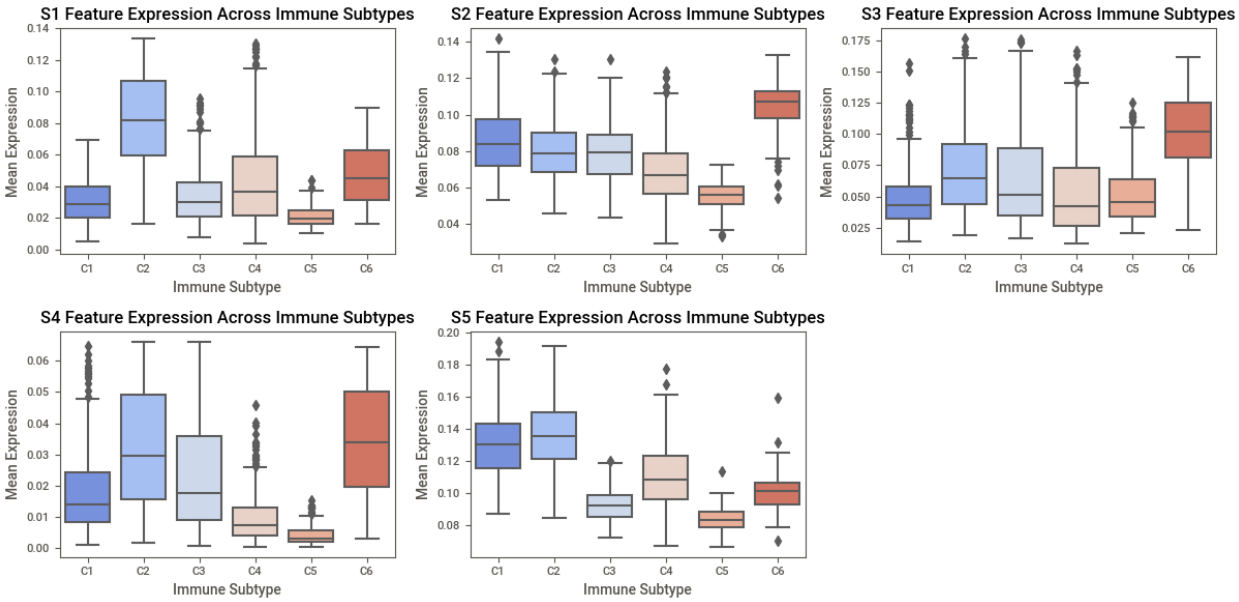
```
# Create the plot
plt.figure(figsize=(10, 6))

# Bar plot using Seaborn
ax = sns.countplot(data=df_final, x='Immune Subtype',
                   order=df_final['Immune
Subtype'].value_counts().index,
                   palette='viridis')

# Add counts on top of bars
for p in ax.patches:
    ax.annotate(f'{p.get_height():.0f}',
                (p.get_x() + p.get_width() / 2., p.get_height()),
                ha='center', va='center',
                xytext=(0, 5),
                textcoords='offset points')

# Customize the plot
plt.title('Distribution of Immune Subtypes', fontsize=14)
plt.xlabel('Immune Subtype', fontsize=12)
plt.ylabel('Count', fontsize=12)
plt.xticks(rotation=45, ha='right')  # Rotate x-labels for better
readability

# Adjust layout to prevent label cutoff
```

```
plt.tight_layout()
plt.show()
```




```
# next steps, remove missingness , nmf - feature reduction,
normalization, standardization, outliers etc
# perform classification



# Group features by suffix (S1, S2, S3, etc.)
feature_groups = {}
for suffix in ['S1', 'S2', 'S3', 'S4', 'S5']:
    feature_groups[suffix] = [col for col in
df_final_normalized.columns if col.endswith(f"_{suffix}")]

# Compute mean expression for each feature group per immune subtype
group_means = pd.DataFrame()
for group, features in feature_groups.items():
    if features:
        group_means[group] =
df_final_normalized[features].mean(axis=1)

# Add immune subtype labels
group_means['Immune Subtype'] = df_final_normalized['Immune Subtype']

# Plot each feature group across different immune subtypes
plt.figure(figsize=(12, 6))
```

```
for i, group in enumerate(feature_groups.keys(), 1):
    plt.subplot(2, 3, i)
    sns.boxplot(x='Immune Subtype', y=group, data=group_means,
palette='coolwarm')
    plt.title(f"{group} Feature Expression Across Immune Subtypes")
    plt.xlabel("Immune Subtype")
    plt.ylabel("Mean Expression")

plt.tight_layout()
plt.show()
```




```
import numpy as np

# Convert categorical immune subtype labels to numerical
immune_subtype_numeric =
LabelEncoder().fit_transform(df_final_normalized['Immune Subtype'])

# Compute correlation for each feature group
correlations = {}
for suffix in ['S1', 'S2', 'S3', 'S4', 'S5']:
    feature_subset = [col for col in df_final_normalized.columns if
col.endswith(f"_{suffix}")]
    if feature_subset:
        mean_expression =
df_final_normalized[feature_subset].mean(axis=1)
        correlations[suffix] = np.corrcoef(mean_expression,
immune_subtype_numeric)[0, 1]

# Print correlation values
for group, corr in correlations.items():
```

```
    print(f"Correlation between {group} features and immune subtype:
{corr:.4f}")
```

```
Correlation between S1 features and immune subtype: -0.1655
Correlation between S2 features and immune subtype: -0.1757
Correlation between S3 features and immune subtype: 0.1368
Correlation between S4 features and immune subtype: -0.1786
Correlation between S5 features and immune subtype: -0.5616
```

```
# Count the number of features for each suffix
feature_counts = {suffix: len(features) for suffix, features in
feature_groups.items()}
print(feature_counts)
```

```
{'S1': 24, 'S2': 80, 'S3': 111, 'S4': 15, 'S5': 210}
```

Scaling and Normalization

```
from sklearn.preprocessing import StandardScaler, MinMaxScaler

# Select only numeric columns for preprocessing
numeric_cols = df_final.select_dtypes(include=['number']).columns

# Standardization (Z-score: mean=0, std=1)
scaler = StandardScaler()
df_final_standardized = df_final.copy()
df_final_standardized[numeric_cols] =
scaler.fit_transform(df_final[numeric_cols])

# Normalization (Min-Max Scaling: range 0-1)
min_max_scaler = MinMaxScaler()
df_final_normalized = df_final.copy()
df_final_normalized[numeric_cols] =
min_max_scaler.fit_transform(df_final[numeric_cols])

df_final_normalized.head()
```

```
                           ACTL6A_S5  ADAM9_S2  ADAMTS1_S5  ADCY7_S3
AIMP2_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.129424  0.169197    0.022398  0.234976
0.180090
TCGA-B5-A3F9                0.123828  0.033567    0.039275  0.306215
0.153894
TCGA-49-6767                0.150459  0.157022    0.003525  0.249774
0.155868
TCGA-AA-A03J                0.083904  0.039410    0.004848  0.088996
0.435516
```

```
TCGA-77-7337                0.305805  0.433055    0.057262  0.140494
0.383001

                           ALKBH7_S5  ALOX5AP_S3  AMPD3_S3
APITD1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.036046    0.025048  0.083360   0.183705

TCGA-B5-A3F9                0.062357    0.019597  0.049824   0.299588

TCGA-49-6767                0.019391    0.051084  0.038690   0.237127

TCGA-AA-A03J                0.143618    0.013858  0.020518   0.097687

TCGA-77-7337                0.020735    0.078867  0.043357   0.325132


                           APOC1_S3   APOE_S3   APOO_S5  ARHGAP1_S2  \
TCGA Participant Barcode
TCGA-WA-A7H4               0.000159  0.001151  0.023024    0.199662
TCGA-B5-A3F9               0.001520  0.004717  0.137673    0.067752
TCGA-49-6767               0.001623  0.005735  0.081950    0.066406
TCGA-AA-A03J               0.000464  0.002793  0.054958    0.165810
TCGA-77-7337               0.000145  0.000614  0.069031    0.239196

                           ARHGAP15_S3  ARHGDIA_S2  ARRB2_S3    B2M_S3
\
TCGA Participant Barcode

TCGA-WA-A7H4                  0.034113    0.193792  0.062812  0.149856

TCGA-B5-A3F9                  0.010200    0.143697  0.062524  0.207232

TCGA-49-6767                  0.055246    0.111708  0.088977  0.166943

TCGA-AA-A03J                  0.009896    0.337684  0.181536  0.125507

TCGA-77-7337                  0.022262    0.188949  0.069010  0.100442


                           BCCIP_S5  BRCA2_S5  BRIP1_S5    BSG_S2
BTK_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.137327  0.135782  0.095438  0.044576
0.034921
TCGA-B5-A3F9               0.263794  0.042710  0.028896  0.060415
0.024206
TCGA-49-6767               0.266651  0.066250  0.052106  0.018203
0.063984
```

```
TCGA-AA-A03J              0.104531  0.088548  0.020854  0.052634
0.010746
TCGA-77-7337              0.243811  0.089324  0.118525  0.032608
0.034656

                          C11orf24_S5  C12orf24_S5  C13orf1_S5
C13orf18_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                 0.152497     0.022158    0.155023
0.016833
TCGA-B5-A3F9                 0.077849     0.031294    0.136670
0.026442
TCGA-49-6767                 0.143469     0.022671    0.191243
0.005027
TCGA-AA-A03J                 0.114090     0.017954    0.247442
0.337895
TCGA-77-7337                 0.159544     0.028050    0.165591
0.005849

                          C13orf27_S5  C16orf61_S5  C19orf48_S5
C1orf54_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                 0.076541     0.045475     0.019964
0.091298
TCGA-B5-A3F9                 0.070951     0.076200     0.041087
0.120955
TCGA-49-6767                 0.065347     0.076236     0.017507
0.080028
TCGA-AA-A03J                 0.044817     0.064950     0.094862
0.036174
TCGA-77-7337                 0.099976     0.078923     0.023613
0.045622

                          C3AR1_S3  C3orf26_S5  CASP1_S3   CBX1_S5
CCBL2_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.033466    0.252371  0.130334  0.104727
0.357476
TCGA-B5-A3F9              0.038793    0.198402  0.018940  0.171738
0.306802
TCGA-49-6767              0.094222    0.139277  0.032013  0.128322
0.402032
TCGA-AA-A03J              0.032520    0.106481  0.075871  0.032466
0.193422
TCGA-77-7337              0.093986    0.367820  0.110604  0.142077
0.304238
```

```
                            CCL2_S3   CCL5_S3  CCRL2_S3   CCT5_S5
CD14_S4  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.017832  0.039156  0.006166  0.139557
0.018676
TCGA-B5-A3F9               0.031532  0.017448  0.004626  0.123400
0.022774
TCGA-49-6767               0.032674  0.100785  0.012335  0.213498
0.021355
TCGA-AA-A03J               0.006116  0.021166  0.065235  0.159937
0.019312
TCGA-77-7337               0.053161  0.019366  0.010812  0.269010
0.025416

                            CD163_S3   CD19_S4  CD247_S4   CD33_S3
CD37_S4  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.009689  0.003777  0.063614  0.033109
0.015664
TCGA-B5-A3F9               0.017677  0.002194  0.035106  0.071084
0.017639
TCGA-49-6767               0.025298  0.000785  0.083596  0.075230
0.042144
TCGA-AA-A03J               0.004955  0.003470  0.044976  0.031795
0.022887
TCGA-77-7337               0.038462  0.001408  0.039633  0.049115
0.016808

                             CD3D_S4   CD3E_S4   CD3G_S4   CD44_S2
CD48_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.040169  0.023715  0.016702  0.075255
0.014924
TCGA-B5-A3F9               0.040378  0.024578  0.005270  0.008014
0.024867
TCGA-49-6767               0.086689  0.076188  0.065392  0.038649
0.040068
TCGA-AA-A03J               0.047478  0.023950  0.020800  0.012806
0.015225
TCGA-77-7337               0.034323  0.032421  0.047515  0.092642
0.020929

                             CD52_S3   CD53_S3   CD59_S2  CD79A_S4
CD79B_S4  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.018208  0.038376  0.113421  0.012376
```

```
0.009356
TCGA-B5-A3F9              0.016722  0.053371  0.043971  0.001537
0.003067
TCGA-49-6767              0.061403  0.100435  0.047856  0.003519
0.003751
TCGA-AA-A03J              0.026161  0.035758  0.014871  0.005998
0.005639
TCGA-77-7337              0.021409  0.102105  0.107698  0.016165
0.005231

                           CD84_S3   CD86_S3   CD8A_S3   CD8B_S4
CD97_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.003355  0.110400  0.004560  0.002090
0.030530
TCGA-B5-A3F9              0.004642  0.049360  0.026804  0.019961
0.052208
TCGA-49-6767              0.010102  0.144443  0.088363  0.008436
0.073799
TCGA-AA-A03J              0.004176  0.036119  0.037436  0.006757
0.248789
TCGA-77-7337              0.018940  0.188385  0.028269  0.008949
0.052528

                          CDCA4_S5   CDH6_S2   CDK2_S5  CECR1_S3
CELF2_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.239033  0.004352  0.029561  0.022926
0.054613
TCGA-B5-A3F9              0.158576  0.003591  0.031723  0.101669
0.044890
TCGA-49-6767              0.182581  0.000571  0.038141  0.054018
0.036824
TCGA-AA-A03J              0.109814  0.001170  0.028389  0.024335
0.097108
TCGA-77-7337              0.242704  0.125795  0.033506  0.036337
0.041916

                          CENPJ_S5  CENPN_S5  CENPO_S5  CENPW_S5
CEP78_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.077392  0.058463  0.176886  0.036657
0.158284
TCGA-B5-A3F9              0.215607  0.022141  0.065257  0.029258
0.187278
TCGA-49-6767              0.028629  0.047939  0.127073  0.022736
0.143665
```

```
TCGA-AA-A03J                0.066650  0.052004  0.166096  0.060270
0.048390
TCGA-77-7337                0.035145  0.109367  0.380965  0.066628
0.336437

                            CHEK1_S5   CKLF_S5  COL16A1_S2  COL1A2_S2  \
TCGA Participant Barcode
TCGA-WA-A7H4                0.341859  0.055799    0.179046   0.049270
TCGA-B5-A3F9                0.168456  0.060539    0.006053   0.005468
TCGA-49-6767                0.319432  0.279732    0.023565   0.011242
TCGA-AA-A03J                0.107756  0.138498    0.022623   0.016113
TCGA-77-7337                0.422631  0.063020    0.028620   0.055869

                            COL3A1_S2  COL6A1_S2  COL6A3_S2
COL8A1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                 0.053382   0.090780   0.168243   0.020451

TCGA-B5-A3F9                 0.011437   0.041633   0.007959   0.000976

TCGA-49-6767                 0.010848   0.083508   0.050570   0.038859

TCGA-AA-A03J                 0.012296   0.036574   0.051404   0.023293

TCGA-77-7337                 0.080138   0.030022   0.215291   0.087967


                            COPS6_S5   COQ2_S5  CORO1A_S3  CORO1C_S5
COTL1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.073187  0.151087   0.039494   0.153347
0.089469
TCGA-B5-A3F9                0.111060  0.135605   0.031174   0.040879
0.102082
TCGA-49-6767                0.484917  0.148317   0.079423   0.212910
0.331710
TCGA-AA-A03J                0.115420  0.190231   0.176253   0.140236
0.068212
TCGA-77-7337                0.089021  0.137881   0.027044   0.342771
0.110654

                            COX17_S5  CPEB4_S5   CPVL_S3  CSF1R_S3
CTNNA1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.077845  0.066910  0.008639  0.035377
0.332514
TCGA-B5-A3F9                0.119659  0.037612  0.009777  0.031854
```

```
0.324371
TCGA-49-6767              0.130252  0.019654  0.016773  0.048437
0.276330
TCGA-AA-A03J              0.087111  0.024473  0.009468  0.032764
0.336343
TCGA-77-7337              0.053663  0.073843  0.012768  0.056842
0.322352

                          CTNNB1_S2   CTSC_S3  CTSL1_S3   CTSS_S3
CXCL10_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.163481  0.095917  0.023556  0.018769
0.003976
TCGA-B5-A3F9               0.128645  0.058378  0.034386  0.040411
0.006545
TCGA-49-6767               0.273783  0.080481  0.025965  0.236131
0.055138
TCGA-AA-A03J               0.202381  0.042648  0.007856  0.069236
0.006464
TCGA-77-7337               0.248077  0.203987  0.044214  0.046477
0.012050

                          CXCR4_S3   CYBB_S3  CYTIP_S3  DAPK1_S3
DBNDD1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.009499  0.018678  0.030886  0.082398
0.015780
TCGA-B5-A3F9              0.083557  0.018557  0.038931  0.041218
0.066342
TCGA-49-6767              0.019742  0.066786  0.049761  0.013979
0.069333
TCGA-AA-A03J              0.008578  0.013780  0.012738  0.026551
0.027064
TCGA-77-7337              0.038591  0.066092  0.060276  0.027985
0.013411

                          DCBLD2_S5    DCK_S5  DCLRE1B_S5  DDX58_S1
DDX60_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.110837  0.160330    0.287794  0.053464
0.072989
TCGA-B5-A3F9               0.069558  0.103752    0.207635  0.026454
0.059039
TCGA-49-6767               0.854769  0.256074    0.232055  0.031037
0.078268
TCGA-AA-A03J               0.016922  0.102256    0.117280  0.027088
0.072232
```

```
TCGA-77-7337                0.101117  0.293041    0.243294  0.014842
0.074897

                            DHFR_S5  DLEU1_S5  DLEU2_S5  DOCK2_S3
DSP_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.085747  0.097426  0.145469  0.044880
0.184312
TCGA-B5-A3F9                0.165736  0.245137  0.103178  0.021801
0.017743
TCGA-49-6767                0.086690  0.111307  0.224930  0.071234
0.036720
TCGA-AA-A03J                0.152287  0.163563  0.053175  0.032179
0.033738
TCGA-77-7337                0.065244  0.207484  0.215501  0.069890
0.218342

                              DUT_S5   DVL1_S2   DVL3_S2  DYNLT1_S5  \
TCGA Participant Barcode
TCGA-WA-A7H4                0.202159  0.284554  0.069952   0.117494
TCGA-B5-A3F9                0.286724  0.189871  0.051480   0.268827
TCGA-49-6767                0.093573  0.239620  0.055469   0.111976
TCGA-AA-A03J                0.357800  0.412409  0.082258   0.106556
TCGA-77-7337                0.082165  0.101505  0.123503   0.112132

                            EBNA1BP2_S5  EFNA5_S2  EIF2AK1_S5
EIF4EBP1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                  0.254913  0.100997    0.180324
0.045329
TCGA-B5-A3F9                  0.138149  0.012586    0.275381
0.067787
TCGA-49-6767                  0.229355  0.003574    0.234314
0.050975
TCGA-AA-A03J                  0.141679  0.001799    0.280938
0.065817
TCGA-77-7337                  0.123693  0.058037    0.311581
0.164195

                            EIF4G1_S5   EMP2_S5   ENO1_S5  EPHA2_S2
EPHB2_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.224737  0.073334  0.176186  0.100160
0.030938
TCGA-B5-A3F9                0.099652  0.037854  0.106735  0.038517
0.027591
TCGA-49-6767                0.234288  0.092133  0.181019  0.065869
```

```
0.127376
TCGA-AA-A03J                0.181465  0.046118  0.087059  0.199168
0.253530
TCGA-77-7337                0.303863  0.117772  0.565930  0.153488
0.135483

                           EPHB3_S2  EPHB4_S2  ERLIN1_S5  EVI2A_S3
EVI2B_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.032859  0.146180   0.144039  0.018113
0.023936
TCGA-B5-A3F9               0.009309  0.103594   0.204526  0.016738
0.019382
TCGA-49-6767               0.014559  0.752040   0.450168  0.041415
0.057620
TCGA-AA-A03J               0.149952  0.288277   0.213134  0.004571
0.012131
TCGA-77-7337               0.059069  0.105869   0.432919  0.048020
0.049250

                           EXOSC8_S5    EZR_S5     F3_S5  FAM167A_S5  \
TCGA Participant Barcode
TCGA-WA-A7H4                0.161982  0.171173  0.210631    0.010754
TCGA-B5-A3F9                0.184557  0.076877  0.011839    0.002479
TCGA-49-6767                0.227998  0.257706  0.033512    0.008293
TCGA-AA-A03J                0.180274  0.154220  0.012739    0.004658
TCGA-77-7337                0.111053  0.132114  0.036959    0.003482

                           FAM89B_S5  FARSA_S5  FARSB_S5  FBXO41_S5
FCER1G_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.204180  0.102053  0.329130   0.047208
0.018273
TCGA-B5-A3F9                0.169258  0.111014  0.328010   0.007366
0.044371
TCGA-49-6767                0.316784  0.065963  0.376409   0.048723
0.077242
TCGA-AA-A03J                0.160079  0.268263  0.287552   0.107019
0.017333
TCGA-77-7337                0.129489  0.113086  0.279407   0.012326
0.045339

                           FCGR1A_S3  FCGR2A_S3  FCGR2B_S3
FCGR3A_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.006308   0.030941   0.012875   0.013803
```

```
TCGA-B5-A3F9                0.028072   0.025621   0.010577   0.024701

TCGA-49-6767                0.066439   0.067265   0.021345   0.071530

TCGA-AA-A03J                0.007035   0.028258   0.003753   0.012462

TCGA-77-7337                0.019803   0.169831   0.030141   0.057074


                            FGD1_S2   FGL2_S3    FGR_S3   FLI1_S3
FLNC_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.020552  0.006781  0.032155  0.083692
0.068794
TCGA-B5-A3F9               0.088063  0.003095  0.017032  0.046122
0.001224
TCGA-49-6767               0.093016  0.011098  0.067096  0.064801
0.152849
TCGA-AA-A03J               0.025836  0.005562  0.030626  0.035335
0.013318
TCGA-77-7337               0.043341  0.011555  0.039693  0.086868
0.002109

                           FMNL1_S3    FN1_S2  FNBP1_S3   FPR3_S3
FYB_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.101362  0.027169  0.046967  0.056917
0.053084
TCGA-B5-A3F9               0.022260  0.001267  0.065727  0.020283
0.041086
TCGA-49-6767               0.359959  0.028097  0.098693  0.104719
0.066201
TCGA-AA-A03J               0.074379  0.003745  0.039936  0.040684
0.010794
TCGA-77-7337               0.059797  0.051011  0.032962  0.125610
0.390670

                             GGH_S5  GIMAP4_S3  GLRX3_S5  GNG11_S5
GNPTAB_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.027823   0.025146  0.160717  0.010046
0.057354
TCGA-B5-A3F9               0.021042   0.039299  0.149249  0.013832
0.051172
TCGA-49-6767               0.010985   0.071292  0.422999  0.003070
0.026608
TCGA-AA-A03J               0.019023   0.024416  0.174956  0.002216
```

```
0.023469
TCGA-77-7337              0.038501   0.049023  0.343412  0.005633
0.087136

                          GPLD1_S5  GPNMB_S3  GSTCD_S5  H2AFZ_S5
HAS2_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.000408  0.053609  0.109785  0.152008
0.037437
TCGA-B5-A3F9              0.000736  0.012891  0.137458  0.199343
0.014518
TCGA-49-6767              0.000730  0.036157  0.102690  0.123665
0.004785
TCGA-AA-A03J              0.000500  0.003611  0.074926  0.145930
0.010080
TCGA-77-7337              0.000756  0.217466  0.158231  0.179624
0.141935

                          HAUS1_S5    HCK_S3  HCLS1_S3    HDC_S4
HERC5_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.020346  0.047622  0.057652  0.001211
0.025456
TCGA-B5-A3F9              0.067098  0.038908  0.086070  0.000030
0.014123
TCGA-49-6767              0.015883  0.099041  0.085216  0.000035
0.009157
TCGA-AA-A03J              0.021343  0.050230  0.035609  0.000945
0.005109
TCGA-77-7337              0.033436  0.065982  0.227667  0.000368
0.036022

                          HERC6_S1  HLA-DMA_S3  HLA-DRB1_S3  HMGN2_S5
\
TCGA Participant Barcode

TCGA-WA-A7H4              0.075657    0.022540     0.047645  0.136413

TCGA-B5-A3F9              0.053794    0.103293     0.086854  0.244664

TCGA-49-6767              0.022453    0.060781     0.204553  0.142709

TCGA-AA-A03J              0.041770    0.053130     0.069789  0.071459

TCGA-77-7337              0.051997    0.052412     0.073172  0.168506


                          HMHA1_S3   HN1L_S5  HNRNPA2B1_S5  HNRNPR_S5
\
```

```
TCGA Participant Barcode

TCGA-WA-A7H4               0.071628  0.271469      0.475610   0.399541

TCGA-B5-A3F9               0.034324  0.280583      0.548851   0.388752

TCGA-49-6767               0.088358  0.161530      0.251151   0.299662

TCGA-AA-A03J               0.163964  0.157864      0.164280   0.170824

TCGA-77-7337               0.019563  0.358311      0.172482   0.437498


                           HSPB11_S5  HSPG2_S2  HYLS1_S5  ICAM1_S2
ID2_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.218957  0.188063  0.109928  0.034119
0.037530
TCGA-B5-A3F9                0.224201  0.064808  0.062105  0.012642
0.209001
TCGA-49-6767                0.096953  0.078204  0.078556  0.136449
0.032776
TCGA-AA-A03J                0.041110  0.171041  0.021903  0.022959
0.106346
TCGA-77-7337                0.073749  0.206996  0.151481  0.066420
0.037849

                             ID3_S5  IFI16_S3  IFI27_S1  IFI30_S3
IFI44_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.030801  0.174878  0.010585  0.038855
0.092889
TCGA-B5-A3F9               0.159889  0.044124  0.030673  0.047912
0.042425
TCGA-49-6767               0.007644  0.064804  0.033371  0.092049
0.082192
TCGA-AA-A03J               0.055888  0.011214  0.056305  0.077159
0.012154
TCGA-77-7337               0.035191  0.190937  0.023718  0.078464
0.022584

                           IFI44L_S1   IFI6_S1  IFIH1_S1  IFIT1_S1
IFIT2_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.008884  0.016181  0.052465  0.012104
0.013984
TCGA-B5-A3F9                0.014465  0.041187  0.044148  0.009991
```

```
0.020951
TCGA-49-6767                0.066570  0.005650  0.088845  0.007317
0.059526
TCGA-AA-A03J                0.010863  0.031855  0.034967  0.009420
0.017772
TCGA-77-7337                0.002674  0.007654  0.061444  0.004512
0.012758

                           IFIT3_S1  IFRD2_S5  IGF2R_S2  IGFBP2_S2
IGFBP3_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.032078  0.174951  0.417760   0.008048
0.011803
TCGA-B5-A3F9               0.022286  0.148126  0.090358   0.046128
0.017299
TCGA-49-6767               0.069046  0.249150  0.292550   0.000552
0.008982
TCGA-AA-A03J               0.046223  0.469290  0.304512   0.004864
0.010374
TCGA-77-7337               0.028578  0.124414  0.286602   0.029685
0.117323

                           IGFBP4_S2  IGFBP5_S2    IGJ_S4  IKZF1_S3
IL10RA_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.076637   0.006510  0.077137  0.030634
0.043147
TCGA-B5-A3F9                0.259275   0.002993  0.003288  0.016425
0.033162
TCGA-49-6767                0.053771   0.007041  0.002262  0.039746
0.075596
TCGA-AA-A03J                0.062825   0.004197  0.009832  0.014251
0.045385
TCGA-77-7337                0.099363   0.027515  0.038158  0.027573
0.042415

                            IL18_S3   IL7R_S3   IMP4_S5   IPO4_S5
ISG15_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.083119  0.079098  0.139524  0.207161
0.014778
TCGA-B5-A3F9               0.005650  0.010949  0.130361  0.212707
0.037194
TCGA-49-6767               0.116826  0.013034  0.160946  0.200339
0.022078
TCGA-AA-A03J               0.025072  0.010332  0.283705  0.306210
0.056967
```

```
TCGA-77-7337              0.051308  0.095548  0.150310  0.249193
0.008037

                          ITGA3_S2  ITGA4_S2  ITGA5_S2  ITGA6_S5
ITGB2_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.276907  0.053032  0.086581  0.373463
0.021288
TCGA-B5-A3F9              0.015699  0.018388  0.012094  0.032617
0.026752
TCGA-49-6767              0.270769  0.034460  0.049861  0.042737
0.100076
TCGA-AA-A03J              0.068338  0.023130  0.024807  0.061921
0.036380
TCGA-77-7337              0.105767  0.067336  0.151642  0.192889
0.038590

                          ITGB3_S2  ITGB5_S2  ITGB8_S2   JUNB_S5
JUP_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.017033  0.202443  0.024935  0.028278
0.081848
TCGA-B5-A3F9              0.001087  0.097532  0.012690  0.029653
0.020165
TCGA-49-6767              0.007168  0.096397  0.027774  0.011750
0.048325
TCGA-AA-A03J              0.003954  0.107455  0.003000  0.051715
0.076755
TCGA-77-7337              0.014997  0.112845  0.013661  0.027507
0.132390

                          KIAA0090_S5   KRR1_S5  LAIR1_S3  LAMA4_S2
LAMB1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                 0.442496  0.266041  0.031268  0.063886
0.134072
TCGA-B5-A3F9                 0.161326  0.163491  0.054181  0.010380
0.166282
TCGA-49-6767                 0.171628  0.144502  0.087119  0.012079
0.102749
TCGA-AA-A03J                 0.081225  0.086434  0.035759  0.020305
0.133157
TCGA-77-7337                 0.672297  0.374044  0.050773  0.065265
0.125134

                          LAPTM5_S3    LCK_S4   LCP1_S3   LCP2_S3
LHFPL2_S3  \
```

```
TCGA Participant Barcode

TCGA-WA-A7H4                0.029875  0.023022  0.057339  0.062984
0.132281
TCGA-B5-A3F9                0.057293  0.023920  0.078932  0.038561
0.092902
TCGA-49-6767                0.151932  0.060208  0.098733  0.108373
0.124139
TCGA-AA-A03J                0.037515  0.026947  0.024536  0.035882
0.090632
TCGA-77-7337                0.083237  0.029194  0.103698  0.121301
0.178464

                           LILRB4_S3  LMNB2_S5  LOXL2_S5   LRMP_S3
LRP1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.008157  0.250072  0.030669  0.004127
0.021220
TCGA-B5-A3F9                0.020459  0.144390  0.008254  0.002365
0.009356
TCGA-49-6767                0.039352  0.172712  0.023336  0.001690
0.010615
TCGA-AA-A03J                0.016934  0.366108  0.024214  0.001860
0.042072
TCGA-77-7337                0.027372  0.233035  0.045484  0.025887
0.048902

                           LRRC17_S2  LRRC40_S5   LSM3_S5   LSM4_S5
LST1_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.060840   0.178373  0.109577  0.100458
0.045123
TCGA-B5-A3F9                0.025620   0.193141  0.163247  0.198695
0.031602
TCGA-49-6767                0.011497   0.233212  0.156403  0.109571
0.061070
TCGA-AA-A03J                0.003852   0.017883  0.069174  0.172504
0.037040
TCGA-77-7337                0.016909   0.136157  0.111263  0.089692
0.026585

                             LTB_S4   LY86_S3   LYAR_S5    LYN_S3
LYZ_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.042660  0.009661  0.210344  0.114421
0.031317
TCGA-B5-A3F9               0.018470  0.030685  0.104753  0.099920
```

```
0.001993
TCGA-49-6767               0.170264  0.040346  0.296658  0.346249
0.016032
TCGA-AA-A03J               0.024858  0.014650  0.189843  0.072093
0.007370
TCGA-77-7337               0.002604  0.019700  0.242711  0.239733
0.016140

                           MAGOHB_S5  MAP3K8_S5  MAPRE1_S5  MARCKSL1_S2
\
TCGA Participant Barcode

TCGA-WA-A7H4                0.075697   0.035770   0.145664     0.040757

TCGA-B5-A3F9                0.107498   0.062074   0.090899     0.385772

TCGA-49-6767                0.104949   0.027099   0.134828     0.170430

TCGA-AA-A03J                0.048341   0.018603   0.160546     0.126210

TCGA-77-7337                0.135993   0.038159   0.314931     0.027627


                           MARVELD2_S5   MCM3_S5   MCM7_S5  MCTS1_S5
MERTK_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                  0.113350  0.216977  0.074077  0.221840
0.010786
TCGA-B5-A3F9                  0.190823  0.152350  0.075562  0.121013
0.046835
TCGA-49-6767                  0.179898  0.115889  0.331314  0.217010
0.033905
TCGA-AA-A03J                  0.227938  0.175003  0.194599  0.101290
0.007017
TCGA-77-7337                  0.180933  0.076036  0.081903  0.082588
0.035576

                             MET_S5  MFSD11_S5   MKKS_S5  MLF1IP_S5
MMP11_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.028999   0.368101  0.106318   0.169909
0.080923
TCGA-B5-A3F9               0.013101   0.097814  0.118294   0.188447
0.146142
TCGA-49-6767               0.322232   0.274009  0.054881   0.168724
0.068895
TCGA-AA-A03J               0.019980   0.164033  0.035276   0.074050
0.062199
```

```
TCGA-77-7337              0.032555   0.316945  0.063507   0.104407
0.017603

                           MMP1_S2  MMP14_S2  MMP17_S2  MMP19_S2
MMP2_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.194250  0.138782  0.028384  0.021184
0.076244
TCGA-B5-A3F9              0.000080  0.054637  0.005161  0.028822
0.067427
TCGA-49-6767              0.021952  0.216173  0.012879  0.011475
0.030269
TCGA-AA-A03J              0.009151  0.054713  0.003805  0.010316
0.044198
TCGA-77-7337              0.134127  0.087733  0.009456  0.012750
0.092904

                           MMP3_S2  MNAT1_S5   MNDA_S3   MPP1_S3
MRPL12_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.100101  0.297820  0.034854  0.028984
0.053989
TCGA-B5-A3F9              0.001807  0.472321  0.023676  0.023693
0.071177
TCGA-49-6767              0.000234  0.363728  0.134574  0.113771
0.053678
TCGA-AA-A03J              0.015790  0.067786  0.015200  0.119624
0.128341
TCGA-77-7337              0.027166  0.351699  0.055352  0.033229
0.046913

                          MRPL37_S5  MRPS16_S5  MRPS28_S5  MRTO4_S5
MS4A1_S4  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.215832   0.081801   0.152298  0.296662
0.001867
TCGA-B5-A3F9               0.149825   0.159533   0.102605  0.261444
0.000091
TCGA-49-6767               0.175349   0.146096   0.056819  0.206756
0.000514
TCGA-AA-A03J               0.340748   0.150107   0.026642  0.162677
0.000202
TCGA-77-7337               0.150464   0.227553   0.115047  0.292974
0.000489

                            MSN_S3   MT1F_S5    MT3_S5   MTA1_S2
MTHFD1_S5  \
```

```
TCGA Participant Barcode

TCGA-WA-A7H4               0.225881  0.001859  0.000017  0.108317
0.053998
TCGA-B5-A3F9               0.102073  0.007738  0.000714  0.205529
0.049750
TCGA-49-6767               0.523028  0.001701  0.000002  0.180721
0.111187
TCGA-AA-A03J               0.046191  0.005101  0.000447  0.161050
0.065647
TCGA-77-7337               0.342708  0.001910  0.000003  0.101209
0.098276

                             MX1_S1  MYBL1_S5  MYBL2_S5  MYCBP_S5
MYL6_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.023341  0.044450  0.036125  0.094650
0.092348
TCGA-B5-A3F9               0.022720  0.011306  0.032607  0.157460
0.113473
TCGA-49-6767               0.022676  0.080992  0.031808  0.093955
0.110403
TCGA-AA-A03J               0.048149  0.005645  0.130239  0.080209
0.084048
TCGA-77-7337               0.018135  0.014082  0.103269  0.062224
0.073702

                           MYO1F_S3  NCEH1_S5   NCF2_S3  NCKAP1L_S3
NCLN_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.060112  0.033299  0.088336    0.041589
0.124847
TCGA-B5-A3F9               0.070195  0.056313  0.018637    0.033273
0.134272
TCGA-49-6767               0.105832  0.160207  0.182093    0.078590
0.109027
TCGA-AA-A03J               0.102991  0.036514  0.035316    0.045108
0.390046
TCGA-77-7337               0.046165  0.096981  0.250913    0.061815
0.095514

                            NEO1_S2   NID1_S2    NLN_S5   NME1_S5
NOP16_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.103046  0.011977  0.239173  0.064625
0.164285
TCGA-B5-A3F9               0.214586  0.011057  0.232128  0.127969
```

```
0.271940
TCGA-49-6767               0.077230  0.008704  0.331265  0.099628
0.135435
TCGA-AA-A03J               0.307042  0.007268  0.205279  0.084575
0.345630
TCGA-77-7337               0.075113  0.034183  0.281952  0.074518
0.102710

                           NOTCH2_S2   NPC2_S3    NPL_S3  NRIP3_S5
NUDT1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.111047  0.015887  0.049165  0.003531
0.080380
TCGA-B5-A3F9                0.120854  0.047381  0.029214  0.000597
0.057976
TCGA-49-6767                0.145194  0.056083  0.026651  0.004892
0.050619
TCGA-AA-A03J                0.028016  0.009288  0.019027  0.000524
0.072168
TCGA-77-7337                0.157209  0.052034  0.165051  0.009000
0.089115

                           NUDT15_S5  NUP107_S5  NUP35_S5  NUP85_S5
NUP93_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.158879   0.083663  0.173388  0.174390
0.121559
TCGA-B5-A3F9                0.124530   0.071152  0.272161  0.137736
0.136196
TCGA-49-6767                0.164204   0.071165  0.173997  0.171342
0.165973
TCGA-AA-A03J                0.106287   0.048762  0.081494  0.186738
0.159651
TCGA-77-7337                0.159017   0.091544  0.314542  0.138004
0.163207

                           NUPL1_S5  NUTF2_S5   OAS1_S1   OAS2_S1
OAS3_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.210659  0.143198  0.035967  0.045707
0.045250
TCGA-B5-A3F9               0.120356  0.307269  0.012002  0.013871
0.020847
TCGA-49-6767               0.183313  0.285336  0.075691  0.046578
0.067696
TCGA-AA-A03J               0.194111  0.328946  0.121539  0.034826
0.094669
```

```
TCGA-77-7337              0.169962  0.194058  0.030596  0.056352
0.051430

                           OASL_S1  OSBPL3_S3  PA2G4_S5  PAICS_S5
PAK1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.018917   0.153589  0.079970  0.166344
0.072260
TCGA-B5-A3F9              0.013454   0.058412  0.160077  0.224232
0.117644
TCGA-49-6767              0.033349   0.112185  0.218809  0.104921
0.116603
TCGA-AA-A03J              0.063285   0.081699  0.217128  0.162434
0.082411
TCGA-77-7337              0.009631   0.063854  0.183514  0.265156
0.127628

                           PAK2_S2  PDAP1_S5  PDIA4_S5  PDLIM7_S5
PFKP_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.241153  0.168676  0.135598   0.250676
0.049983
TCGA-B5-A3F9              0.126378  0.210840  0.335843   0.075176
0.032116
TCGA-49-6767              0.351109  0.800304  0.456125   0.106854
0.132821
TCGA-AA-A03J              0.083170  0.215826  0.174828   0.117111
0.132418
TCGA-77-7337              0.646806  0.126817  0.178679   0.053024
0.158858

                           PFN1_S5   PGM2_S5  PHF19_S5  PIK3CG_S3
PITPNC1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.255600  0.210470  0.191977   0.024237
0.076846
TCGA-B5-A3F9              0.191584  0.054117  0.053123   0.002969
0.131413
TCGA-49-6767              0.170042  0.119041  0.348527   0.017677
0.040962
TCGA-AA-A03J              0.303984  0.054715  0.385093   0.008843
0.051307
TCGA-77-7337              0.110226  0.268565  0.115697   0.037040
0.103141

                           PLAT_S2  PLAUR_S5  PLCG2_S3   PLEK_S3
PLG_S5  \
```

```
TCGA Participant Barcode

TCGA-WA-A7H4               0.001419  0.039017  0.006507  0.023230
0.000005
TCGA-B5-A3F9               0.001808  0.030971  0.004564  0.030490
0.000152
TCGA-49-6767               0.009950  0.122240  0.003650  0.070551
0.000005
TCGA-AA-A03J               0.002175  0.120793  0.001889  0.016255
0.000029
TCGA-77-7337               0.011542  0.076734  0.009946  0.057403
0.000005

                            PLK4_S5  PLOD2_S5    PNN_S5   PNO1_S5
POLE2_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.272897  0.015910  0.147280  0.161602
0.136847
TCGA-B5-A3F9               0.139299  0.019097  0.197264  0.096395
0.096535
TCGA-49-6767               0.198844  0.011098  0.099750  0.184627
0.152285
TCGA-AA-A03J               0.082470  0.003602  0.090495  0.163267
0.073111
TCGA-77-7337               0.208554  0.063720  0.111271  0.379271
0.182561

                           POLR3K_S5   PPIH_S5  PSMA7_S5  PSMC3_S5
PSMD12_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.089785  0.144393  0.125138  0.167270
0.139899
TCGA-B5-A3F9                0.137554  0.232093  0.073134  0.156777
0.062331
TCGA-49-6767                0.077284  0.175007  0.135764  0.108752
0.113334
TCGA-AA-A03J                0.214049  0.200183  0.410415  0.104846
0.035758
TCGA-77-7337                0.186561  0.084329  0.185575  0.178016
0.241631

                           PSMD14_S5  PSMD2_S5  PTPLB_S5  PTPRC_S3
PXN_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.268905  0.157737  0.084729  0.024735
0.135201
TCGA-B5-A3F9                0.238535  0.063664  0.011875  0.016258
```

```
0.030791
TCGA-49-6767                0.216624  0.125575  0.066794  0.041973
0.152908
TCGA-AA-A03J                0.154648  0.094678  0.020744  0.009985
0.161869
TCGA-77-7337                0.551728  0.376199  0.166360  0.049721
0.091718

                           RAB3B_S5   RAC1_S2  RASSF2_S3  RASSF4_S3
RBM14_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.001371  0.153141   0.006755   0.025563
0.282242
TCGA-B5-A3F9               0.000577  0.167645   0.012288   0.028804
0.387320
TCGA-49-6767               0.005981  0.161858   0.002001   0.053798
0.278413
TCGA-AA-A03J               0.001737  0.194394   0.001996   0.026893
0.635406
TCGA-77-7337               0.003655  0.343638   0.004547   0.006769
0.382809

                            RFC3_S5   RGS8_S5   RHOA_S2   RHOB_S2
RHOC_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.095780  0.000000  0.254724  0.040654
0.129998
TCGA-B5-A3F9               0.117169  0.000000  0.276097  0.170179
0.119639
TCGA-49-6767               0.106650  0.000000  0.417517  0.012437
0.230437
TCGA-AA-A03J               0.145781  0.014786  0.276152  0.050956
0.191826
TCGA-77-7337               0.145882  0.002612  0.318228  0.039787
0.131475

                            RHOG_S2   RHOQ_S2  RMND5B_S5  RNASE6_S3
RND3_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.174512  0.107443   0.097324   0.034370
0.133599
TCGA-B5-A3F9               0.117589  0.108968   0.181132   0.029793
0.149408
TCGA-49-6767               0.159244  0.082045   0.153058   0.042139
0.097058
TCGA-AA-A03J               0.243694  0.011838   0.278400   0.034181
0.030692
```

```
TCGA-77-7337              0.115530  0.086344   0.073182   0.027792
0.374773

                          RNF138_S5  RNF41_S5   RPN1_S5  RPP40_S5
RSAD2_S1  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.087150  0.089728  0.169853  0.203581
0.011339
TCGA-B5-A3F9               0.197635  0.124705  0.339800  0.126313
0.006195
TCGA-49-6767               0.104717  0.115594  0.218984  0.175488
0.021158
TCGA-AA-A03J               0.063504  0.164837  0.201962  0.157057
0.043230
TCGA-77-7337               0.152272  0.178470  0.492370  0.242419
0.020195

                           RTP4_S1  RUNX3_S3  RUVBL1_S5  SAMD9_S1
SAMHD1_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.022262  0.040771   0.176109  0.080709
0.052966
TCGA-B5-A3F9              0.047336  0.011941   0.304404  0.007888
0.130442
TCGA-49-6767              0.032420  0.012582   0.228971  0.010268
0.081837
TCGA-AA-A03J              0.030058  0.072303   0.205084  0.012855
0.067619
TCGA-77-7337              0.063709  0.064088   0.423121  0.048694
0.115436

                          SAMSN1_S3  SAR1A_S5  SAR1B_S5   SDC1_S5
SELL_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.047820  0.206790  0.100884  0.098025
0.004675
TCGA-B5-A3F9               0.053131  0.188092  0.083722  0.013992
0.004051
TCGA-49-6767               0.083075  0.158464  0.069516  0.060385
0.005374
TCGA-AA-A03J               0.031994  0.049852  0.055348  0.040424
0.001718
TCGA-77-7337               0.079174  0.354664  0.111418  0.131077
0.004864

                          SELPLG_S3  SEMA3F_S2  SERPINE1_S2  SH2B3_S3
\
```

```
TCGA Participant Barcode

TCGA-WA-A7H4                0.037912   0.107456     0.047018  0.071774

TCGA-B5-A3F9                0.029117   0.075353     0.000925  0.022738

TCGA-49-6767                0.081667   0.044402     0.011848  0.141412

TCGA-AA-A03J                0.058516   0.041377     0.001889  0.051041

TCGA-77-7337                0.052901   0.132073     0.053159  0.074056


                           SH3BP5L_S5   SKA1_S5   SKA2_S5
SLC16A1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                 0.397680  0.023298  0.150416    0.187154

TCGA-B5-A3F9                 0.161268  0.027082  0.163312    0.101799

TCGA-49-6767                 0.147696  0.032137  0.151942    0.170208

TCGA-AA-A03J                 0.289520  0.039170  0.019166    0.068127

TCGA-77-7337                 0.161067  0.065265  0.160582    0.392069


                           SLC1A3_S3  SLC25A40_S5  SLC25A5_S5
SLC7A7_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.004853     0.113213    0.066678
0.008365
TCGA-B5-A3F9                0.001806     0.050642    0.025590
0.008739
TCGA-49-6767                0.001690     0.296410    0.051780
0.030765
TCGA-AA-A03J                0.000242     0.103375    0.114527
0.017580
TCGA-77-7337                0.018085     0.194974    0.062755
0.017081

                            SMC2_S5    SMO_S2    SMS_S5  SMURF2_S5
SNRPA1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.111563  0.061107  0.010822   0.117046
0.098420
TCGA-B5-A3F9               0.059632  0.223859  0.042159   0.080097
0.150227
```

```
TCGA-49-6767              0.122536  0.120447  0.044985   0.469200
0.155771
TCGA-AA-A03J              0.029403  0.041830  0.026457   0.055882
0.132591
TCGA-77-7337              0.305418  0.065579  0.028911   0.139213
0.228911

                          SNRPA_S5  SNRPB_S5  SNRPC_S5  SNRPD1_S5
SNRPE_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.129965  0.110974  0.189107   0.082965
0.033202
TCGA-B5-A3F9              0.357502  0.147041  0.194742   0.211044
0.106333
TCGA-49-6767              0.129030  0.153509  0.222232   0.130903
0.081369
TCGA-AA-A03J              0.353250  0.135313  0.214041   0.078309
0.063577
TCGA-77-7337              0.135565  0.076978  0.224079   0.174427
0.061940

                          SNX17_S5  SP140_S3  SPAG17_S5  SPARC_S2
SRM_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.217200  0.025957   0.016643  0.045811
0.102953
TCGA-B5-A3F9              0.230067  0.012544   0.115356  0.010303
0.081692
TCGA-49-6767              0.169501  0.047222   0.000769  0.012549
0.132022
TCGA-AA-A03J              0.174395  0.009958   0.000777  0.012536
0.129299
TCGA-77-7337              0.178775  0.015842   0.027190  0.056416
0.118792

                          STAT1_S1  STK17A_S5  STRA13_S5    SYK_S3
TAGLN_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.044481   0.241405   0.076512  0.059110
0.038161
TCGA-B5-A3F9              0.055906   0.059119   0.099121  0.168666
0.003984
TCGA-49-6767              0.118732   0.286971   0.054861  0.053159
0.007983
TCGA-AA-A03J              0.043672   0.024340   0.104535  0.138534
0.025773
TCGA-77-7337              0.099685   0.158866   0.026717  0.132907
```

```
0.027327

                            TAP1_S1  TBXAS1_S3  TCEB1_S5  TCF7L2_S2
THBS1_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.078096   0.035709  0.130174   0.099239
0.178477
TCGA-B5-A3F9               0.086456   0.087330  0.103268   0.162975
0.012453
TCGA-49-6767               0.153990   0.039003  0.143786   0.199046
0.012716
TCGA-AA-A03J               0.231884   0.302471  0.052949   0.228019
0.031254
TCGA-77-7337               0.064915   0.038927  0.159103   0.164192
0.093766

                            THBS2_S2  TIMP1_S2   TLR2_S3  TMEM130_S5
TNC_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.065712  0.028905  0.148018    0.000175
0.422519
TCGA-B5-A3F9               0.001430  0.027609  0.075820    0.000108
0.015413
TCGA-49-6767               0.028705  0.038713  0.048070    0.000505
0.152232
TCGA-AA-A03J               0.018995  0.047791  0.023434    0.000256
0.017496
TCGA-77-7337               0.158186  0.041740  0.283652    0.000990
0.302027

                            TNFAIP3_S3  TNFRSF12A_S5  TNFRSF1A_S2
TNFRSF1B_S3  \
TCGA Participant Barcode

TCGA-WA-A7H4                  0.080005      0.168056     0.252215
0.044209
TCGA-B5-A3F9                  0.100122      0.044841     0.219536
0.089376
TCGA-49-6767                  0.054746      0.121907     0.350215
0.129051
TCGA-AA-A03J                  0.058073      0.108840     0.273136
0.162439
TCGA-77-7337                  0.114188      0.049923     0.433645
0.040264

                            TNFSF13B_S3  TOMM40_S5   TPI1_S5
TPM1_S5  \
TCGA Participant Barcode
```

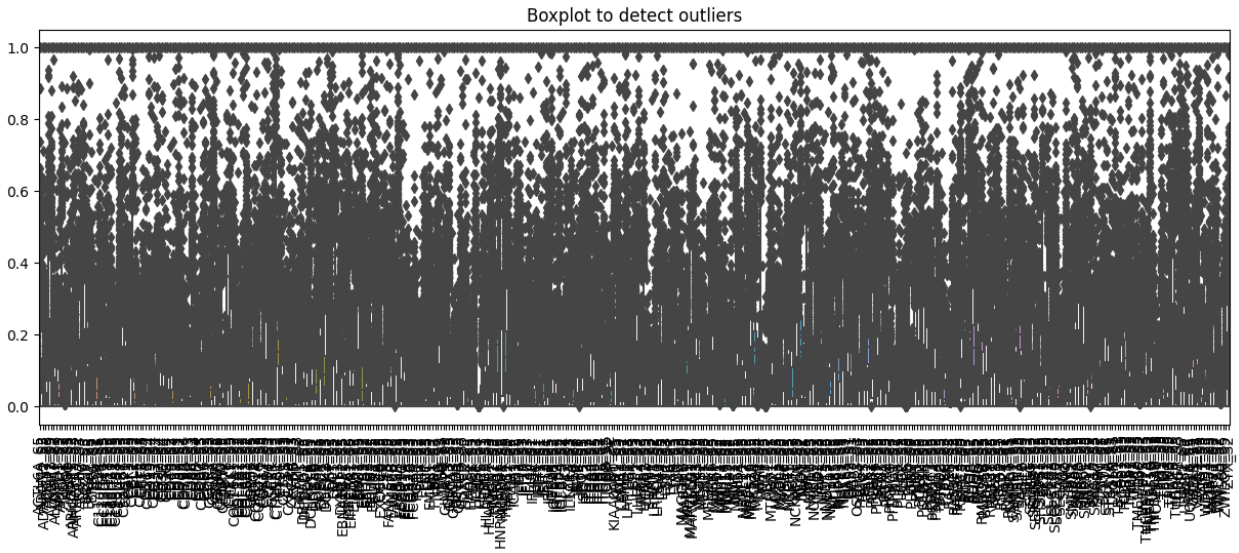
```
TCGA-WA-A7H4                  0.007273   0.030447  0.055560  0.085022

TCGA-B5-A3F9                  0.005195   0.039365  0.121625  0.019840

TCGA-49-6767                  0.027343   0.044709  0.170366  0.031936

TCGA-AA-A03J                  0.007791   0.085939  0.142167  0.048872

TCGA-77-7337                  0.019876   0.041386  0.247876  0.027684


                             TPM2_S5   TPM3_S5  TPRKB_S5  TRA2B_S5
TUBA4A_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.076884  0.248836  0.224479  0.370910
0.153193
TCGA-B5-A3F9              0.010281  0.191961  0.432516  0.307541
0.002220
TCGA-49-6767              0.029378  0.219096  0.172608  0.150605
0.040278
TCGA-AA-A03J              0.013824  0.219590  0.089482  0.170531
0.110547
TCGA-77-7337              0.008869  0.276453  0.239914  0.384092
0.063399

                          TUBG1_S5   UAP1_S5  UBE2J1_S5   UMPS_S5
UQCR10_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.249644  0.107880   0.182237  0.252513
0.037382
TCGA-B5-A3F9              0.227805  0.080434   0.249096  0.145911
0.072831
TCGA-49-6767              0.238908  0.078863   0.267850  0.215546
0.063526
TCGA-AA-A03J              0.293324  0.038459   0.137453  0.194995
0.042015
TCGA-77-7337              0.435219  0.129999   0.277074  0.312816
0.022331

                          USPL1_S5   VCAN_S2  VDAC1_S5  VSIG4_S3
VTA1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4              0.201070  0.072346  0.195607  0.003054
0.269151
TCGA-B5-A3F9              0.190156  0.108592  0.229437  0.013031
0.175206
```

```
TCGA-49-6767                0.178736  0.040453  0.227530  0.023227
0.184954
TCGA-AA-A03J                0.102779  0.018265  0.353334  0.005992
0.102899
TCGA-77-7337                0.054357  0.206705  0.262335  0.017925
0.348754

                            WDHD1_S5  WDR54_S5  WDR77_S5  WIPF1_S3
WNT2B_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.079282  0.072920  0.102072  0.089859
0.010692
TCGA-B5-A3F9                0.034451  0.188849  0.633403  0.067037
0.006442
TCGA-49-6767                0.094871  0.037184  0.124078  0.091530
0.002030
TCGA-AA-A03J                0.018202  0.034451  0.221328  0.026653
0.012035
TCGA-77-7337                0.114240  0.052860  0.092856  0.126317
0.094993

                            WNT8B_S2   WSB2_S5  ZWILCH_S5    ZYX_S2  \
TCGA Participant Barcode
TCGA-WA-A7H4                0.009518  0.124660   0.154684  0.191966
TCGA-B5-A3F9                0.002612  0.089291   0.144194  0.126821
TCGA-49-6767                0.003011  0.142595   0.099884  0.293115
TCGA-AA-A03J                0.003367  0.134369   0.079119  0.256544
TCGA-77-7337                0.000520  0.167126   0.160216  0.132789

                           Immune Subtype
TCGA Participant Barcode
TCGA-WA-A7H4                           C1
TCGA-B5-A3F9                           C1
TCGA-49-6767                           C1
TCGA-AA-A03J                           C1
TCGA-77-7337                           C1

df_final_normalized.to_csv("df_final_normalized.csv")
```

Outliers Handling

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import NMF
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
```

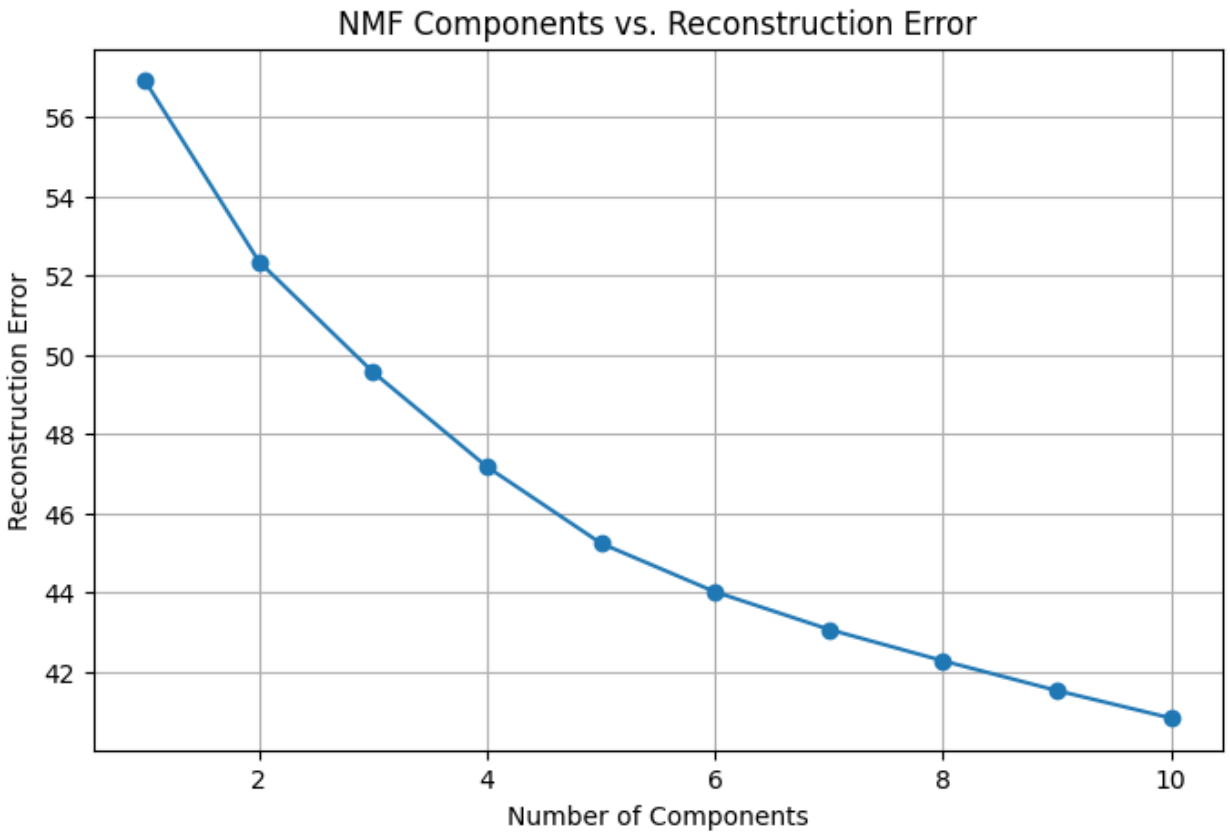
```python
# Visualize Outliers
def visualize_outliers(df):
    plt.figure(figsize=(15, 5))
    sns.boxplot(data=df)
    plt.xticks(rotation=90)
    plt.title("Boxplot to detect outliers")
    plt.show()

# Outlier Treatment using IQR Method
# Outlier Treatment using IQR Method (Handle Non-Numeric Columns)
def remove_outliers_iqr(df):
    numeric_df = df.select_dtypes(include=[np.number])  # Keep only
numeric columns
    Q1 = numeric_df.quantile(0.25)
    Q3 = numeric_df.quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    numeric_df = numeric_df.clip(lower=lower_bound, upper=upper_bound,
axis=1)

    # Merge back non-numeric columns
    df[numeric_df.columns] = numeric_df
    return df


# Load Data (Assuming df_final_normalized is pre-loaded)
# df_final_normalized = pd.read_csv("your_data.csv")  # Uncomment if
loading from file

# Visualize and Remove Outliers
visualize_outliers(df_final_normalized)
df_final_normalized = remove_outliers_iqr(df_final_normalized)
```

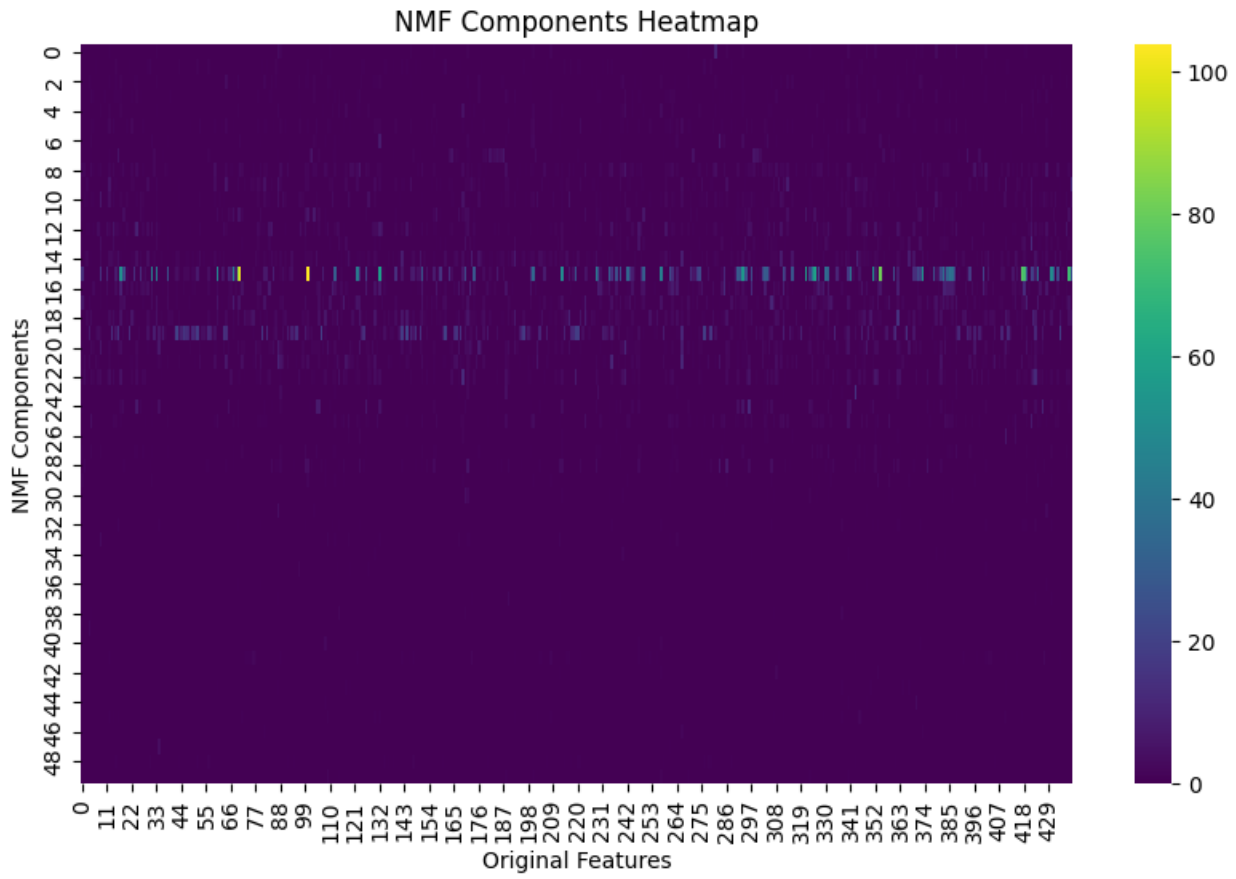


Feature Reduction- Performing NMF

```
# Section 3: Perform NMF (excluding output variable)
X = df_final_normalized.drop(columns=['Immune Subtype'])  # Exclude
target variable
nmf = NMF(n_components=50, random_state=42)  # Choose appropriate
components
X_nmf = nmf.fit_transform(X)

# Explained Reconstruction Error Plot
reconstruction_errors = []
components_range = range(1, 11)
for n in components_range:
    nmf_test = NMF(n_components=n, random_state=42)
    nmf_test.fit(X)
    reconstruction_errors.append(nmf_test.reconstruction_err_)

plt.figure(figsize=(8, 5))
plt.plot(components_range, reconstruction_errors, marker='o')
plt.xlabel('Number of Components')
plt.ylabel('Reconstruction Error')
plt.title('NMF Components vs. Reconstruction Error')
plt.grid()
plt.show()

# Visualizing NMF Components
plt.figure(figsize=(10, 6))
sns.heatmap(nmf.components_, cmap='viridis', annot=False)
plt.xlabel("Original Features")
plt.ylabel("NMF Components")
plt.title("NMF Components Heatmap")
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/
_nmf.py:1665: ConvergenceWarning: Maximum number of iterations 200
reached. Increase it to improve convergence.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/_nmf.py:
1665: ConvergenceWarning: Maximum number of iterations 200 reached.
Increase it to improve convergence.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/_nmf.py:
1665: ConvergenceWarning: Maximum number of iterations 200 reached.
Increase it to improve convergence.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/_nmf.py:
1665: ConvergenceWarning: Maximum number of iterations 200 reached.
Increase it to improve convergence.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/_nmf.py:
1665: ConvergenceWarning: Maximum number of iterations 200 reached.
Increase it to improve convergence.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/_nmf.py:
1665: ConvergenceWarning: Maximum number of iterations 200 reached.
Increase it to improve convergence.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/_nmf.py:
1665: ConvergenceWarning: Maximum number of iterations 200 reached.
Increase it to improve convergence.
  warnings.warn(
```


NMF Components vs. Reconstruction Error
Reconstruction Error
56
54
52
50
48
46
44
42
2
4
6
8
10
Number of Components



-knn,svc, xgb,nn different hyperparamters, cross val score, f1, recall -pros/cons for everything

- upto 6 minutes -pdf file will have comments as well its just a py version of it

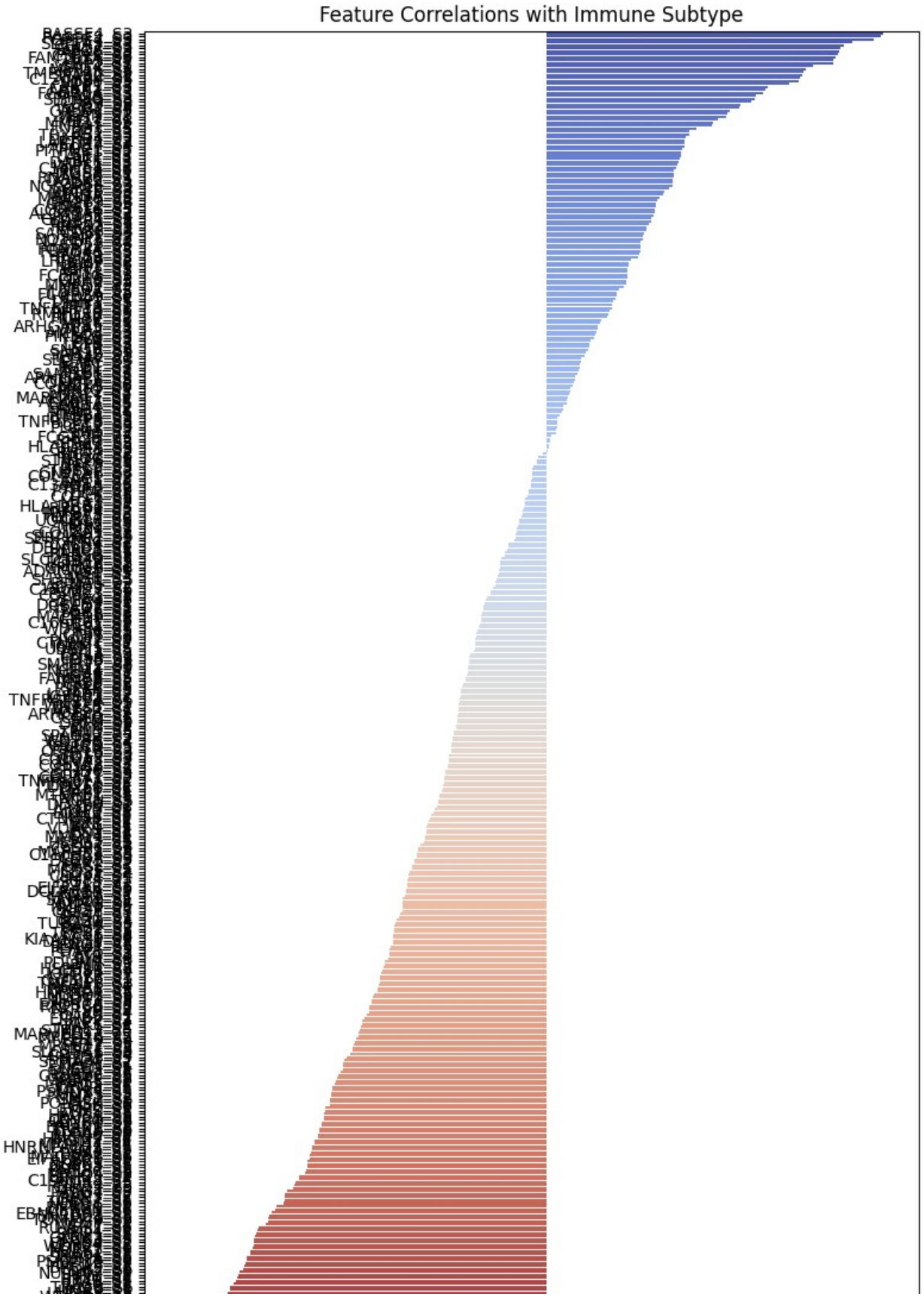
Feature Correlation Plots

```
# Encode target variable
le = LabelEncoder()
y = le.fit_transform(df_final_normalized['Immune Subtype'])  # Encode
categorical labels

# Convert y to DataFrame
y_df = pd.DataFrame(y, columns=['Immune Subtype'])

# Drop target from X
X = df_final_normalized.drop(columns=['Immune Subtype'])

# Reset indices and concatenate
X = X.reset_index(drop=True)
y_df = y_df.reset_index(drop=True)
```

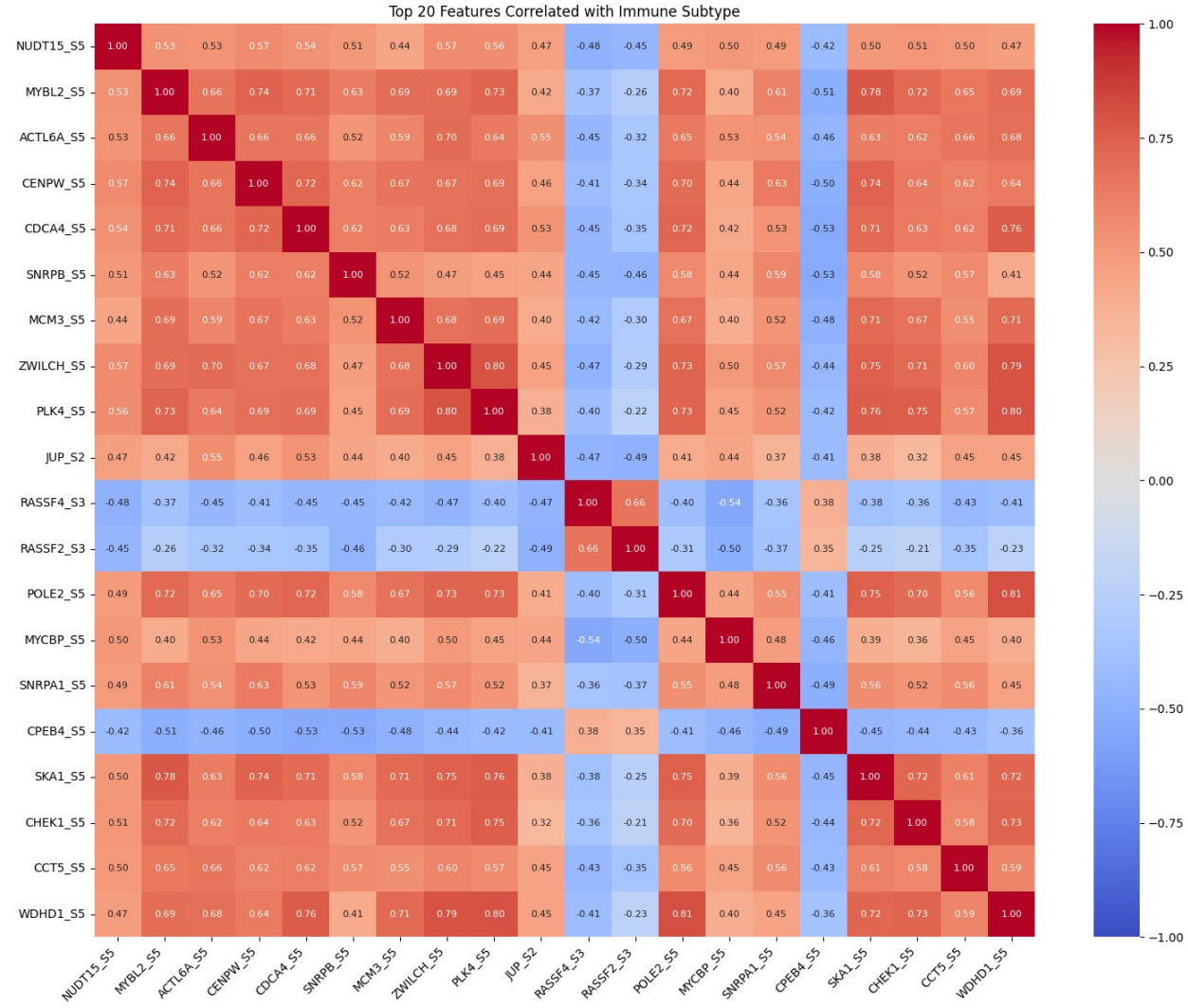
```
df = pd.concat([X, y_df], axis=1)

# Compute correlation matrix
corr_matrix = df.corr()

# 1. First look specifically at Immune Subtype correlations
print("Top correlations with Immune Subtype:")
print(corr_matrix["Immune
Subtype"].sort_values(ascending=False).head(10))
print("\nBottom correlations with Immune Subtype:")
print(corr_matrix["Immune Subtype"].sort_values().head(10))

# 2. Better visualization focused on Immune Subtype
plt.figure(figsize=(8, 12))
immune_corrs = corr_matrix["Immune
Subtype"].sort_values(ascending=False)
immune_corrs = immune_corrs[immune_corrs.index != "Immune Subtype"]  #
Remove self-correlation
sns.barplot(x=immune_corrs.values, y=immune_corrs.index,
palette="coolwarm")
plt.title("Feature Correlations with Immune Subtype")
plt.xlabel("Correlation Coefficient")
plt.tight_layout()
plt.show()

# 3. Filtered heatmap version (adjusted threshold)
plt.figure(figsize=(15, 12))
threshold = 0.3  # More reasonable biological threshold

# Get top N features correlated with Immune Subtype
top_features = corr_matrix["Immune
Subtype"].abs().sort_values(ascending=False).index[1:21]  # Top 20
excluding self
filtered_corr = corr_matrix.loc[top_features, top_features]

sns.heatmap(filtered_corr, annot=True, cmap='coolwarm',
            fmt='.2f', vmin=-1, vmax=1, center=0,
            annot_kws={"size": 8})
plt.title("Top 20 Features Correlated with Immune Subtype")
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

```
Top correlations with Immune Subtype:
Immune Subtype    1.000000
RASSF4_S3         0.471393
RASSF2_S3         0.467887
CPEB4_S5          0.456713
SLC1A3_S3         0.427795
CELF2_S3          0.415925
```

```
EVI2A_S3          0.410678
APOE_S3           0.408811
CSF1R_S3          0.407040
FAM167A_S5        0.404026
Name: Immune Subtype, dtype: float64

Bottom correlations with Immune Subtype:
NUDT15_S5   -0.514211
MYBL2_S5    -0.513117
ACTL6A_S5   -0.505404
CENPW_S5    -0.496627
CDCA4_S5    -0.491714
SNRPB_S5    -0.487020
MCM3_S5     -0.479625
ZWILCH_S5   -0.479406
PLK4_S5     -0.476722
JUP_S2      -0.476225
Name: Immune Subtype, dtype: float64
```


Feature Correlations with Immune Subtype

```
/usr/local/lib/python3.10/dist-packages/matplotlib/colors.py:721:
RuntimeWarning: invalid value encountered in less
  xa[xa < 0] = -1
```




```
X_selected = df_final_normalized[top_features]
X_selected.head()
```

TCGA Participant Barcode	NUDT15_S5	MYBL2_S5	ACTL6A_S5	CENPW_S5	CDCA4_S5
TCGA-WA-A7H4	0.158879	0.036125	0.129424	0.036657	0.239033
TCGA-B5-A3F9	0.124530	0.032607	0.123828	0.029258	0.158576
TCGA-49-6767	0.164204	0.031808	0.150459	0.022736	

```
0.182581
TCGA-AA-A03J                0.106287  0.108375   0.083904  0.049675
0.109814
TCGA-77-7337                0.159017  0.103269   0.251188  0.049675
0.242704

                           SNRPB_S5   MCM3_S5  ZWILCH_S5   PLK4_S5
JUP_S2  \
TCGA Participant Barcode

TCGA-WA-A7H4               0.110974  0.216977   0.154684  0.272897
0.081848
TCGA-B5-A3F9               0.147041  0.152350   0.144194  0.139299
0.020165
TCGA-49-6767               0.153509  0.115889   0.099884  0.198844
0.048325
TCGA-AA-A03J               0.135313  0.175003   0.079119  0.082470
0.076755
TCGA-77-7337               0.076978  0.076036   0.160216  0.208554
0.102031

                           RASSF4_S3  RASSF2_S3  POLE2_S5  MYCBP_S5
SNRPA1_S5  \
TCGA Participant Barcode

TCGA-WA-A7H4                0.025563   0.006755  0.136847  0.094650
0.098420
TCGA-B5-A3F9                0.028804   0.012288  0.096535  0.157460
0.150227
TCGA-49-6767                0.053798   0.002001  0.152285  0.093955
0.155771
TCGA-AA-A03J                0.026893   0.001996  0.073111  0.080209
0.132591
TCGA-77-7337                0.006769   0.004547  0.182561  0.062224
0.228911

                           CPEB4_S5   SKA1_S5  CHEK1_S5   CCT5_S5
WDHD1_S5
TCGA Participant Barcode

TCGA-WA-A7H4               0.066910  0.023298  0.341859  0.139557
0.079282
TCGA-B5-A3F9               0.037612  0.027082  0.168456  0.123400
0.034451
TCGA-49-6767               0.019654  0.032137  0.319432  0.213498
0.094871
TCGA-AA-A03J               0.024473  0.039170  0.107756  0.159937
0.018202
TCGA-77-7337               0.073843  0.065265  0.422631  0.222876
0.114240
```

```
df.head()
```

```
   ACTL6A_S5  ADAM9_S2  ADAMTS1_S5  ADCY7_S3  AIMP2_S5  ALKBH7_S5
ALOX5AP_S3  \
0   0.129424  0.169197    0.022398  0.234976  0.180090   0.036046
0.025048
1   0.123828  0.033567    0.039275  0.306215  0.153894   0.062357
0.019597
2   0.150459  0.157022    0.003525  0.249774  0.155868   0.019391
0.051084
3   0.083904  0.039410    0.004848  0.088996  0.413569   0.143618
0.013858
4   0.251188  0.217491    0.057262  0.140494  0.383001   0.020735
0.078867

   AMPD3_S3  APITD1_S5  APOC1_S3   APOE_S3   APOO_S5  ARHGAP1_S2
ARHGAP15_S3  \
0  0.083360   0.183705  0.000159  0.001151  0.023024    0.199662
0.034113
1  0.049824   0.299588  0.001520  0.004717  0.137673    0.067752
0.010200
2  0.038690   0.237127  0.001623  0.005735  0.081950    0.066406
0.055246
3  0.020518   0.097687  0.000464  0.002793  0.054958    0.165810
0.009896
4  0.043357   0.325132  0.000145  0.000614  0.069031    0.239196
0.022262

   ARHGDIA_S2  ARRB2_S3    B2M_S3  BCCIP_S5  BRCA2_S5  BRIP1_S5
BSG_S2  \
0    0.193792  0.062812  0.149856  0.137327  0.135782  0.095438
0.044576
1    0.143697  0.062524  0.207232  0.263794  0.042710  0.028896
0.060415
2    0.111708  0.088977  0.166943  0.266651  0.066250  0.052106
0.018203
3    0.310192  0.181536  0.125507  0.104531  0.088548  0.020854
0.052634
4    0.188949  0.069010  0.100442  0.243811  0.089324  0.118525
0.032608

     BTK_S3  C11orf24_S5  C12orf24_S5  C13orf1_S5  C13orf18_S3
C13orf27_S5  \
0  0.034921     0.152497     0.022158    0.155023     0.016833
0.076541
1  0.024206     0.077849     0.031294    0.136670     0.026442
0.070951
2  0.063984     0.143469     0.022671    0.191243     0.005027
0.065347
3  0.010746     0.114090     0.017954    0.247442     0.043303
```

```
0.044817
4  0.034656     0.159544     0.028050    0.165591     0.005849
0.099976

   C16orf61_S5  C19orf48_S5  C1orf54_S3  C3AR1_S3  C3orf26_S5
CASP1_S3  \
0     0.045475     0.019964    0.091298  0.033466    0.252371
0.130334
1     0.076200     0.041087    0.120955  0.038793    0.198402
0.018940
2     0.076236     0.017507    0.080028  0.094222    0.139277
0.032013
3     0.064950     0.076582    0.036174  0.032520    0.106481
0.075871
4     0.078923     0.023613    0.045622  0.093986    0.300077
0.110604

    CBX1_S5  CCBL2_S5   CCL2_S3   CCL5_S3  CCRL2_S3   CCT5_S5
CD14_S4  \
0  0.104727  0.357476  0.017832  0.039156  0.006166  0.139557
0.018676
1  0.171738  0.306802  0.031532  0.017448  0.004626  0.123400
0.022774
2  0.128322  0.402032  0.032674  0.069883  0.012335  0.213498
0.021355
3  0.032466  0.193422  0.006116  0.021166  0.044520  0.159937
0.019312
4  0.142077  0.304238  0.053161  0.019366  0.010812  0.222876
0.025416

   CD163_S3   CD19_S4  CD247_S4   CD33_S3   CD37_S4   CD3D_S4
CD3E_S4  \
0  0.009689  0.003777  0.063614  0.033109  0.015664  0.040169
0.023715
1  0.017677  0.002194  0.035106  0.071084  0.017639  0.040378
0.024578
2  0.025298  0.000785  0.083596  0.075230  0.042144  0.086689
0.076188
3  0.004955  0.003470  0.044976  0.031795  0.022887  0.047478
0.023950
4  0.038462  0.001408  0.039633  0.049115  0.016808  0.034323
0.032421

    CD3G_S4   CD44_S2   CD48_S3   CD52_S3   CD53_S3   CD59_S2
CD79A_S4  \
0  0.016702  0.075255  0.014924  0.018208  0.038376  0.113421
0.012376
1  0.005270  0.008014  0.024867  0.016722  0.053371  0.043971
0.001537
2  0.065392  0.038649  0.040068  0.061403  0.100435  0.047856
```

```
0.003519
3  0.020800  0.012806  0.015225  0.026161  0.035758  0.014871
0.005998
4  0.047515  0.092642  0.020929  0.021409  0.102105  0.107698
0.016165

   CD79B_S4   CD84_S3   CD86_S3   CD8A_S3   CD8B_S4   CD97_S3
CDCA4_S5  \
0  0.009356  0.003355  0.110400  0.004560  0.002090  0.030530
0.239033
1  0.003067  0.004642  0.049360  0.026804  0.016076  0.052208
0.158576
2  0.003751  0.010102  0.144443  0.077024  0.008436  0.073799
0.182581
3  0.005639  0.004176  0.036119  0.037436  0.006757  0.201381
0.109814
4  0.005231  0.018940  0.188385  0.028269  0.008949  0.052528
0.242704

    CDH6_S2   CDK2_S5  CECR1_S3  CELF2_S3  CENPJ_S5  CENPN_S5
CENPO_S5  \
0  0.004352  0.029561  0.022926  0.054613  0.077392  0.058463
0.176886
1  0.003591  0.031723  0.101669  0.044890  0.183160  0.022141
0.065257
2  0.000571  0.038141  0.054018  0.036824  0.028629  0.047939
0.127073
3  0.001170  0.028389  0.024335  0.097108  0.066650  0.052004
0.166096
4  0.028240  0.033506  0.036337  0.041916  0.035145  0.106980
0.339721

   CENPW_S5  CEP78_S5  CHEK1_S5   CKLF_S5  COL16A1_S2  COL1A2_S2
COL3A1_S2  \
0  0.036657  0.158284  0.341859  0.055799    0.130675   0.049270
0.053382
1  0.029258  0.187278  0.168456  0.060539    0.006053   0.005468
0.011437
2  0.022736  0.143665  0.319432  0.279732    0.023565   0.011242
0.010848
3  0.049675  0.048390  0.107756  0.138498    0.022623   0.016113
0.012296
4  0.049675  0.336437  0.422631  0.063020    0.028620   0.055869
0.059424

   COL6A1_S2  COL6A3_S2  COL8A1_S2  COPS6_S5   COQ2_S5  CORO1A_S3
CORO1C_S5  \
0   0.090780   0.168243   0.020451  0.073187  0.151087   0.039494
0.153347
1   0.041633   0.007959   0.000976  0.111060  0.135605   0.031174
```

```
0.040879
2   0.083508   0.050570   0.038859  0.235384  0.148317   0.079423
0.212910
3   0.036574   0.051404   0.023293  0.115420  0.190231   0.173819
0.140236
4   0.030022   0.203530   0.087967  0.089021  0.137881   0.027044
0.337473

   COTL1_S5  COX17_S5  CPEB4_S5   CPVL_S3  CSF1R_S3  CTNNA1_S2
CTNNB1_S2  \
0  0.089469  0.077845  0.066910  0.008639  0.035377   0.332514
0.163481
1  0.102082  0.119659  0.037612  0.009777  0.031854   0.324371
0.128645
2  0.274675  0.130252  0.019654  0.016773  0.048437   0.276330
0.273783
3  0.068212  0.087111  0.024473  0.009468  0.032764   0.336343
0.202381
4  0.110654  0.053663  0.073843  0.012768  0.056842   0.322352
0.248077

    CTSC_S3  CTSL1_S3   CTSS_S3  CXCL10_S1  CXCR4_S3   CYBB_S3
CYTIP_S3  \
0  0.095917  0.023556  0.018769   0.003976  0.009499  0.018678
0.030886
1  0.058378  0.034386  0.040411   0.006545  0.083557  0.018557
0.038931
2  0.080481  0.025965  0.171797   0.032839  0.019742  0.066786
0.049761
3  0.042648  0.007856  0.069236   0.006464  0.008578  0.013780
0.012738
4  0.191799  0.044214  0.046477   0.012050  0.038591  0.066092
0.060276

   DAPK1_S3  DBNDD1_S5  DCBLD2_S5    DCK_S5  DCLRE1B_S5  DDX58_S1
DDX60_S1  \
0  0.082398   0.015780   0.107223  0.160330    0.287794  0.053464
0.072989
1  0.041218   0.066342   0.069558  0.103752    0.207635  0.026454
0.059039
2  0.013979   0.069333   0.107223  0.256074    0.232055  0.031037
0.078268
3  0.026551   0.027064   0.016922  0.102256    0.117280  0.027088
0.072232
4  0.027985   0.013411   0.101117  0.293041    0.243294  0.014842
0.074897

    DHFR_S5  DLEU1_S5  DLEU2_S5  DOCK2_S3    DSP_S2    DUT_S5
DVL1_S2  \
0  0.085747  0.097426  0.145469  0.044880  0.118441  0.202159
```

```
0.284554
1  0.165736  0.245137  0.103178  0.021801  0.017743  0.286724
0.189871
2  0.086690  0.111307  0.224930  0.071234  0.036720  0.093573
0.239620
3  0.152287  0.163563  0.053175  0.032179  0.033738  0.357800
0.412409
4  0.065244  0.207484  0.215501  0.069890  0.118441  0.082165
0.101505

    DVL3_S2  DYNLT1_S5  EBNA1BP2_S5  EFNA5_S2  EIF2AK1_S5  EIF4EBP1_S5
\
0  0.069952   0.117494     0.254913  0.100997    0.180324     0.045329

1  0.051480   0.268827     0.138149  0.012586    0.275381     0.067787

2  0.055469   0.111976     0.229355  0.003574    0.234314     0.050975

3  0.082258   0.106556     0.141679  0.001799    0.280938     0.065817

4  0.123503   0.112132     0.123693  0.058037    0.311581     0.112676


   EIF4G1_S5   EMP2_S5   ENO1_S5  EPHA2_S2  EPHB2_S2  EPHB3_S2
EPHB4_S2  \
0   0.224737  0.073334  0.176186  0.100160  0.030938  0.032859
0.146180
1   0.099652  0.037854  0.106735  0.038517  0.027591  0.009309
0.103594
2   0.234288  0.092133  0.181019  0.065869  0.119018  0.014559
0.315124
3   0.181465  0.046118  0.087059  0.199168  0.119018  0.072677
0.288277
4   0.303863  0.117772  0.311886  0.153488  0.119018  0.059069
0.105869

   ERLIN1_S5  EVI2A_S3  EVI2B_S3  EXOSC8_S5    EZR_S5     F3_S5
FAM167A_S5  \
0   0.144039  0.018113  0.023936   0.161982  0.171173  0.084577
0.010754
1   0.204526  0.016738  0.019382   0.184557  0.076877  0.011839
0.002479
2   0.411496  0.041415  0.057620   0.227998  0.257706  0.033512
0.008293
3   0.213134  0.004571  0.012131   0.180274  0.154220  0.012739
0.004658
4   0.411496  0.048020  0.049250   0.111053  0.132114  0.036959
0.003482

   FAM89B_S5  FARSA_S5  FARSB_S5  FBXO41_S5  FCER1G_S3  FCGR1A_S3
```

```
FCGR2A_S3  \
0   0.204180  0.102053  0.329130   0.047208   0.018273   0.006308
0.030941
1   0.169258  0.111014  0.328010   0.007366   0.044371   0.028072
0.025621
2   0.316784  0.065963  0.376409   0.048723   0.077242   0.066439
0.067265
3   0.160079  0.230402  0.287552   0.107019   0.017333   0.007035
0.028258
4   0.129489  0.113086  0.279407   0.012326   0.045339   0.019803
0.161422

   FCGR2B_S3  FCGR3A_S3   FGD1_S2   FGL2_S3    FGR_S3   FLI1_S3
FLNC_S5  \
0   0.012875   0.013803  0.020552  0.006781  0.032155  0.083692
0.031307
1   0.010577   0.024701  0.088063  0.003095  0.017032  0.046122
0.001224
2   0.021345   0.071530  0.093016  0.011098  0.067096  0.064801
0.031307
3   0.003753   0.012462  0.025836  0.005562  0.030626  0.035335
0.013318
4   0.030141   0.057074  0.043341  0.011555  0.039693  0.086868
0.002109

   FMNL1_S3    FN1_S2  FNBP1_S3   FPR3_S3    FYB_S3    GGH_S5
GIMAP4_S3  \
0  0.101362  0.027169  0.046967  0.056917  0.053084  0.027823
0.025146
1  0.022260  0.001267  0.065727  0.020283  0.041086  0.021042
0.039299
2  0.272864  0.028097  0.098693  0.104719  0.066201  0.010985
0.071292
3  0.074379  0.003745  0.039936  0.040684  0.010794  0.019023
0.024416
4  0.059797  0.051011  0.032962  0.125610  0.181051  0.038501
0.049023

   GLRX3_S5  GNG11_S5  GNPTAB_S3  GPLD1_S5  GPNMB_S3  GSTCD_S5
H2AFZ_S5  \
0  0.160717  0.010046   0.057354  0.000408  0.053609  0.109785
0.152008
1  0.149249  0.013832   0.051172  0.000736  0.012891  0.137458
0.199343
2  0.331305  0.003070   0.026608  0.000730  0.036157  0.102690
0.123665
3  0.174956  0.002216   0.023469  0.000500  0.003611  0.074926
0.145930
4  0.331305  0.005633   0.087136  0.000756  0.101994  0.158231
0.179624
```

```
    HAS2_S5  HAUS1_S5    HCK_S3  HCLS1_S3    HDC_S4  HERC5_S1
HERC6_S1  \
0  0.037437  0.020346  0.047622  0.057652  0.001211  0.025456
0.075657
1  0.014518  0.067098  0.038908  0.086070  0.000030  0.014123
0.053794
2  0.004785  0.015883  0.099041  0.085216  0.000035  0.009157
0.022453
3  0.010080  0.021343  0.050230  0.035609  0.000945  0.005109
0.041770
4  0.061951  0.033436  0.065982  0.227667  0.000368  0.036022
0.051997

   HLA-DMA_S3  HLA-DRB1_S3  HMGN2_S5  HMHA1_S3   HN1L_S5  HNRNPA2B1_S5
\
0    0.022540     0.047645  0.136413  0.071628  0.271469      0.475610

1    0.103293     0.086854  0.244664  0.034324  0.280583      0.548851

2    0.060781     0.204553  0.142709  0.088358  0.161530      0.251151

3    0.053130     0.069789  0.071459  0.163964  0.157864      0.164280

4    0.052412     0.073172  0.168506  0.019563  0.358311      0.172482


   HNRNPR_S5  HSPB11_S5  HSPG2_S2  HYLS1_S5  ICAM1_S2    ID2_S5
ID3_S5  \
0   0.399541   0.218957  0.188063  0.109928  0.034119  0.037530
0.030801
1   0.388752   0.224201  0.064808  0.062105  0.012642  0.209001
0.159889
2   0.299662   0.096953  0.078204  0.078556  0.118333  0.032776
0.007644
3   0.170824   0.041110  0.171041  0.021903  0.022959  0.106346
0.055888
4   0.437498   0.073749  0.206996  0.151481  0.066420  0.037849
0.035191

   IFI16_S3  IFI27_S1  IFI30_S3  IFI44_S1  IFI44L_S1   IFI6_S1
IFIH1_S1  \
0  0.174878  0.010585  0.038855  0.092889   0.008884  0.016181
0.052465
1  0.044124  0.030673  0.047912  0.042425   0.014465  0.041187
0.044148
2  0.064804  0.033371  0.092049  0.082192   0.066570  0.005650
0.088845
3  0.011214  0.056305  0.077159  0.012154   0.010863  0.031855
0.034967
```

```
4  0.190937  0.023718  0.078464  0.022584   0.002674  0.007654
0.061444

   IFIT1_S1  IFIT2_S1  IFIT3_S1  IFRD2_S5  IGF2R_S2  IGFBP2_S2
IGFBP3_S2  \
0  0.012104  0.013984  0.032078  0.174951  0.417760   0.008048
0.011803
1  0.009991  0.020951  0.022286  0.148126  0.090358   0.046128
0.017299
2  0.007317  0.059526  0.069046  0.249150  0.292550   0.000552
0.008982
3  0.009420  0.017772  0.046223  0.406637  0.304512   0.004864
0.010374
4  0.004512  0.012758  0.028578  0.124414  0.286602   0.029685
0.073992

   IGFBP4_S2  IGFBP5_S2    IGJ_S4  IKZF1_S3  IL10RA_S3   IL18_S3
IL7R_S3  \
0   0.076637   0.006510  0.055104  0.030634   0.043147  0.083119
0.079098
1   0.259275   0.002993  0.003288  0.016425   0.033162  0.005650
0.010949
2   0.053771   0.007041  0.002262  0.039746   0.075596  0.109420
0.013034
3   0.062825   0.004197  0.009832  0.014251   0.045385  0.025072
0.010332
4   0.099363   0.027515  0.038158  0.027573   0.042415  0.051308
0.094333

    IMP4_S5   IPO4_S5  ISG15_S1  ITGA3_S2  ITGA4_S2  ITGA5_S2
ITGA6_S5  \
0  0.139524  0.207161  0.014778  0.251271  0.053032  0.086581
0.139174
1  0.130361  0.212707  0.037194  0.015699  0.018388  0.012094
0.032617
2  0.160946  0.200339  0.022078  0.251271  0.034460  0.049861
0.042737
3  0.283705  0.306210  0.056967  0.068338  0.023130  0.024807
0.061921
4  0.150310  0.249193  0.008037  0.105767  0.067336  0.151642
0.139174

   ITGB2_S2  ITGB3_S2  ITGB5_S2  ITGB8_S2   JUNB_S5    JUP_S2
KIAA0090_S5  \
0  0.021288  0.017033  0.202443  0.024935  0.028278  0.081848
0.409317
1  0.026752  0.001087  0.097532  0.012690  0.029653  0.020165
0.161326
2  0.100076  0.007168  0.096397  0.027774  0.011750  0.048325
0.171628
```

```
3  0.036380  0.003954  0.107455  0.003000  0.051715  0.076755
0.081225
4  0.038590  0.014997  0.112845  0.013661  0.027507  0.102031
0.409317

    KRR1_S5  LAIR1_S3  LAMA4_S2  LAMB1_S2  LAPTM5_S3    LCK_S4
LCP1_S3  \
0  0.266041  0.031268  0.063886  0.134072   0.029875  0.023022
0.057339
1  0.163491  0.054181  0.010380  0.166282   0.057293  0.023920
0.078932
2  0.144502  0.087119  0.012079  0.102749   0.151932  0.060208
0.098733
3  0.086434  0.035759  0.020305  0.133157   0.037515  0.026947
0.024536
4  0.355570  0.050773  0.065265  0.125134   0.083237  0.029194
0.103698

    LCP2_S3  LHFPL2_S3  LILRB4_S3  LMNB2_S5  LOXL2_S5   LRMP_S3
LRP1_S2  \
0  0.062984   0.132281   0.008157  0.250072  0.030669  0.004127
0.021220
1  0.038561   0.092902   0.020459  0.144390  0.008254  0.002365
0.009356
2  0.108373   0.124139   0.039352  0.172712  0.023336  0.001690
0.010615
3  0.035882   0.090632   0.016934  0.366108  0.024214  0.001860
0.042072
4  0.121301   0.178464   0.027372  0.233035  0.045484  0.025887
0.048902

   LRRC17_S2  LRRC40_S5   LSM3_S5   LSM4_S5   LST1_S3    LTB_S4
LY86_S3  \
0   0.060840   0.178373  0.109577  0.100458  0.045123  0.042660
0.009661
1   0.025620   0.193141  0.163247  0.198695  0.031602  0.018470
0.030685
2   0.011497   0.233212  0.156403  0.109571  0.061070  0.057014
0.040346
3   0.003852   0.017883  0.069174  0.172504  0.037040  0.024858
0.014650
4   0.016909   0.136157  0.111263  0.089692  0.026585  0.002604
0.019700

    LYAR_S5    LYN_S3    LYZ_S3  MAGOHB_S5  MAP3K8_S5  MAPRE1_S5
MARCKSL1_S2  \
0  0.210344  0.114421  0.031317   0.075697   0.035770   0.145664
0.040757
1  0.104753  0.099920  0.001993   0.107498   0.062074   0.090899
0.385772
```

```
2  0.229836  0.346249  0.016032   0.104949   0.027099   0.134828
0.170430
3  0.189843  0.072093  0.007370   0.048341   0.018603   0.160546
0.126210
4  0.229836  0.239733  0.016140   0.135993   0.038159   0.246152
0.027627

   MARVELD2_S5   MCM3_S5   MCM7_S5  MCTS1_S5  MERTK_S3    MET_S5
MFSD11_S5  \
0     0.113350  0.216977  0.074077  0.217205  0.010786  0.028999
0.368101
1     0.190823  0.152350  0.075562  0.121013  0.046835  0.013101
0.097814
2     0.179898  0.115889  0.199362  0.217010  0.033905  0.072149
0.274009
3     0.227938  0.175003  0.194599  0.101290  0.007017  0.019980
0.164033
4     0.180933  0.076036  0.081903  0.082588  0.035576  0.032555
0.316945

    MKKS_S5  MLF1IP_S5  MMP11_S2   MMP1_S2  MMP14_S2  MMP17_S2
MMP19_S2  \
0  0.106318   0.169909  0.073249  0.007719  0.138782  0.028384
0.021184
1  0.118294   0.188447  0.073249  0.000080  0.054637  0.005161
0.028822
2  0.054881   0.168724  0.068895  0.007719  0.169203  0.012879
0.011475
3  0.035276   0.074050  0.062199  0.007719  0.054713  0.003805
0.010316
4  0.063507   0.104407  0.017603  0.007719  0.087733  0.009456
0.012750

    MMP2_S2   MMP3_S2  MNAT1_S5   MNDA_S3   MPP1_S3  MRPL12_S5
MRPL37_S5  \
0  0.076244  0.004697  0.297820  0.034854  0.028984   0.053989
0.215832
1  0.067427  0.001807  0.460132  0.023676  0.023693   0.071177
0.149825
2  0.030269  0.000234  0.363728  0.134574  0.113771   0.053678
0.175349
3  0.044198  0.004697  0.067786  0.015200  0.119624   0.128341
0.340748
4  0.092904  0.004697  0.351699  0.055352  0.033229   0.046913
0.150464

   MRPS16_S5  MRPS28_S5  MRTO4_S5  MS4A1_S4    MSN_S3   MT1F_S5
MT3_S5  \
0   0.081801   0.152298  0.296662  0.001867  0.225881  0.001859
0.000017
```

```
1   0.159533   0.102605  0.261444  0.000091  0.102073  0.007738
0.000714
2   0.146096   0.056819  0.206756  0.000514  0.523028  0.001701
0.000002
3   0.150107   0.026642  0.162677  0.000202  0.046191  0.005101
0.000447
4   0.227553   0.115047  0.292974  0.000489  0.342708  0.001910
0.000003

    MTA1_S2  MTHFD1_S5    MX1_S1  MYBL1_S5  MYBL2_S5  MYCBP_S5
MYL6_S5  \
0  0.108317   0.053998  0.023341  0.044450  0.036125  0.094650
0.092348
1  0.205529   0.049750  0.022720  0.011306  0.032607  0.157460
0.113473
2  0.180721   0.111187  0.022676  0.059719  0.031808  0.093955
0.110403
3  0.161050   0.065647  0.048149  0.005645  0.108375  0.080209
0.084048
4  0.101209   0.098276  0.018135  0.014082  0.103269  0.062224
0.073702

   MYO1F_S3  NCEH1_S5   NCF2_S3  NCKAP1L_S3   NCLN_S5   NEO1_S2
NID1_S2  \
0  0.060112  0.033299  0.088336    0.041589  0.124847  0.103046
0.011977
1  0.070195  0.056313  0.018637    0.033273  0.134272  0.214586
0.011057
2  0.105832  0.160207  0.182093    0.078590  0.109027  0.077230
0.008704
3  0.102991  0.036514  0.035316    0.045108  0.249041  0.307042
0.007268
4  0.046165  0.096981  0.237740    0.061815  0.095514  0.075113
0.034183

     NLN_S5   NME1_S5  NOP16_S5  NOTCH2_S2   NPC2_S3    NPL_S3
NRIP3_S5  \
0  0.239173  0.064625  0.164285   0.111047  0.015887  0.049165
0.003531
1  0.232128  0.127969  0.271940   0.120854  0.047381  0.029214
0.000597
2  0.331265  0.099628  0.135435   0.145194  0.056083  0.026651
0.004892
3  0.205279  0.084575  0.345630   0.028016  0.009288  0.019027
0.000524
4  0.281952  0.074518  0.102710   0.157209  0.052034  0.165051
0.009000

   NUDT1_S5  NUDT15_S5  NUP107_S5  NUP35_S5  NUP85_S5  NUP93_S5
NUPL1_S5  \
```

```
0  0.080380   0.158879   0.083663  0.173388  0.174390  0.121559
0.210659
1  0.057976   0.124530   0.071152  0.272161  0.137736  0.136196
0.120356
2  0.050619   0.164204   0.071165  0.173997  0.171342  0.165973
0.183313
3  0.072168   0.106287   0.048762  0.081494  0.186738  0.159651
0.194111
4  0.089115   0.159017   0.091544  0.314542  0.138004  0.163207
0.169962

   NUTF2_S5   OAS1_S1   OAS2_S1   OAS3_S1   OASL_S1  OSBPL3_S3
PA2G4_S5  \
0  0.143198  0.035967  0.045707  0.045250  0.018917   0.153589
0.079970
1  0.307269  0.012002  0.013871  0.020847  0.013454   0.058412
0.160077
2  0.285336  0.075691  0.046578  0.067696  0.033349   0.112185
0.218809
3  0.328946  0.121539  0.034826  0.094669  0.063285   0.081699
0.217128
4  0.194058  0.030596  0.056352  0.051430  0.009631   0.063854
0.183514

   PAICS_S5   PAK1_S2   PAK2_S2  PDAP1_S5  PDIA4_S5  PDLIM7_S5
PFKP_S5  \
0  0.166344  0.072260  0.241153  0.168676  0.135598   0.207377
0.049983
1  0.224232  0.117644  0.126378  0.210840  0.335843   0.075176
0.032116
2  0.104921  0.116603  0.351109  0.331346  0.456125   0.106854
0.132821
3  0.162434  0.082411  0.083170  0.215826  0.174828   0.117111
0.132418
4  0.265156  0.127628  0.389034  0.126817  0.178679   0.053024
0.158858

    PFN1_S5   PGM2_S5  PHF19_S5  PIK3CG_S3  PITPNC1_S5   PLAT_S2
PLAUR_S5  \
0  0.255600  0.176835  0.191977   0.024237    0.076846  0.001419
0.039017
1  0.191584  0.054117  0.053123   0.002969    0.131413  0.001808
0.030971
2  0.170042  0.119041  0.348527   0.017677    0.040962  0.009950
0.122240
3  0.303984  0.054715  0.385093   0.008843    0.051307  0.002175
0.120793
4  0.110226  0.176835  0.115697   0.037040    0.103141  0.011542
0.076734
```

```
   PLCG2_S3   PLEK_S3    PLG_S5   PLK4_S5  PLOD2_S5    PNN_S5
PNO1_S5  \
0  0.006507  0.023230  0.000005  0.272897  0.015910  0.147280
0.161602
1  0.004564  0.030490  0.000023  0.139299  0.019097  0.197264
0.096395
2  0.003650  0.070551  0.000005  0.198844  0.011098  0.099750
0.184627
3  0.001889  0.016255  0.000023  0.082470  0.003602  0.090495
0.163267
4  0.009946  0.057403  0.000005  0.208554  0.048598  0.111271
0.322468

   POLE2_S5  POLR3K_S5   PPIH_S5  PSMA7_S5  PSMC3_S5  PSMD12_S5
PSMD14_S5  \
0  0.136847   0.089785  0.144393  0.125138  0.167270   0.139899
0.268905
1  0.096535   0.137554  0.232093  0.073134  0.156777   0.062331
0.238535
2  0.152285   0.077284  0.175007  0.135764  0.108752   0.113334
0.216624
3  0.073111   0.214049  0.200183  0.229206  0.104846   0.035758
0.154648
4  0.182561   0.186561  0.084329  0.185575  0.178016   0.194763
0.482800

   PSMD2_S5  PTPLB_S5  PTPRC_S3    PXN_S2  RAB3B_S5   RAC1_S2
RASSF2_S3  \
0  0.157737  0.084729  0.024735  0.135201  0.001371  0.153141
0.006755
1  0.063664  0.011875  0.016258  0.030791  0.000577  0.167645
0.012288
2  0.125575  0.066794  0.041973  0.139315  0.005981  0.161858
0.002001
3  0.094678  0.020744  0.009985  0.139315  0.001737  0.194394
0.001996
4  0.223142  0.147774  0.049721  0.091718  0.003655  0.343638
0.004547

   RASSF4_S3  RBM14_S5   RFC3_S5   RGS8_S5   RHOA_S2   RHOB_S2
RHOC_S5  \
0   0.025563  0.282242  0.095780  0.000000  0.254724  0.040654
0.129998
1   0.028804  0.387320  0.117169  0.000000  0.276097  0.170179
0.119639
2   0.053798  0.278413  0.106650  0.000000  0.417517  0.012437
0.230437
3   0.026893  0.587550  0.145781  0.014786  0.276152  0.050956
0.191826
4   0.006769  0.382809  0.145882  0.002612  0.318228  0.039787
```

```
0.131475

    RHOG_S2   RHOQ_S2  RMND5B_S5  RNASE6_S3   RND3_S2  RNF138_S5
RNF41_S5  \
0  0.174512  0.107443   0.097324   0.034370  0.133599   0.087150
0.089728
1  0.117589  0.108968   0.181132   0.029793  0.149408   0.197635
0.124705
2  0.159244  0.082045   0.153058   0.042139  0.097058   0.104717
0.115594
3  0.243694  0.011838   0.278400   0.034181  0.030692   0.063504
0.164837
4  0.115530  0.086344   0.073182   0.027792  0.198128   0.152272
0.178470

    RPN1_S5  RPP40_S5  RSAD2_S1   RTP4_S1  RUNX3_S3  RUVBL1_S5
SAMD9_S1  \
0  0.169853  0.203581  0.011339  0.022262  0.040771   0.176109
0.065363
1  0.339800  0.126313  0.006195  0.047336  0.011941   0.304404
0.007888
2  0.218984  0.175488  0.021158  0.032420  0.012582   0.228971
0.010268
3  0.201962  0.157057  0.043230  0.030058  0.069775   0.205084
0.012855
4  0.482965  0.242419  0.020195  0.063709  0.064088   0.419433
0.048694

   SAMHD1_S3  SAMSN1_S3  SAR1A_S5  SAR1B_S5   SDC1_S5   SELL_S3
SELPLG_S3  \
0   0.052966   0.047820  0.206790  0.100884  0.098025  0.004675
0.037912
1   0.130442   0.053131  0.188092  0.083722  0.013992  0.004051
0.029117
2   0.081837   0.083075  0.158464  0.069516  0.060385  0.005374
0.081667
3   0.067619   0.031994  0.049852  0.055348  0.040424  0.001718
0.058516
4   0.115436   0.079174  0.354664  0.111418  0.131077  0.004864
0.052901

   SEMA3F_S2  SERPINE1_S2  SH2B3_S3  SH3BP5L_S5   SKA1_S5   SKA2_S5  \
0   0.107456     0.047018  0.071774    0.397680  0.023298  0.150416
1   0.075353     0.000925  0.022738    0.161268  0.027082  0.163312
2   0.044402     0.011848  0.141412    0.147696  0.032137  0.151942
3   0.041377     0.001889  0.051041    0.289520  0.039170  0.019166
4   0.132073     0.053159  0.074056    0.161067  0.065265  0.160582

   SLC16A1_S5  SLC1A3_S3  SLC25A40_S5  SLC25A5_S5  SLC7A7_S3   SMC2_S5
\
```

```
0    0.187154   0.004853     0.113213    0.066678   0.008365  0.111563

1    0.101799   0.001806     0.050642    0.025590   0.008739  0.059632

2    0.170208   0.001690     0.296410    0.051780   0.030765  0.122536

3    0.068127   0.000242     0.103375    0.096260   0.017580  0.029403

4    0.342128   0.018085     0.194974    0.062755   0.017081  0.181299


     SMO_S2    SMS_S5  SMURF2_S5  SNRPA1_S5  SNRPA_S5  SNRPB_S5
SNRPC_S5  \
0  0.061107  0.010822   0.117046   0.098420  0.129965  0.110974
0.189107
1  0.223859  0.042159   0.080097   0.150227  0.357502  0.147041
0.194742
2  0.120447  0.044985   0.240120   0.155771  0.129030  0.153509
0.222232
3  0.041830  0.026457   0.055882   0.132591  0.353250  0.135313
0.214041
4  0.065579  0.028911   0.139213   0.228911  0.135565  0.076978
0.224079

   SNRPD1_S5  SNRPE_S5  SNX17_S5  SP140_S3  SPAG17_S5  SPARC_S2
SRM_S5  \
0   0.082965  0.033202  0.217200  0.025957   0.016643  0.045811
0.102953
1   0.211044  0.106333  0.230067  0.012544   0.066766  0.010303
0.081692
2   0.130903  0.081369  0.169501  0.047222   0.000769  0.012549
0.132022
3   0.078309  0.063577  0.174395  0.009958   0.000777  0.012536
0.129299
4   0.174427  0.061940  0.178775  0.015842   0.027190  0.056416
0.118792

   STAT1_S1  STK17A_S5  STRA13_S5    SYK_S3  TAGLN_S5   TAP1_S1
TBXAS1_S3  \
0  0.044481   0.229143   0.076512  0.059110  0.038161  0.078096
0.035709
1  0.055906   0.059119   0.099121  0.168666  0.003984  0.086456
0.087330
2  0.118732   0.229143   0.054861  0.053159  0.007983  0.153990
0.039003
3  0.043672   0.024340   0.104535  0.138534  0.025773  0.173724
0.236709
4  0.099685   0.158866   0.026717  0.132907  0.027327  0.064915
0.038927
```

```
   TCEB1_S5  TCF7L2_S2  THBS1_S2  THBS2_S2  TIMP1_S2   TLR2_S3
TMEM130_S5  \
0  0.130174   0.099239  0.178477  0.065712  0.028905  0.148018
0.000175
1  0.103268   0.162975  0.012453  0.001430  0.027609  0.075820
0.000108
2  0.143786   0.199046  0.012716  0.028705  0.038713  0.048070
0.000505
3  0.052949   0.228019  0.031254  0.018995  0.047791  0.023434
0.000256
4  0.159103   0.164192  0.093766  0.106684  0.041740  0.263277
0.000990

     TNC_S2  TNFAIP3_S3  TNFRSF12A_S5  TNFRSF1A_S2  TNFRSF1B_S3
TNFSF13B_S3  \
0  0.115225    0.080005      0.168056     0.252215     0.044209
0.007273
1  0.015413    0.100122      0.044841     0.219536     0.089376
0.005195
2  0.115225    0.054746      0.121907     0.350215     0.129051
0.027343
3  0.017496    0.058073      0.108840     0.273136     0.162439
0.007791
4  0.115225    0.114188      0.049923     0.433645     0.040264
0.019876

   TOMM40_S5   TPI1_S5   TPM1_S5   TPM2_S5   TPM3_S5  TPRKB_S5
TRA2B_S5  \
0   0.030447  0.055560  0.085022  0.054904  0.248836  0.224479
0.370910
1   0.039365  0.121625  0.019840  0.010281  0.191961  0.427167
0.307541
2   0.044709  0.170366  0.031936  0.029378  0.219096  0.172608
0.150605
3   0.085939  0.142167  0.048872  0.013824  0.219590  0.089482
0.170531
4   0.041386  0.247876  0.027684  0.008869  0.276453  0.239914
0.384092

   TUBA4A_S5  TUBG1_S5   UAP1_S5  UBE2J1_S5   UMPS_S5  UQCR10_S5
USPL1_S5  \
0   0.097268  0.249644  0.107880   0.182237  0.252513   0.037382
0.201070
1   0.002220  0.227805  0.080434   0.249096  0.145911   0.072831
0.190156
2   0.040278  0.238908  0.078863   0.267850  0.215546   0.063526
0.178736
3   0.097268  0.293324  0.038459   0.137453  0.194995   0.042015
0.102779
4   0.063399  0.435219  0.129999   0.277074  0.312816   0.022331
```

```
0.054357

    VCAN_S2  VDAC1_S5  VSIG4_S3   VTA1_S5  WDHD1_S5  WDR54_S5
WDR77_S5  \
0  0.072346  0.195607  0.003054  0.269151  0.079282  0.072920
0.102072
1  0.108592  0.229437  0.013031  0.175206  0.034451  0.188849
0.240956
2  0.040453  0.227530  0.023227  0.184954  0.094871  0.037184
0.124078
3  0.018265  0.353334  0.005992  0.102899  0.018202  0.034451
0.221328
4  0.206705  0.262335  0.017925  0.348754  0.114240  0.052860
0.092856

   WIPF1_S3  WNT2B_S2  WNT8B_S2   WSB2_S5  ZWILCH_S5    ZYX_S2  Immune
Subtype
0  0.089859  0.010692  0.008595  0.124660   0.154684  0.191966
0
1  0.067037  0.006442  0.002612  0.089291   0.144194  0.126821
0
2  0.091530  0.002030  0.003011  0.142595   0.099884  0.293115
0
3  0.026653  0.012035  0.003367  0.134369   0.079119  0.256544
0
4  0.126317  0.041299  0.000520  0.167126   0.160216  0.132789
0

corr_matrix["Immune Subtype"].head()

ACTL6A_S5    -0.505404
ADAM9_S2     -0.072987
ADAMTS1_S5   -0.066917
ADCY7_S3      0.031620
AIMP2_S5     -0.157777
Name: Immune Subtype, dtype: float64
```

Comparision of different feature reduction methods

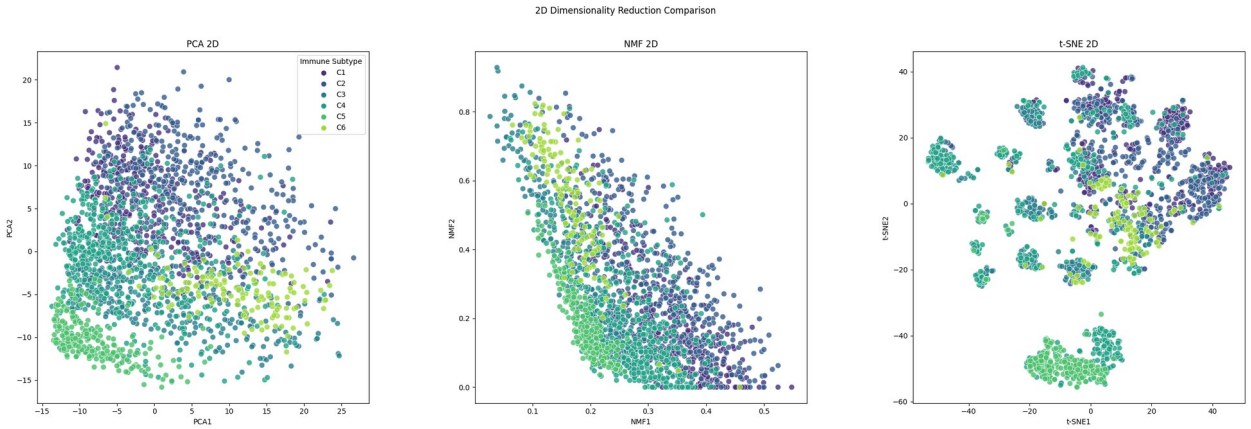
PCA VS NMF VS T-SNE

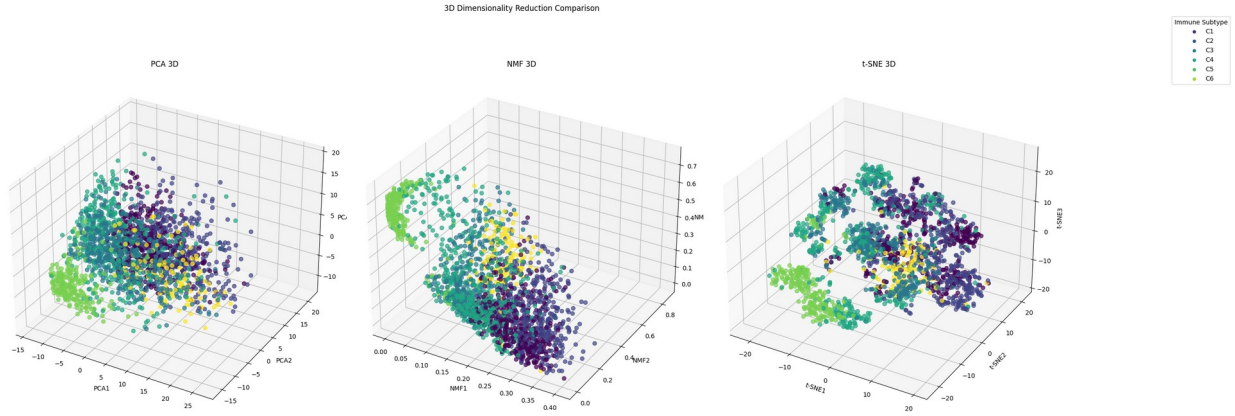
```
# Data preparation
X_full = df_final_normalized.drop(columns=['Immune Subtype'])
y = df_final_normalized['Immune Subtype']

# Scaling
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_full)
minmax_scaler = MinMaxScaler()
```

```
X_nonneg = minmax_scaler.fit_transform(X_full)

# Dimensionality reductions (create all versions but don't visualize
100D)
## PCA
pca = PCA(n_components=0.90)
X_pca = pca.fit_transform(X_scaled)  # For later use
X_pca_2d = PCA(n_components=2).fit_transform(X_scaled)
X_pca_3d = PCA(n_components=3).fit_transform(X_scaled)
X_pca_100d = PCA(n_components=100).fit_transform(X_scaled)  # Created
but not visualized

## NMF
nmf = NMF(n_components=pca.n_components_, init='nndsvda',
random_state=42)
X_nmf = nmf.fit_transform(X_nonneg)  # For later use
X_nmf_2d = NMF(n_components=2, init='nndsvda',
random_state=42).fit_transform(X_nonneg)
X_nmf_3d = NMF(n_components=3, init='nndsvda',
random_state=42).fit_transform(X_nonneg)
X_nmf_100d = NMF(n_components=100, init='nndsvda',
random_state=42).fit_transform(X_nonneg)  # Created but not visualized

## t-SNE
X_tsne_2d = TSNE(n_components=2, perplexity=30,
random_state=42).fit_transform(X_scaled)
X_tsne_3d = TSNE(n_components=3, perplexity=30, random_state=42,
method='exact').fit_transform(X_scaled)
# Skip 100D t-SNE as it's not recommended

# Visualization function for 2D plots
def plot_2d(X, y, title, ax):
    sns.scatterplot(x=X[:, 0], y=X[:, 1], hue=y,
                    palette='viridis', alpha=0.8, s=60, ax=ax)
    ax.set_title(title)
    ax.set_xlabel(f'{title.split()[0]}1')
    ax.set_ylabel(f'{title.split()[0]}2')

# Visualization function for 3D plots
def plot_3d(X, y, title, ax):
    scatter = ax.scatter(X[:, 0], X[:, 1], X[:, 2],
                         c=y.astype('category').cat.codes,
                         cmap='viridis', alpha=0.7, s=40)
    ax.set_title(title)
    ax.set_xlabel(f'{title.split()[0]}1')
    ax.set_ylabel(f'{title.split()[0]}2')
    ax.set_zlabel(f'{title.split()[0]}3')
    return scatter

# Create 2D comparison plot
```

```
fig_2d, axes_2d = plt.subplots(1, 3, figsize=(24, 8))
plot_2d(X_pca_2d, y, 'PCA 2D', axes_2d[0])
plot_2d(X_nmf_2d, y, 'NMF 2D', axes_2d[1])
plot_2d(X_tsne_2d, y, 't-SNE 2D', axes_2d[2])
axes_2d[0].legend(title='Immune Subtype', bbox_to_anchor=(1, 1))
axes_2d[1].legend().remove()
axes_2d[2].legend().remove()
plt.suptitle('2D Dimensionality Reduction Comparison', y=1.02)
plt.tight_layout()
plt.show()

# Create 3D comparison plot
fig_3d = plt.figure(figsize=(24, 10))
ax1 = fig_3d.add_subplot(1, 3, 1, projection='3d')
plot_3d(X_pca_3d, y, 'PCA 3D', ax1)

ax2 = fig_3d.add_subplot(1, 3, 2, projection='3d')
plot_3d(X_nmf_3d, y, 'NMF 3D', ax2)

ax3 = fig_3d.add_subplot(1, 3, 3, projection='3d')
plot_3d(X_tsne_3d, y, 't-SNE 3D', ax3)

# Add single legend
legend_labels = y.unique()
for i, label in enumerate(legend_labels):
    ax1.scatter([], [], [], label=label,
color=sns.color_palette('viridis')[i])
fig_3d.legend(title='Immune Subtype', bbox_to_anchor=(1.1, 1),
loc='upper left')

plt.suptitle('3D Dimensionality Reduction Comparison', y=1.02)
plt.tight_layout()
plt.show()

# All transformed datasets are available for later use:
# X_pca, X_pca_100d, X_nmf, X_nmf_100d (but not visualized)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/
_nmf.py:1665: ConvergenceWarning: Maximum number of iterations 200
reached. Increase it to improve convergence.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/_nmf.py:
1665: ConvergenceWarning: Maximum number of iterations 200 reached.
Increase it to improve convergence.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/decomposition/_nmf.py:
1665: ConvergenceWarning: Maximum number of iterations 200 reached.
Increase it to improve convergence.
  warnings.warn(
```





Best Feature Reduction:

t-SNE → Best separation, distinct clusters.

PCA → Retains global structure but overlaps.

NMF → More structured than PCA but less separation than t-SNE.

For interpretability → PCA/NMF For classification → t-SNE

EDA

```
pca_df = pd.DataFrame(X_pca_3d)

y = le.fit_transform(df_final_normalized['Immune Subtype'])  # Encode categorical labels

# Convert y to DataFrame
y_df = pd.DataFrame(y, columns=['Immune Subtype'])

# Drop target from X

# Reset indices and concatenate
#X = X.reset_index(drop=True)
#y_df = y_df.reset_index(drop=True)
```

```python
df = pd.concat([pca_df, y_df], axis=1)
import sweetviz as sv

# Generate a report
report = sv.analyze(pd.DataFrame(X_pca))

# Display inline in Kaggle
report.show_html("report.html")  # Save if needed
report.show_notebook()  # Inline display
```

{"model_id":"13b099be6ed44d488c8a96267550b88f","version_major":2,"version_minor":0}

```
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py:219: PerformanceWarning: DataFrame is highly fragmented.  This is usually the result of calling `frame.insert` many times, which has poor performance.  Consider joining all columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe = frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py:219: PerformanceWarning: DataFrame is highly fragmented.  This is usually the result of calling `frame.insert` many times, which has poor performance.  Consider joining all columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe = frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py:219: PerformanceWarning: DataFrame is highly fragmented.  This is usually the result of calling `frame.insert` many times, which has poor performance.  Consider joining all columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe = frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py:219: PerformanceWarning: DataFrame is highly fragmented.  This is usually the result of calling `frame.insert` many times, which has poor performance.  Consider joining all columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe = frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py:219: PerformanceWarning: DataFrame is highly fragmented.  This is usually the result of calling `frame.insert` many times, which has poor performance.  Consider joining all columns at once using pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe = frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py:219: PerformanceWarning: DataFrame is highly fragmented.  This is
```

```
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
```

```
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
```

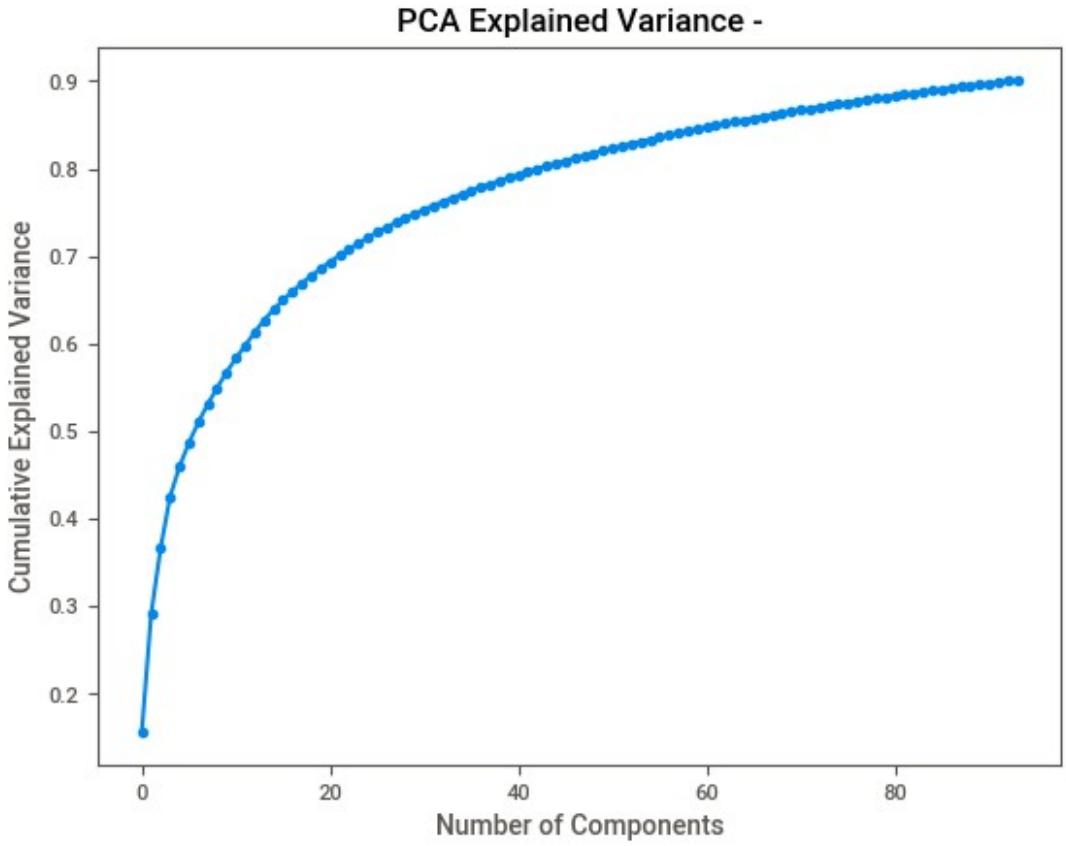
```
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
```

```
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
```

```
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)
/usr/local/lib/python3.10/dist-packages/sweetviz/graph_associations.py
:219: PerformanceWarning: DataFrame is highly fragmented.  This is
usually the result of calling `frame.insert` many times, which has
poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe
= frame.copy()`
  new_dataframe[feature] = pd.Series(dtype=float)

Report report.html was generated! NOTEBOOK/COLAB USERS: the web
browser MAY not pop up, regardless, the report IS saved in your
notebook/colab files.

<IPython.core.display.HTML object>
```

```python
# X_full: Full dataset without dimensionality reduction
X_full = df_final_normalized.drop(columns=['Immune Subtype'])  #
Assuming 'Immune Subtype' is the target variable
# X_pca: PCA-reduced dataset
from sklearn.decomposition import PCA

pca = PCA(n_components=0.90)
X_pca = pca.fit_transform(X_full)

explained_var_ratio = np.cumsum(pca.explained_variance_ratio_)
```

```
plt.plot(explained_var_ratio, marker='o')
plt.xlabel("Number of Components")
plt.ylabel("Cumulative Explained Variance")
plt.title(f"PCA Explained Variance -")
plt.show()
```



Section 2 - Classification Model Development

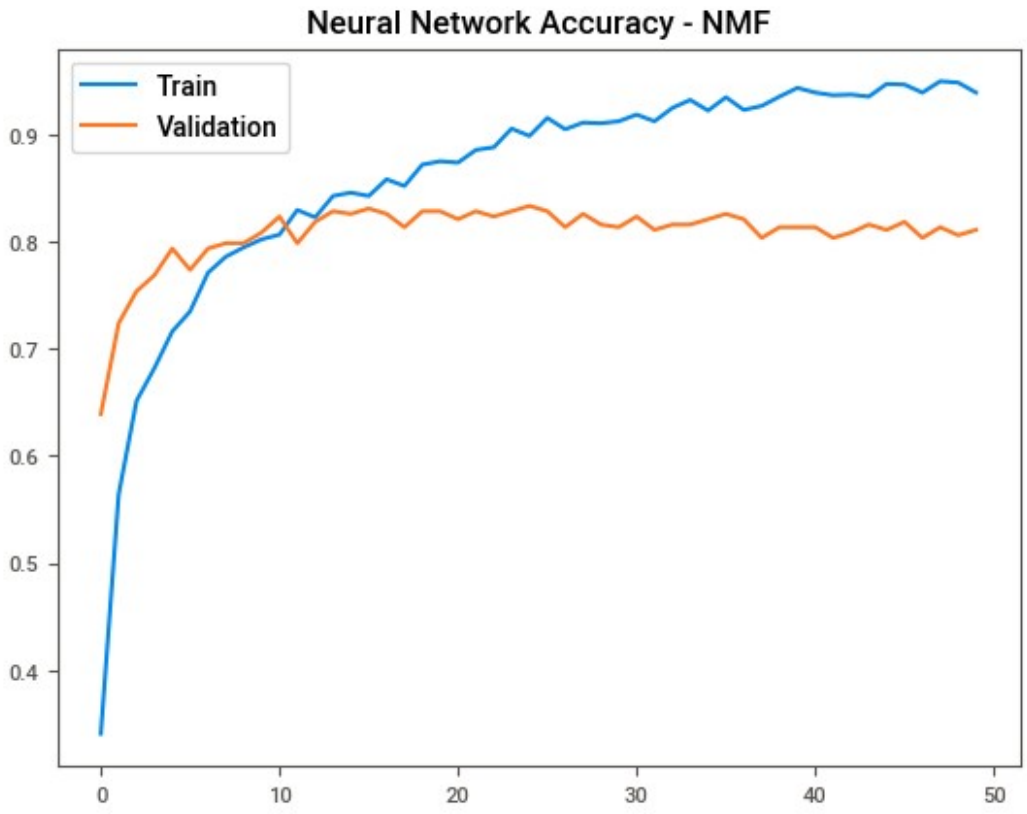
```
# Label Encoding
y = LabelEncoder().fit_transform(df_final_normalized['Immune
Subtype'])

# PCA Transformation (90% Variance Retention)
#pca = PCA(n_components=0.90)
#X_pca = pca.fit_transform(X_full)
#X_FE=df_features.drop(columns=['Immune Subtype'])

# PCA and Kernel PCA
pca = PCA(n_components=0.90)
kpca = KernelPCA(n_components=50, kernel='rbf')
```

```
X_pca = pca.fit_transform(X_full)
X_kpca = kpca.fit_transform(X_full)
datasets = {
    "NMF": X_nmf,
    "X_pca": X_pca,
    "X_kpca": X_kpca,
    #"X_FE":X_FE,
   "X_nmf_100d": X_nmf_100d,
    #"tsne_100d": tsne_100d,
    "X_full": X_full,
   "X_tsne_3d": X_tsne_3d,
   # "Top_20_Correlated_Features": X_selected
}

# Scale Data
scaler = StandardScaler()

def train_and_evaluate_model(model, model_name, X_train, X_test,
y_train, y_test):
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)

    print(f'\n{model_name} Accuracy: {accuracy:.4f}')
    print(classification_report(y_test, y_pred,
target_names=LabelEncoder().fit(df_final_normalized['Immune
Subtype']).classes_))

    # Confusion Matrix
    cm = confusion_matrix(y_test, y_pred)
    plt.figure(figsize=(6, 5))
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
    plt.xlabel('Predicted')
    plt.ylabel('Actual')
    plt.title(f'Confusion Matrix - {model_name}')
    plt.show()

    return accuracy

def create_nn(input_dim):
    model = Sequential([
        Dense(64, activation='relu', input_shape=(input_dim,)),
        Dropout(0.3),
        Dense(32, activation='relu'),
        Dropout(0.3),
        Dense(len(set(y)), activation='softmax')
    ])
    model.compile(optimizer=Adam(learning_rate=0.001),
loss='sparse_categorical_crossentropy', metrics=['accuracy'])
    return model
```

```
results = {}

for name, X in datasets.items():
    print(f"\n===== Training Models on {name} Dataset =====")

    X_train, X_test, y_train, y_test = train_test_split(X, y,
stratify=y,test_size=0.2, random_state=42)
    print(f"Training set: {X_train.shape}, Test set: {X_test.shape}")
    X_train = scaler.fit_transform(X_train)
    X_test = scaler.transform(X_test)

    # Neural Network Training
    nn_model = create_nn(X_train.shape[1])
    history = nn_model.fit(X_train, y_train, validation_data=(X_test,
y_test), epochs=50, batch_size=16, verbose=0)

    plt.plot(history.history['accuracy'], label='Train')
    plt.plot(history.history['val_accuracy'], label='Validation')
    plt.title(f'Neural Network Accuracy - {name}')
    plt.legend()
    plt.show()

    y_pred_nn = nn_model.predict(X_test).argmax(axis=1)
    print(f"\nNeural Network - {name} Classification Report:\n",
classification_report(y_test, y_pred_nn))

    # Random Forest Classifier
    rf_clf = RandomForestClassifier(n_estimators=200, max_depth=10,
random_state=42)
    results[f"RandomForest_{name}"] = train_and_evaluate_model(rf_clf,
"Random Forest", X_train, X_test, y_train, y_test)

    # XGBoost with Hyperparameter Tuning
    xgb_params = {'max_depth': [1,2], 'learning_rate': [0.01, 0.1],
'n_estimators': [100, 200]}
    xgb_clf = GridSearchCV(XGBClassifier(eval_metric='logloss',
random_state=42), xgb_params, cv=3)
    xgb_clf.fit(X_train, y_train)
    print(f"Best XGBoost Parameters for {name}:
{xgb_clf.best_params_}")
    results[f"XGBoost_{name}"] =
train_and_evaluate_model(xgb_clf.best_estimator_, "XGBoost", X_train,
X_test, y_train, y_test)
     # SVM with Hyperparameter Tuning
    svm_params = {
        'C': [0.1, 1, 10, 50, 100],
        'gamma': ['scale', 'auto', 0.001, 0.01, 0.1],
        'kernel': ['linear', 'rbf', 'poly'],
```

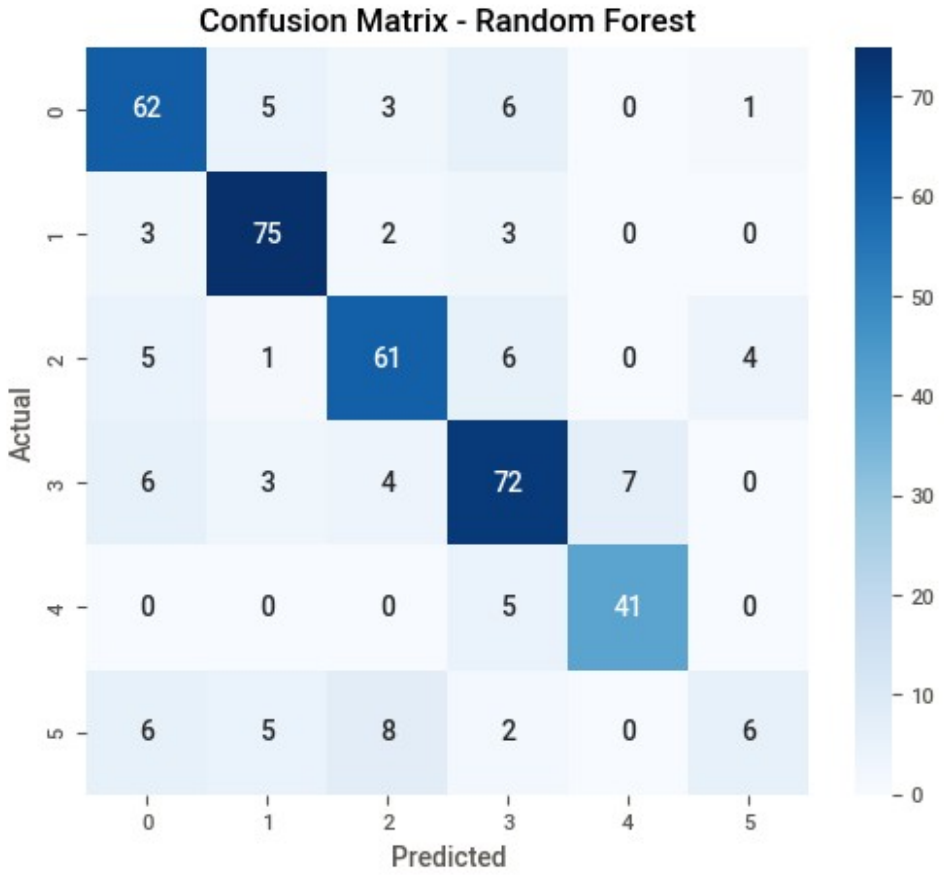
```
        'class_weight': ['balanced', None]
    }
    svm_clf = GridSearchCV(SVC(), svm_params, cv=3, verbose=1,
n_jobs=-1)
    svm_clf.fit(X_train, y_train)
    print(f"Best SVM Parameters for {name}: {svm_clf.best_params_}")
    results[f"SVM_{name}"] =
train_and_evaluate_model(svm_clf.best_estimator_, "SVM", X_train,
X_test, y_train, y_test)

    # Bagging SVM
    bagging_svm =
BaggingClassifier(base_estimator=SVC(**svm_clf.best_params_),
n_estimators=10, random_state=42)

    results[f"BaggingSVM_{name}"] =
train_and_evaluate_model(bagging_svm, "Bagging SVM", X_train, X_test,
y_train, y_test)

    # Logistic Regression
    log_reg = LogisticRegression(max_iter=500)
    results[f"LogisticRegression_{name}"] =
train_and_evaluate_model(log_reg, "Logistic Regression", X_train,
X_test, y_train, y_test)
```

```
===== Training Models on NMF Dataset =====
Training set: (1607, 138), Test set: (402, 138)

/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/
dense.py:87: UserWarning: Do not pass an `input_shape`/`input_dim`
argument to a layer. When using Sequential models, prefer using an
`Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer,
**kwargs)
```



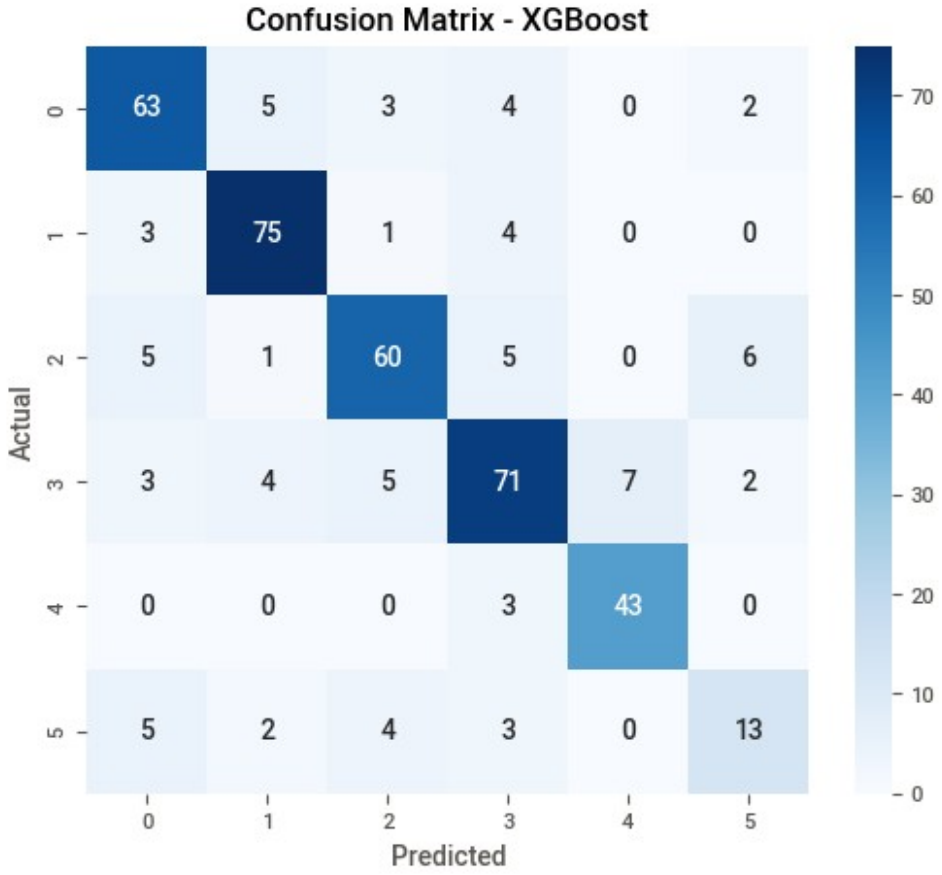

```
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 9ms/step

Neural Network - NMF Classification Report:
              precision    recall  f1-score   support

           0       0.76      0.87      0.81        77
           1       0.89      0.92      0.90        83
           2       0.81      0.75      0.78        77
           3       0.81      0.75      0.78        92
           4       0.83      0.93      0.88        46
           5       0.65      0.48      0.55        27

    accuracy                           0.81       402
   macro avg       0.79      0.78      0.78       402
weighted avg       0.81      0.81      0.81       402


Random Forest Accuracy: 0.7886
              precision    recall  f1-score   support

          C1       0.76      0.81      0.78        77
          C2       0.84      0.90      0.87        83
          C3       0.78      0.79      0.79        77
          C4       0.77      0.78      0.77        92
          C5       0.85      0.89      0.87        46
```

```
              C6       0.55      0.22      0.32        27

    accuracy                           0.79       402
   macro avg       0.76      0.73      0.73       402
weighted avg       0.78      0.79      0.78       402
```

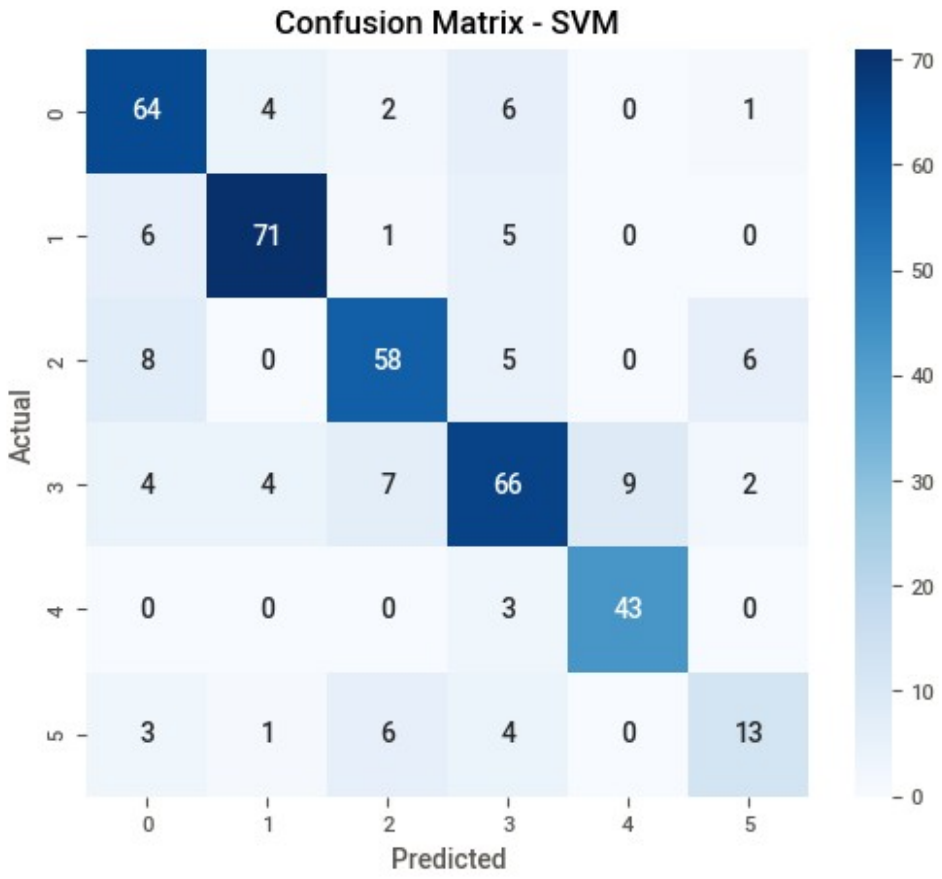



```
Best XGBoost Parameters for NMF: {'learning_rate': 0.1, 'max_depth':
2, 'n_estimators': 200}

XGBoost Accuracy: 0.8085
              precision    recall  f1-score   support

          C1       0.80      0.82      0.81        77
          C2       0.86      0.90      0.88        83
          C3       0.82      0.78      0.80        77
          C4       0.79      0.77      0.78        92
          C5       0.86      0.93      0.90        46
          C6       0.57      0.48      0.52        27

    accuracy                           0.81       402
   macro avg       0.78      0.78      0.78       402
```

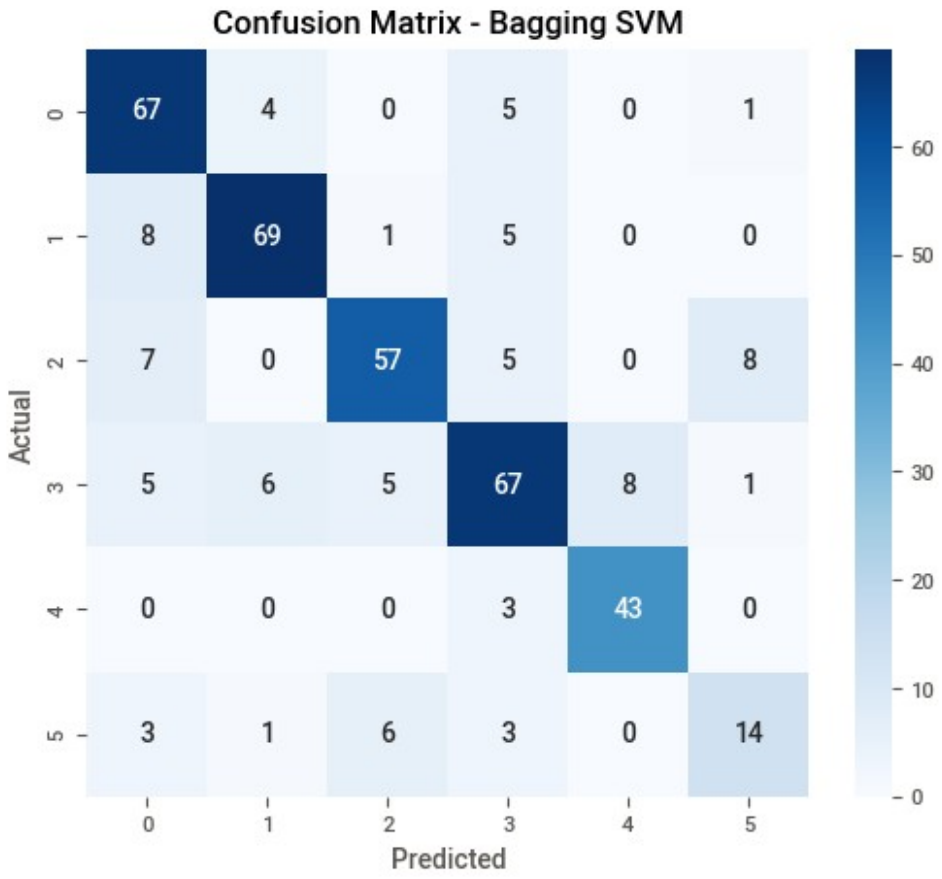
```
weighted avg       0.81      0.81      0.81       402
```



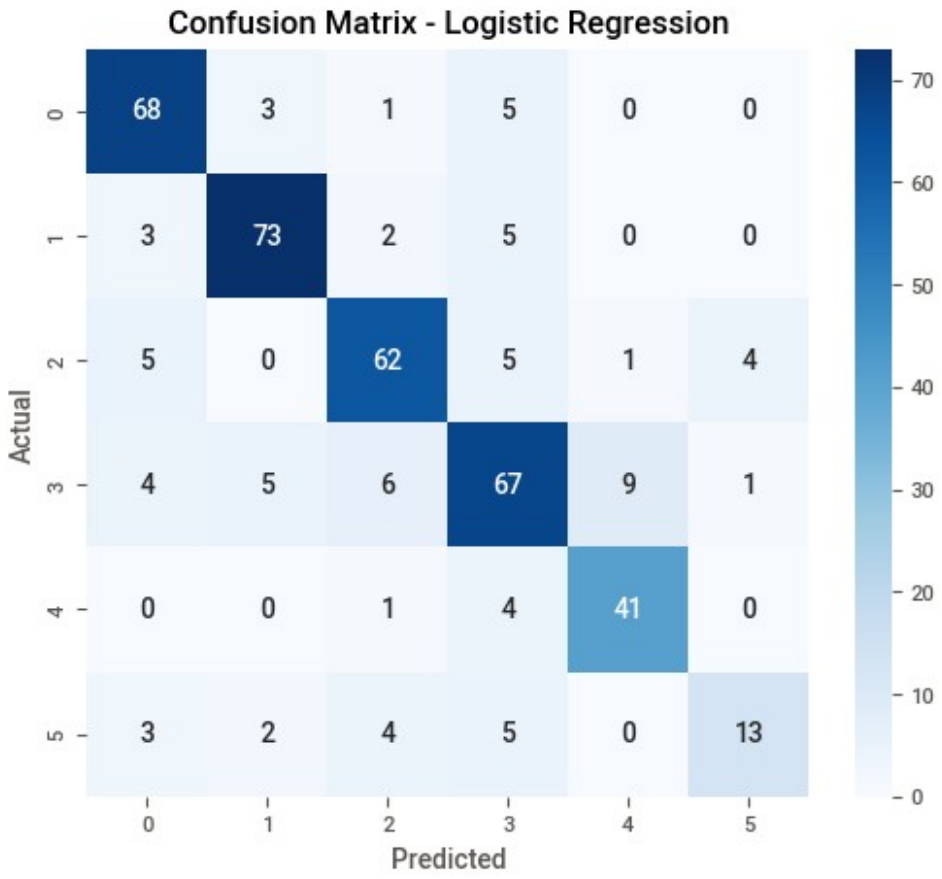

```
Fitting 3 folds for each of 150 candidates, totalling 450 fits
Best SVM Parameters for NMF: {'C': 10, 'class_weight': 'balanced',
'gamma': 'scale', 'kernel': 'rbf'}

SVM Accuracy: 0.7836
              precision    recall  f1-score   support

          C1       0.75      0.83      0.79        77
          C2       0.89      0.86      0.87        83
          C3       0.78      0.75      0.77        77
          C4       0.74      0.72      0.73        92
          C5       0.83      0.93      0.88        46
          C6       0.59      0.48      0.53        27

    accuracy                           0.78       402
   macro avg       0.76      0.76      0.76       402
weighted avg       0.78      0.78      0.78       402
```



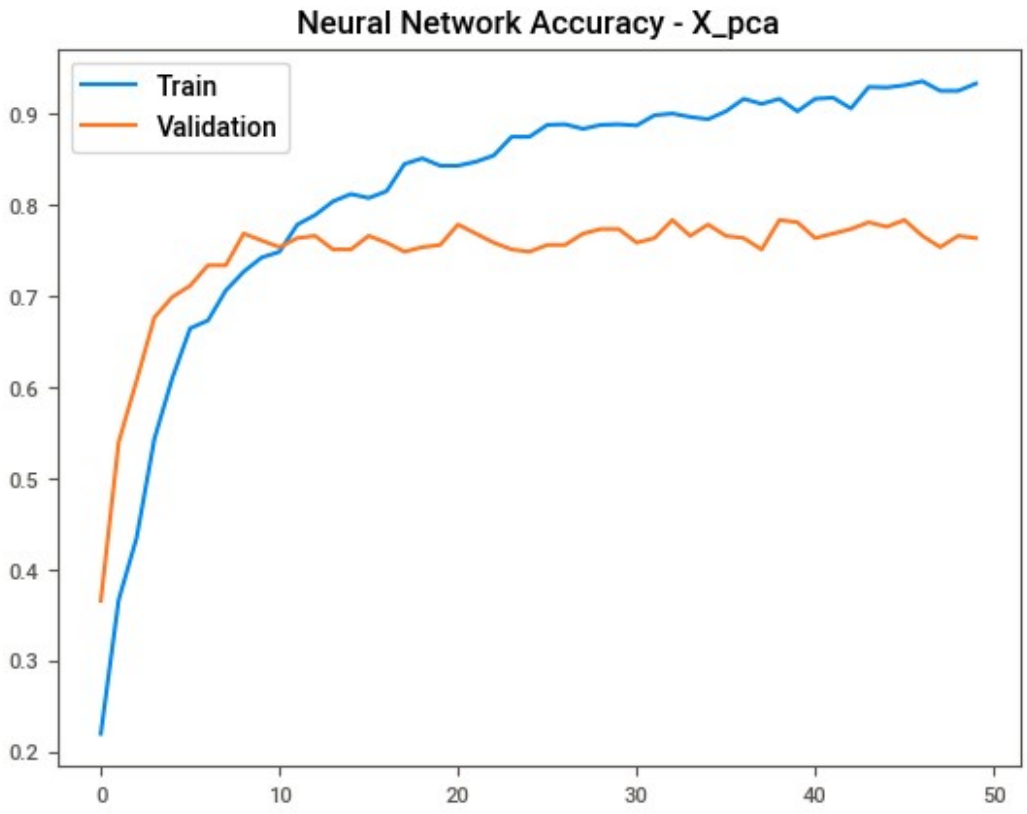

```
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_base.py:166:
FutureWarning: `base_estimator` was renamed to `estimator` in version
1.2 and will be removed in 1.4.
  warnings.warn(


Bagging SVM Accuracy: 0.7886
              precision    recall  f1-score   support

          C1       0.74      0.87      0.80        77
          C2       0.86      0.83      0.85        83
          C3       0.83      0.74      0.78        77
          C4       0.76      0.73      0.74        92
          C5       0.84      0.93      0.89        46
          C6       0.58      0.52      0.55        27

    accuracy                           0.79       402
   macro avg       0.77      0.77      0.77       402
weighted avg       0.79      0.79      0.79       402
```




```
Logistic Regression Accuracy: 0.8060
              precision    recall  f1-score   support

          C1       0.82      0.88      0.85        77
          C2       0.88      0.88      0.88        83
          C3       0.82      0.81      0.81        77
          C4       0.74      0.73      0.73        92
          C5       0.80      0.89      0.85        46
          C6       0.72      0.48      0.58        27

    accuracy                           0.81       402
   macro avg       0.80      0.78      0.78       402
weighted avg       0.80      0.81      0.80       402
```

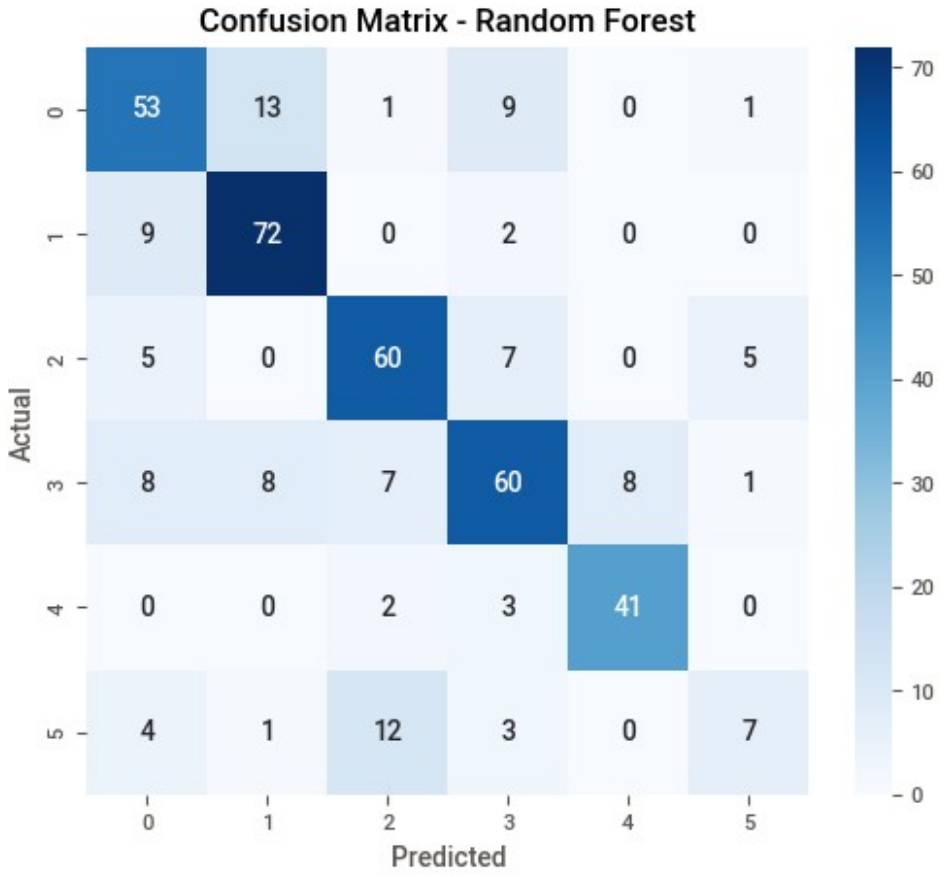



```
===== Training Models on X_pca Dataset =====
Training set: (1607, 94), Test set: (402, 94)

/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/
dense.py:87: UserWarning: Do not pass an `input_shape`/`input_dim`
argument to a layer. When using Sequential models, prefer using an
`Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer,
**kwargs)
```



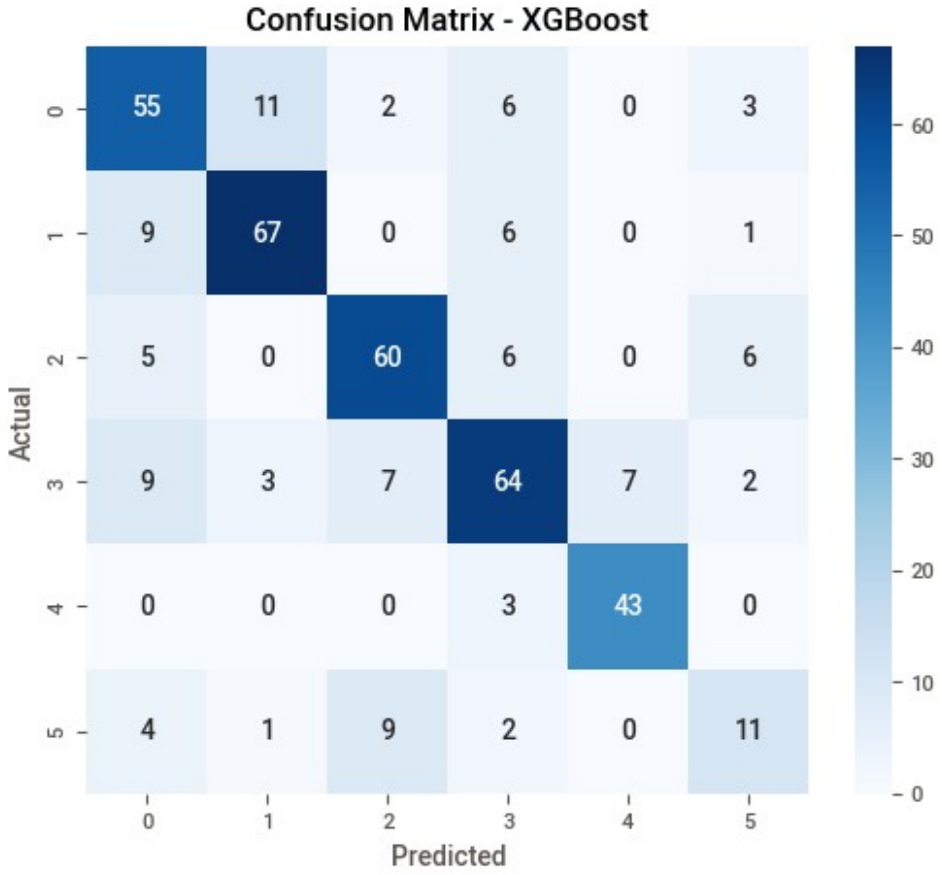

```
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step

Neural Network - X_pca Classification Report:
              precision    recall  f1-score   support

           0       0.74      0.69      0.71        77
           1       0.81      0.87      0.84        83
           2       0.79      0.70      0.74        77
           3       0.69      0.76      0.73        92
           4       0.88      0.91      0.89        46
           5       0.67      0.59      0.63        27

    accuracy                           0.76       402
   macro avg       0.76      0.75      0.76       402
weighted avg       0.76      0.76      0.76       402


Random Forest Accuracy: 0.7289
              precision    recall  f1-score   support

          C1       0.67      0.69      0.68        77
          C2       0.77      0.87      0.81        83
          C3       0.73      0.78      0.75        77
          C4       0.71      0.65      0.68        92
          C5       0.84      0.89      0.86        46
```

```
              C6       0.50      0.26      0.34        27

    accuracy                           0.73       402
   macro avg       0.70      0.69      0.69       402
weighted avg       0.72      0.73      0.72       402
```

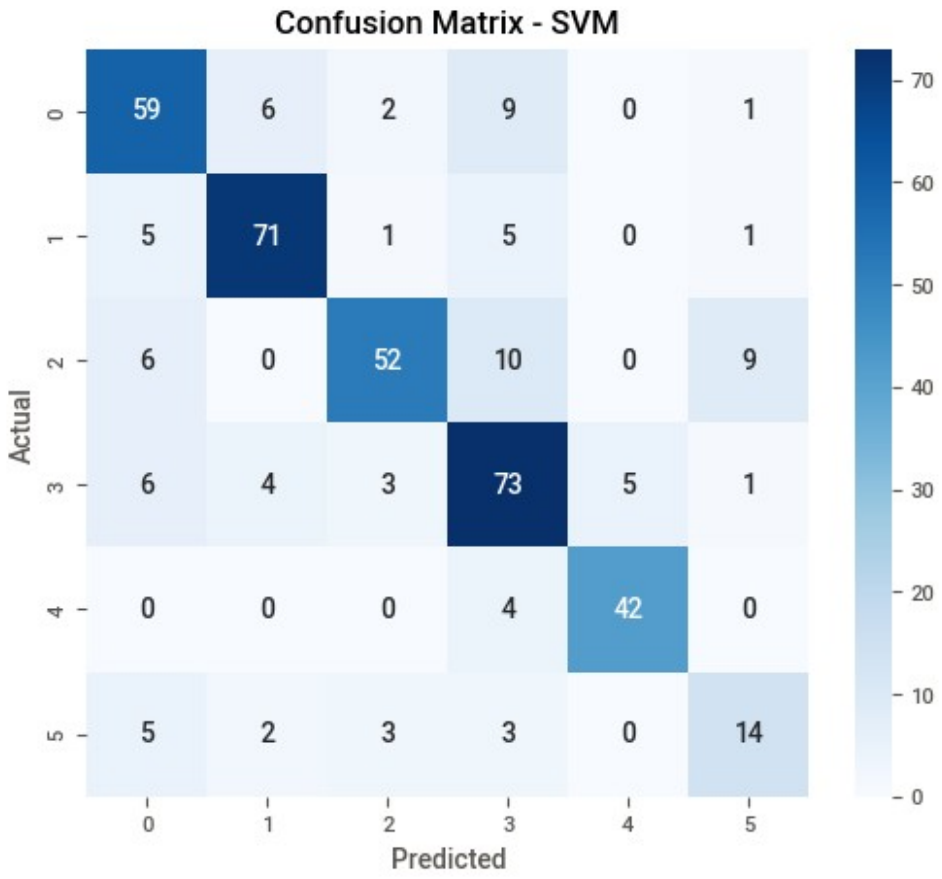



```
Best XGBoost Parameters for X_pca: {'learning_rate': 0.1, 'max_depth':
2, 'n_estimators': 200}

XGBoost Accuracy: 0.7463
              precision    recall  f1-score   support

          C1       0.67      0.71      0.69        77
          C2       0.82      0.81      0.81        83
          C3       0.77      0.78      0.77        77
          C4       0.74      0.70      0.72        92
          C5       0.86      0.93      0.90        46
          C6       0.48      0.41      0.44        27

    accuracy                           0.75       402
   macro avg       0.72      0.72      0.72       402
```

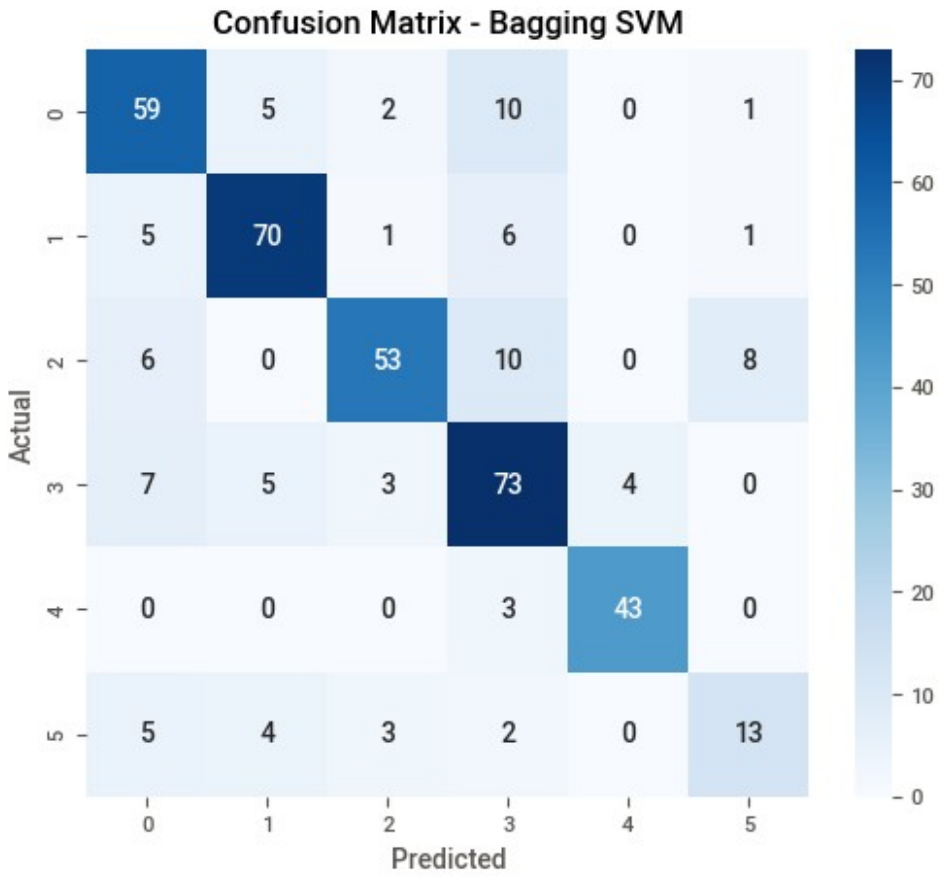
```
weighted avg       0.74      0.75      0.74       402
```



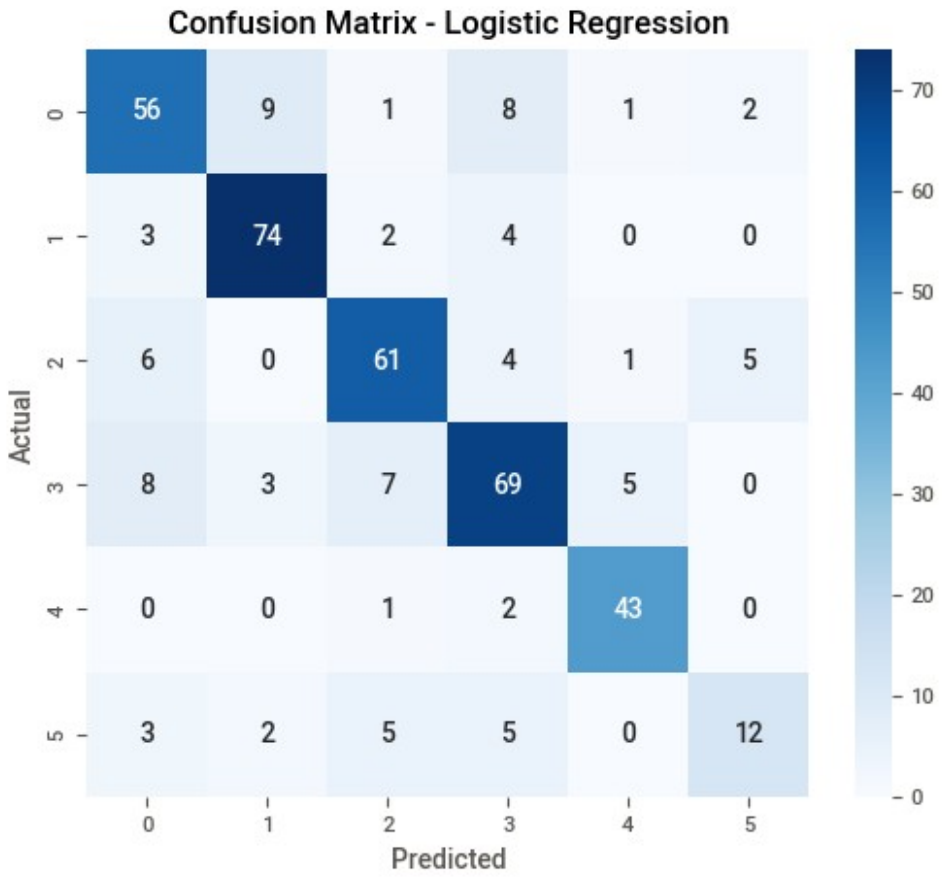

```
Fitting 3 folds for each of 150 candidates, totalling 450 fits
Best SVM Parameters for X_pca: {'C': 10, 'class_weight': None,
'gamma': 0.01, 'kernel': 'rbf'}

SVM Accuracy: 0.7736
              precision    recall  f1-score   support

          C1       0.73      0.77      0.75        77
          C2       0.86      0.86      0.86        83
          C3       0.85      0.68      0.75        77
          C4       0.70      0.79      0.74        92
          C5       0.89      0.91      0.90        46
          C6       0.54      0.52      0.53        27

    accuracy                           0.77       402
   macro avg       0.76      0.75      0.76       402
weighted avg       0.78      0.77      0.77       402
```



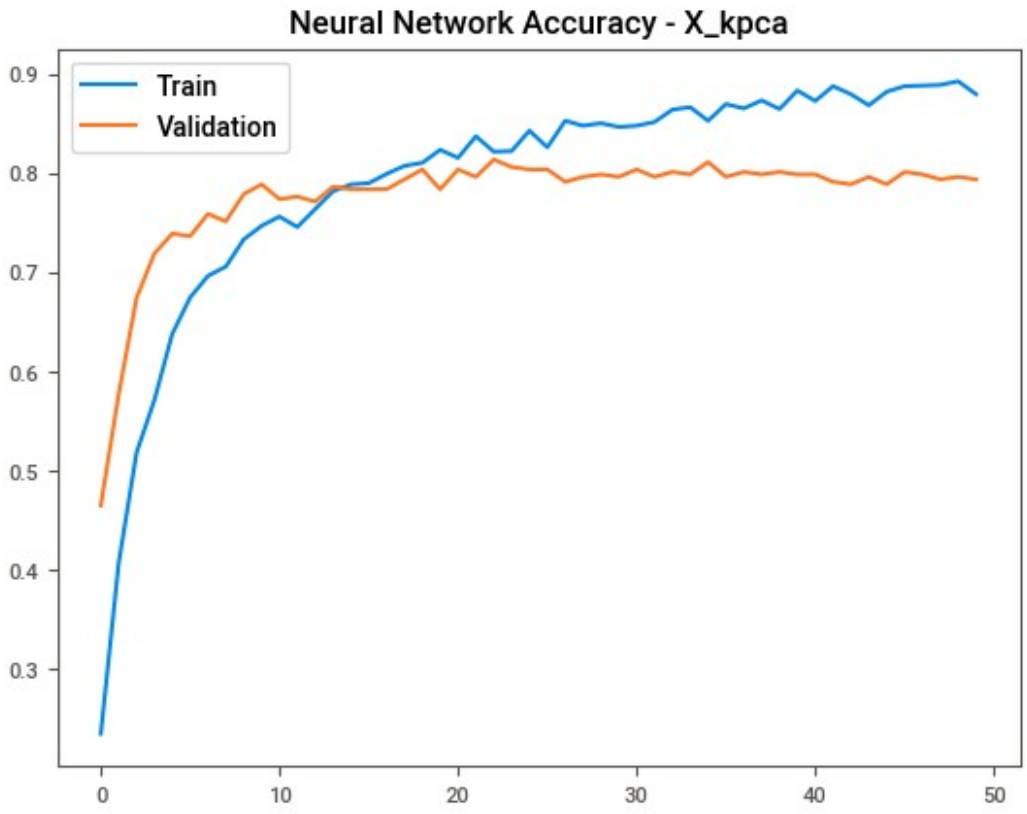

```
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_base.py:166:
FutureWarning: `base_estimator` was renamed to `estimator` in version
1.2 and will be removed in 1.4.
  warnings.warn(


Bagging SVM Accuracy: 0.7736
              precision    recall  f1-score   support

          C1       0.72      0.77      0.74        77
          C2       0.83      0.84      0.84        83
          C3       0.85      0.69      0.76        77
          C4       0.70      0.79      0.74        92
          C5       0.91      0.93      0.92        46
          C6       0.57      0.48      0.52        27

    accuracy                           0.77       402
   macro avg       0.76      0.75      0.76       402
weighted avg       0.78      0.77      0.77       402
```




```
Logistic Regression Accuracy: 0.7836
              precision    recall  f1-score   support

          C1       0.74      0.73      0.73        77
          C2       0.84      0.89      0.87        83
          C3       0.79      0.79      0.79        77
          C4       0.75      0.75      0.75        92
          C5       0.86      0.93      0.90        46
          C6       0.63      0.44      0.52        27

    accuracy                           0.78       402
   macro avg       0.77      0.76      0.76       402
weighted avg       0.78      0.78      0.78       402
```

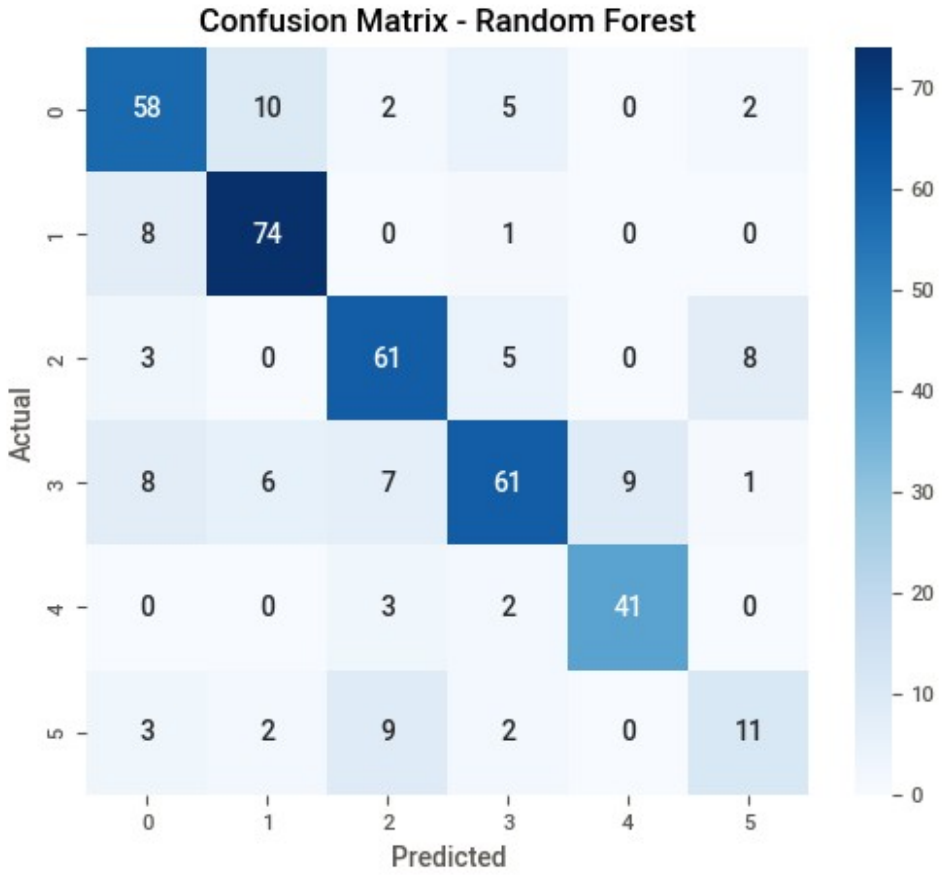



```
===== Training Models on X_kpca Dataset =====
Training set: (1607, 50), Test set: (402, 50)

/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/
dense.py:87: UserWarning: Do not pass an `input_shape`/`input_dim`
argument to a layer. When using Sequential models, prefer using an
`Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer,
**kwargs)
```



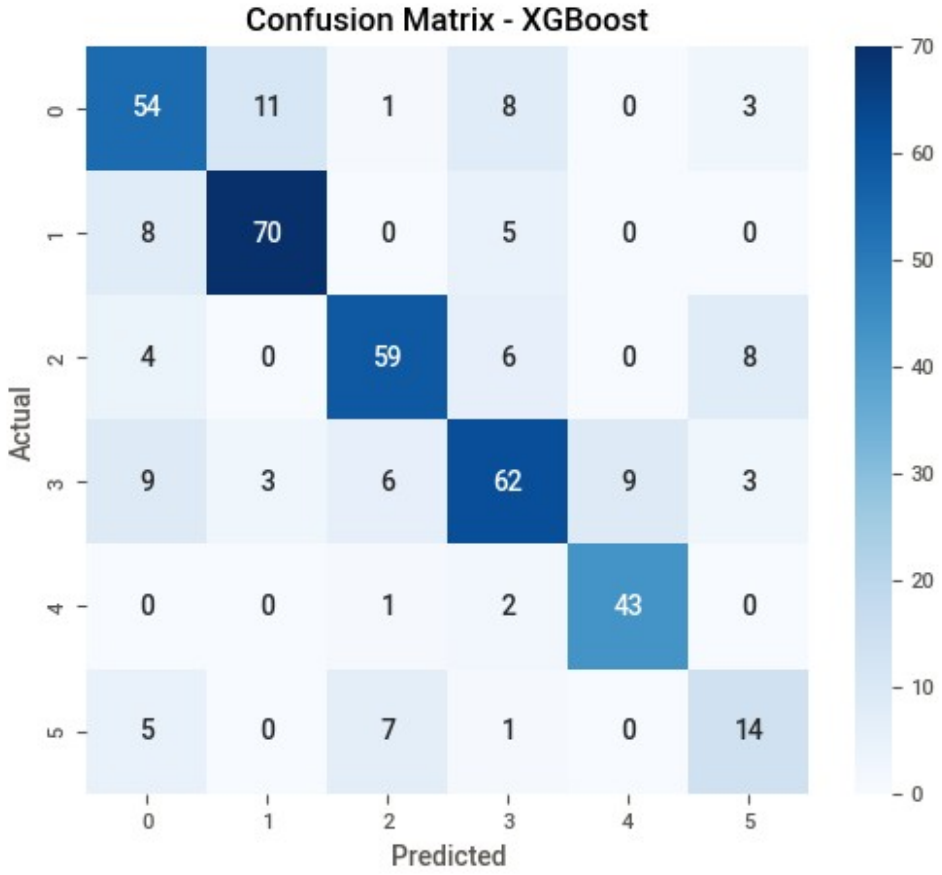

```
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 10ms/step

Neural Network - X_kpca Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.82      0.81        77
           1       0.89      0.86      0.87        83
           2       0.84      0.74      0.79        77
           3       0.71      0.76      0.73        92
           4       0.81      0.93      0.87        46
           5       0.62      0.56      0.59        27

    accuracy                           0.79       402
   macro avg       0.78      0.78      0.78       402
weighted avg       0.80      0.79      0.79       402


Random Forest Accuracy: 0.7612
              precision    recall  f1-score   support

          C1       0.72      0.75      0.74        77
          C2       0.80      0.89      0.85        83
          C3       0.74      0.79      0.77        77
          C4       0.80      0.66      0.73        92
          C5       0.82      0.89      0.85        46
```

```
              C6       0.50      0.41      0.45        27

    accuracy                           0.76       402
   macro avg       0.73      0.73      0.73       402
weighted avg       0.76      0.76      0.76       402
```

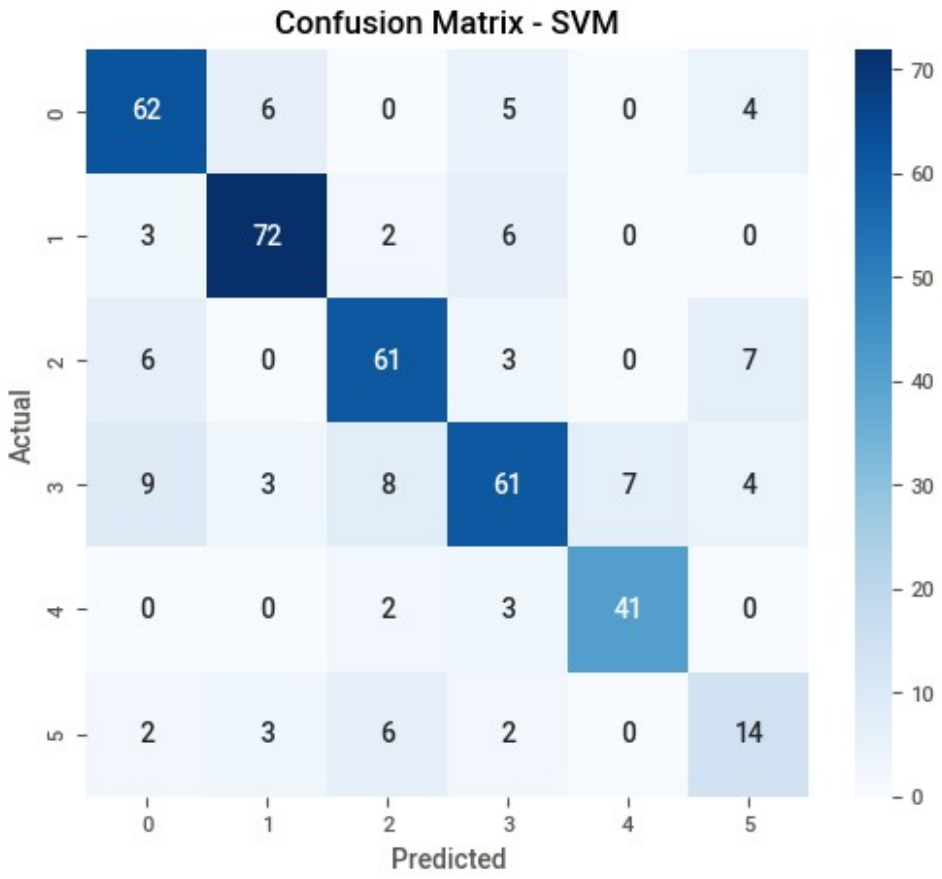



```
Best XGBoost Parameters for X_kpca: {'learning_rate': 0.1,
'max_depth': 2, 'n_estimators': 200}

XGBoost Accuracy: 0.7512
              precision    recall  f1-score   support

          C1       0.68      0.70      0.69        77
          C2       0.83      0.84      0.84        83
          C3       0.80      0.77      0.78        77
          C4       0.74      0.67      0.70        92
          C5       0.83      0.93      0.88        46
          C6       0.50      0.52      0.51        27

    accuracy                           0.75       402
   macro avg       0.73      0.74      0.73       402
```

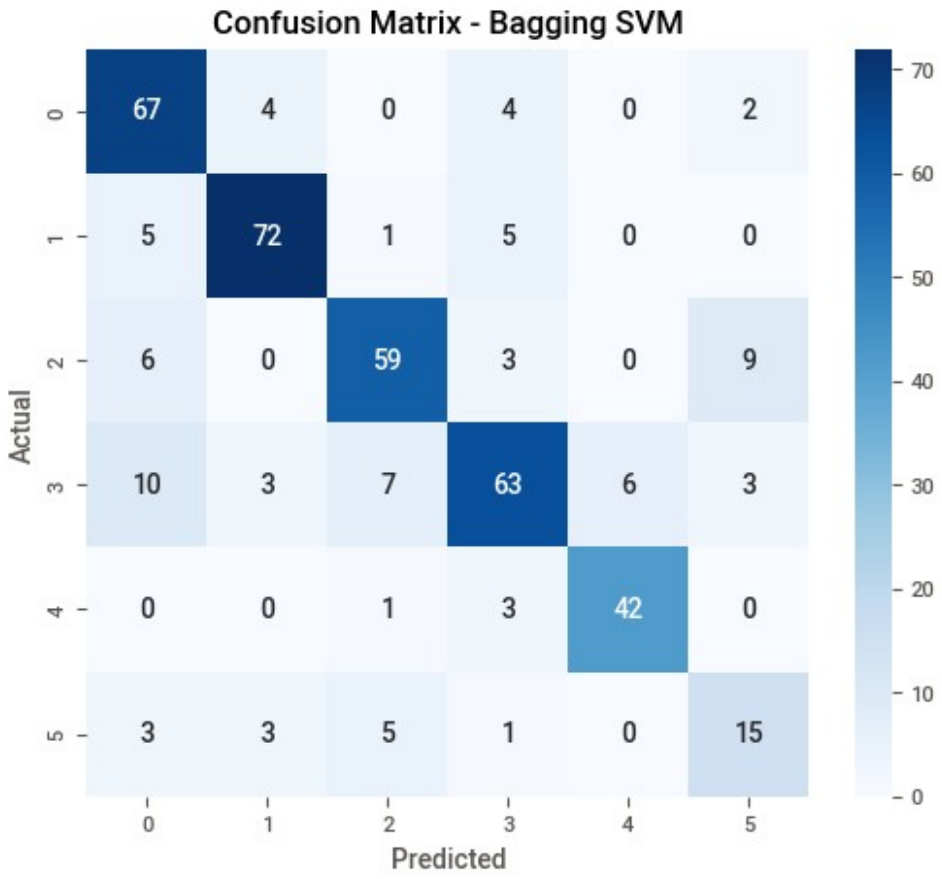
```
weighted avg       0.75      0.75      0.75       402
```



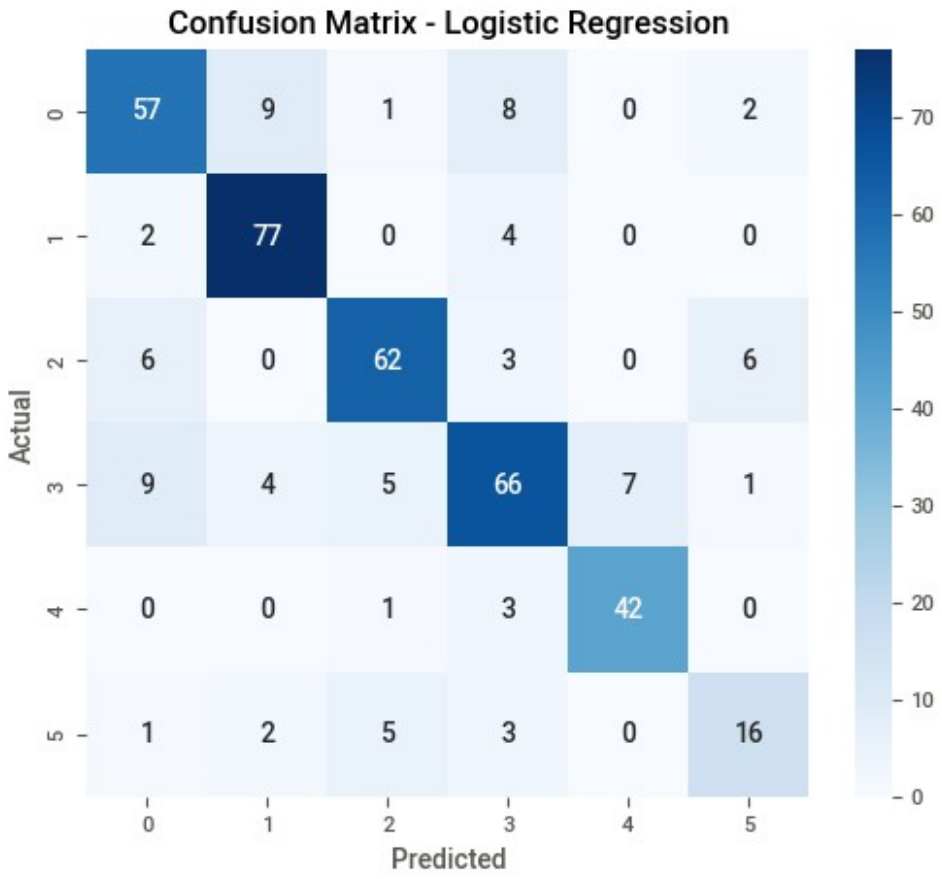

```
Fitting 3 folds for each of 150 candidates, totalling 450 fits
Best SVM Parameters for X_kpca: {'C': 50, 'class_weight': None,
'gamma': 0.001, 'kernel': 'rbf'}

SVM Accuracy: 0.7736
              precision    recall  f1-score   support

          C1       0.76      0.81      0.78        77
          C2       0.86      0.87      0.86        83
          C3       0.77      0.79      0.78        77
          C4       0.76      0.66      0.71        92
          C5       0.85      0.89      0.87        46
          C6       0.48      0.52      0.50        27

    accuracy                           0.77       402
   macro avg       0.75      0.76      0.75       402
weighted avg       0.77      0.77      0.77       402
```



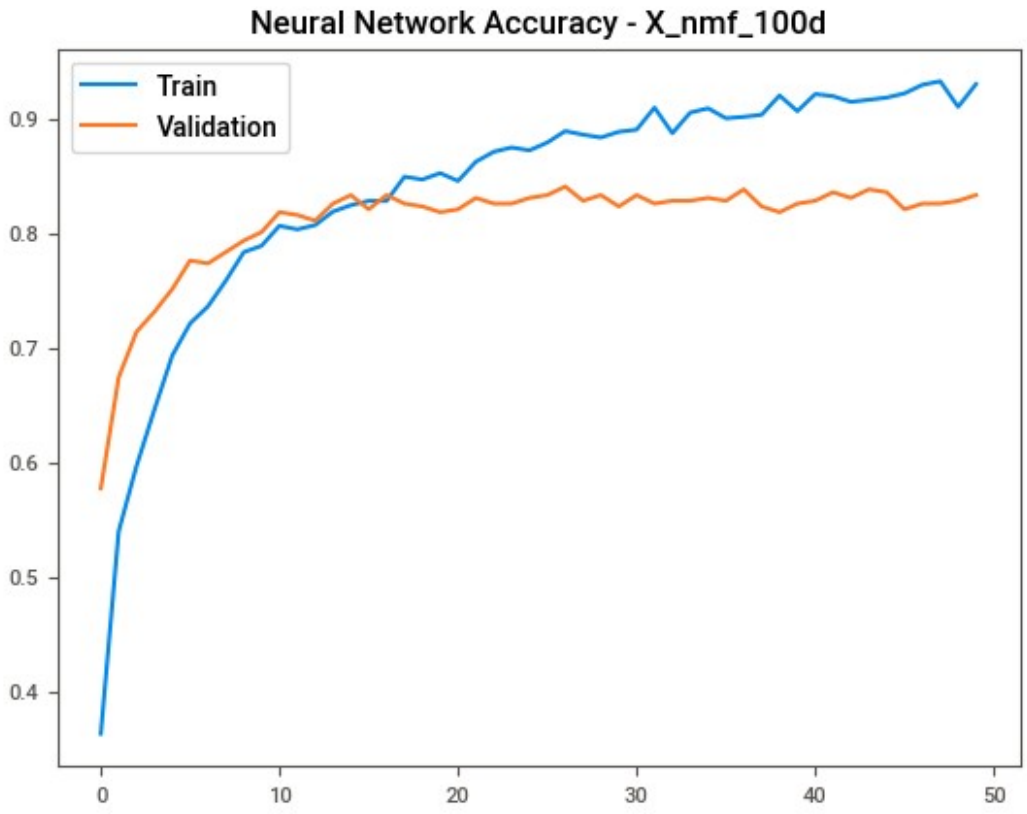

```
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_base.py:166:
FutureWarning: `base_estimator` was renamed to `estimator` in version
1.2 and will be removed in 1.4.
  warnings.warn(


Bagging SVM Accuracy: 0.7910
              precision    recall  f1-score   support

          C1       0.74      0.87      0.80        77
          C2       0.88      0.87      0.87        83
          C3       0.81      0.77      0.79        77
          C4       0.80      0.68      0.74        92
          C5       0.88      0.91      0.89        46
          C6       0.52      0.56      0.54        27

    accuracy                           0.79       402
   macro avg       0.77      0.78      0.77       402
weighted avg       0.79      0.79      0.79       402
```




```
Logistic Regression Accuracy: 0.7960
              precision    recall  f1-score   support

          C1       0.76      0.74      0.75        77
          C2       0.84      0.93      0.88        83
          C3       0.84      0.81      0.82        77
          C4       0.76      0.72      0.74        92
          C5       0.86      0.91      0.88        46
          C6       0.64      0.59      0.62        27

    accuracy                           0.80       402
   macro avg       0.78      0.78      0.78       402
weighted avg       0.79      0.80      0.79       402
```

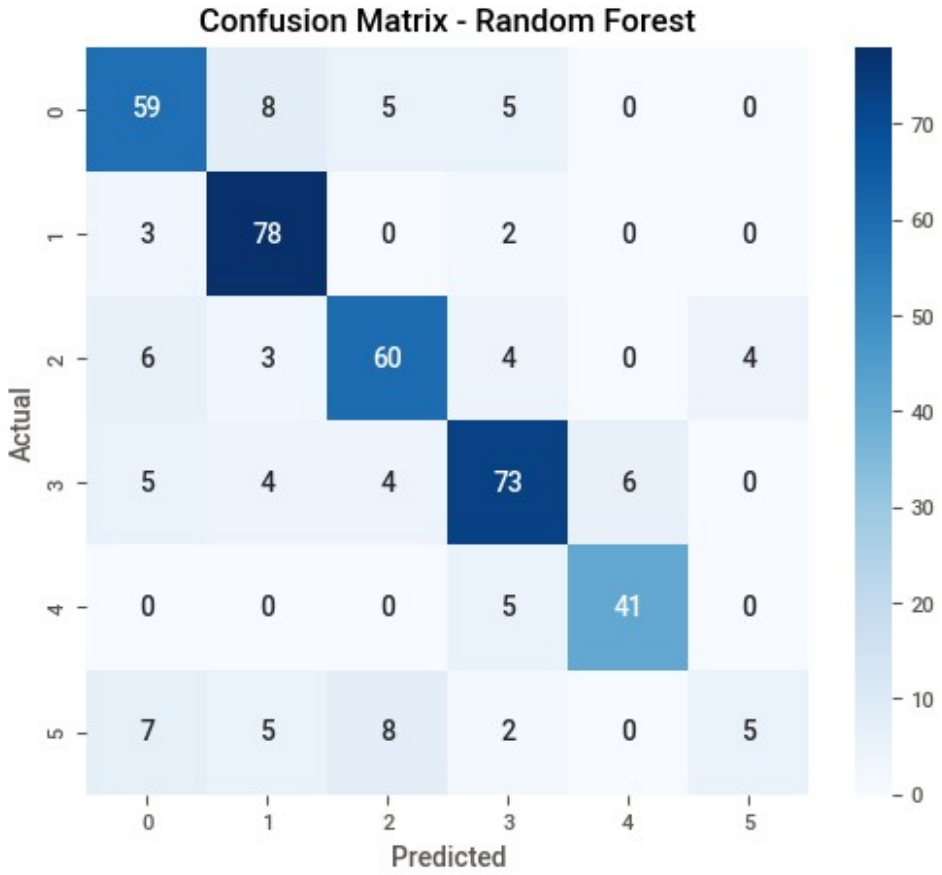



```
===== Training Models on X_nmf_100d Dataset =====
Training set: (1607, 100), Test set: (402, 100)

/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/
dense.py:87: UserWarning: Do not pass an `input_shape`/`input_dim`
argument to a layer. When using Sequential models, prefer using an
`Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer,
**kwargs)
```



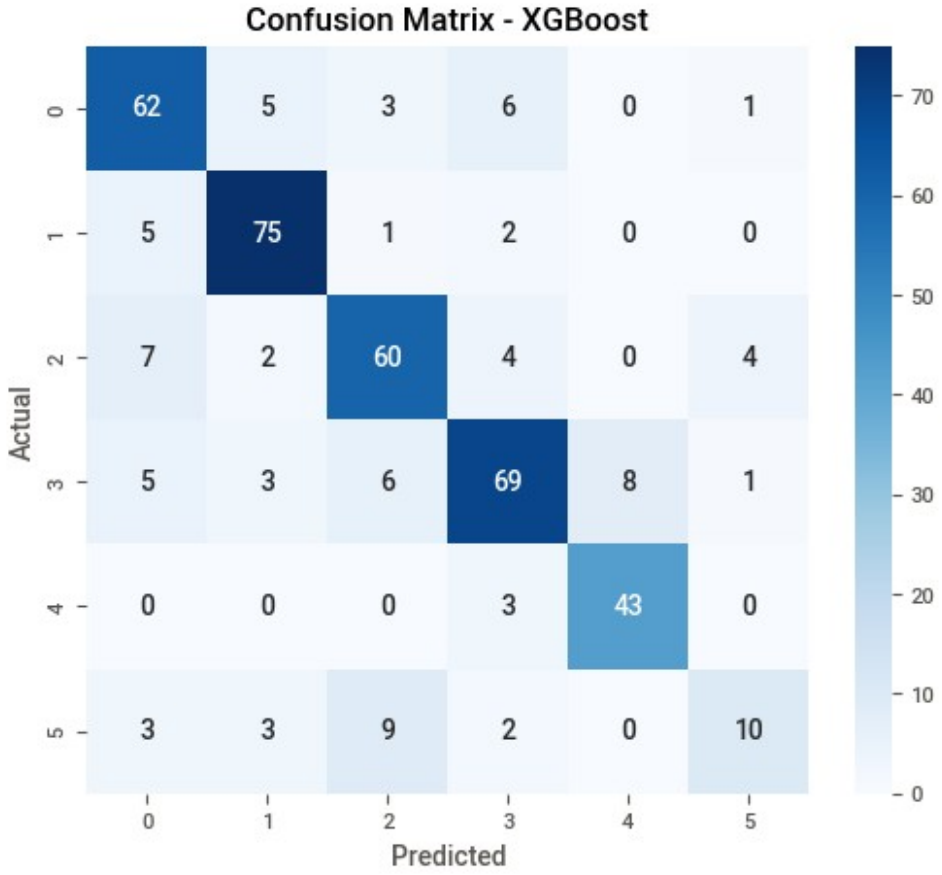

```
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 9ms/step

Neural Network - X_nmf_100d Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.78      0.79        77
           1       0.94      0.89      0.91        83
           2       0.85      0.87      0.86        77
           3       0.75      0.83      0.79        92
           4       0.84      0.93      0.89        46
           5       0.83      0.56      0.67        27

    accuracy                           0.83       402
   macro avg       0.84      0.81      0.82       402
weighted avg       0.84      0.83      0.83       402


Random Forest Accuracy: 0.7861
              precision    recall  f1-score   support

          C1       0.74      0.77      0.75        77
          C2       0.80      0.94      0.86        83
          C3       0.78      0.78      0.78        77
          C4       0.80      0.79      0.80        92
          C5       0.87      0.89      0.88        46
```

```
              C6       0.56      0.19      0.28        27

    accuracy                           0.79       402
   macro avg       0.76      0.73      0.73       402
weighted avg       0.78      0.79      0.77       402
```

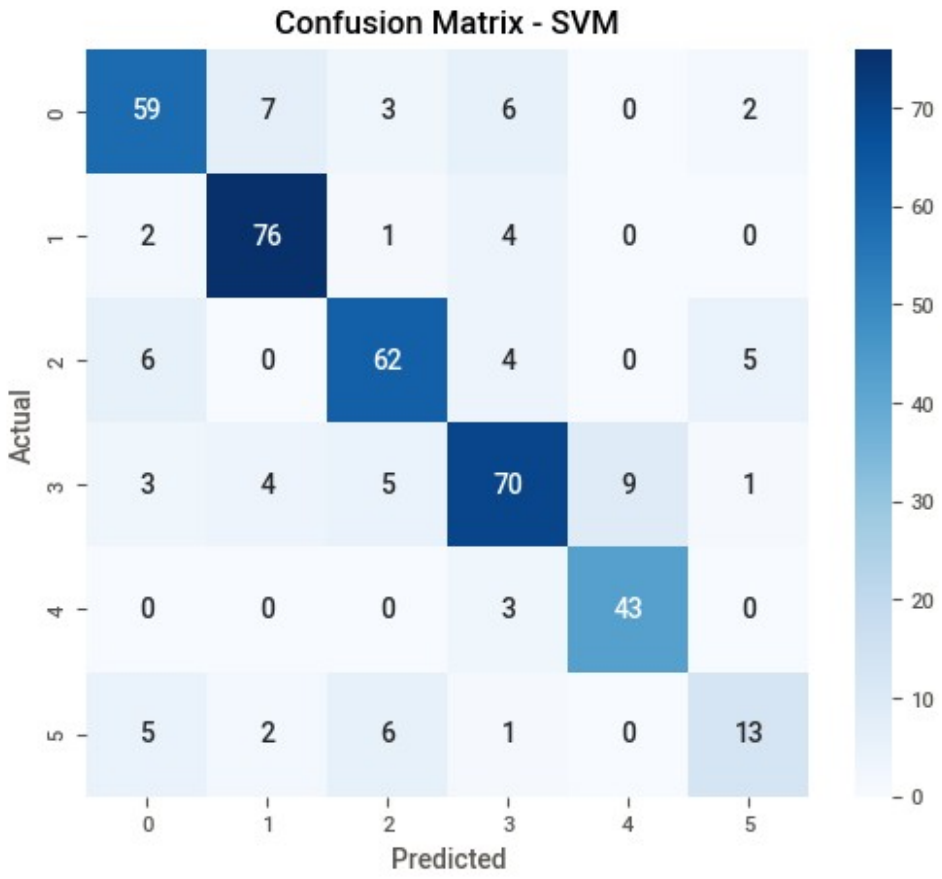



```
Best XGBoost Parameters for X_nmf_100d: {'learning_rate': 0.1,
'max_depth': 2, 'n_estimators': 200}

XGBoost Accuracy: 0.7935
              precision    recall  f1-score   support

          C1       0.76      0.81      0.78        77
          C2       0.85      0.90      0.88        83
          C3       0.76      0.78      0.77        77
          C4       0.80      0.75      0.78        92
          C5       0.84      0.93      0.89        46
          C6       0.62      0.37      0.47        27

    accuracy                           0.79       402
   macro avg       0.77      0.76      0.76       402
```

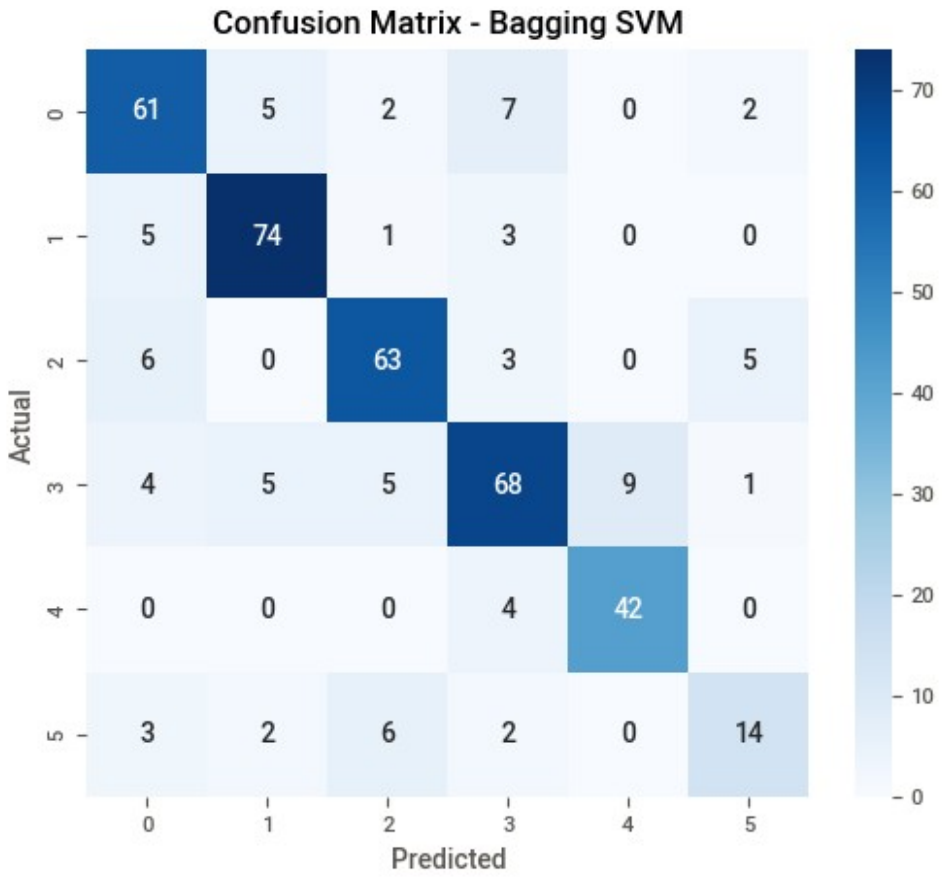
```
weighted avg       0.79      0.79      0.79       402
```



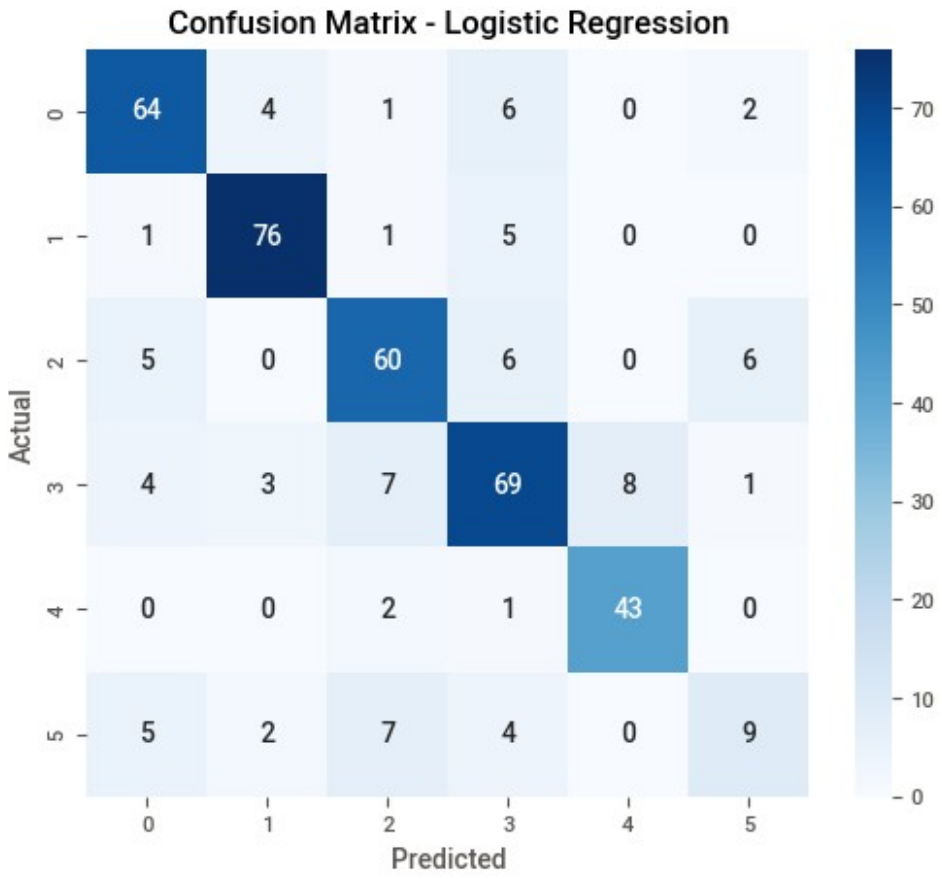

```
Fitting 3 folds for each of 150 candidates, totalling 450 fits
Best SVM Parameters for X_nmf_100d: {'C': 50, 'class_weight':
'balanced', 'gamma': 'auto', 'kernel': 'rbf'}

SVM Accuracy: 0.8035
              precision    recall  f1-score   support

          C1       0.79      0.77      0.78        77
          C2       0.85      0.92      0.88        83
          C3       0.81      0.81      0.81        77
          C4       0.80      0.76      0.78        92
          C5       0.83      0.93      0.88        46
          C6       0.62      0.48      0.54        27

    accuracy                           0.80       402
   macro avg       0.78      0.78      0.78       402
weighted avg       0.80      0.80      0.80       402
```



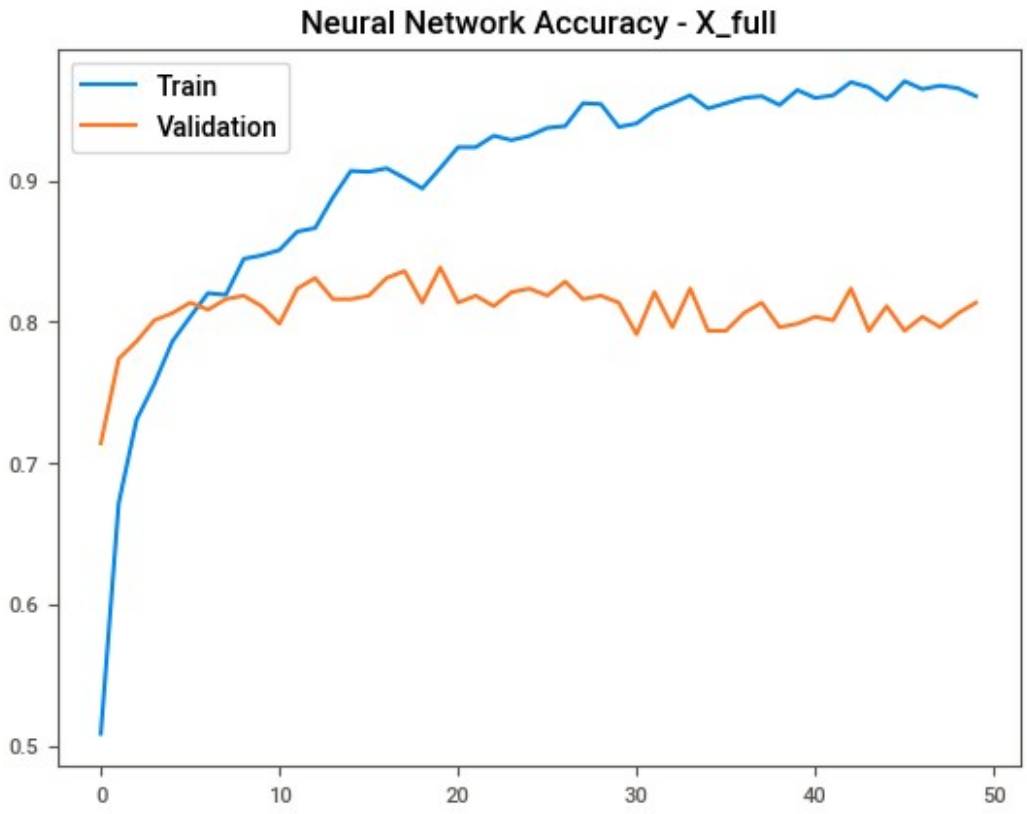

```
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_base.py:166:
FutureWarning: `base_estimator` was renamed to `estimator` in version
1.2 and will be removed in 1.4.
  warnings.warn(


Bagging SVM Accuracy: 0.8010
              precision    recall  f1-score   support

          C1       0.77      0.79      0.78        77
          C2       0.86      0.89      0.88        83
          C3       0.82      0.82      0.82        77
          C4       0.78      0.74      0.76        92
          C5       0.82      0.91      0.87        46
          C6       0.64      0.52      0.57        27

    accuracy                           0.80       402
   macro avg       0.78      0.78      0.78       402
weighted avg       0.80      0.80      0.80       402
```




```
Logistic Regression Accuracy: 0.7985
              precision    recall  f1-score   support

          C1       0.81      0.83      0.82        77
          C2       0.89      0.92      0.90        83
          C3       0.77      0.78      0.77        77
          C4       0.76      0.75      0.75        92
          C5       0.84      0.93      0.89        46
          C6       0.50      0.33      0.40        27

    accuracy                           0.80       402
   macro avg       0.76      0.76      0.76       402
weighted avg       0.79      0.80      0.79       402
```

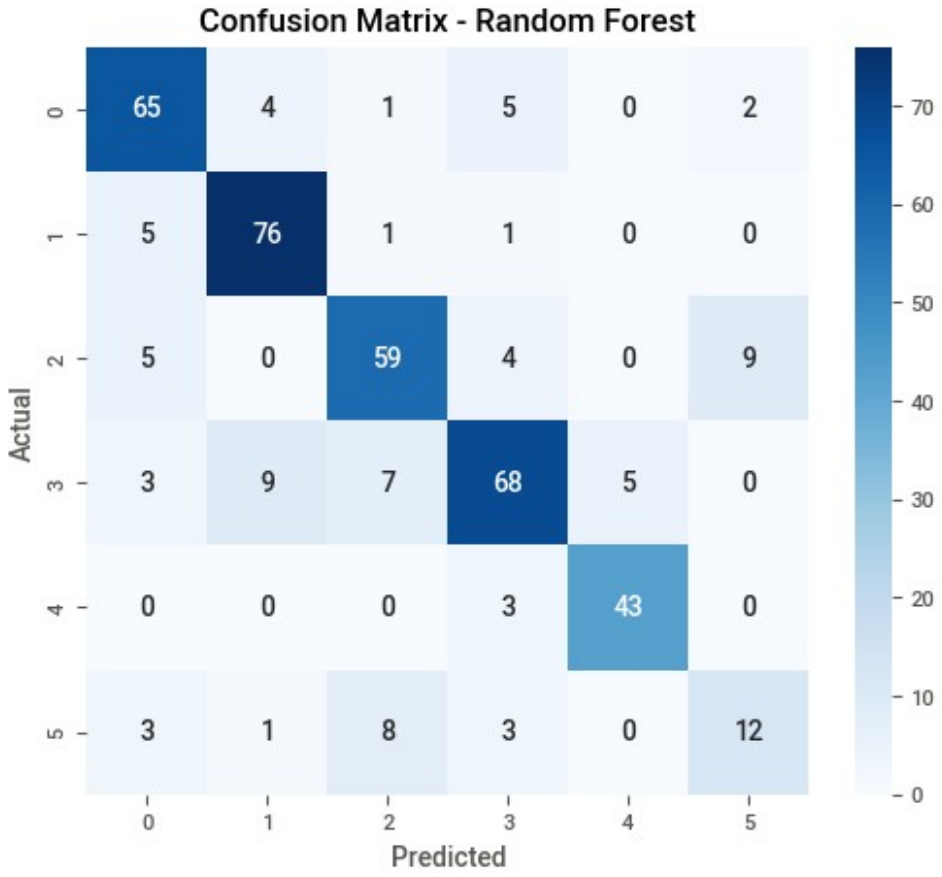



```
===== Training Models on X_full Dataset =====
Training set: (1607, 440), Test set: (402, 440)

/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/
dense.py:87: UserWarning: Do not pass an `input_shape`/`input_dim`
argument to a layer. When using Sequential models, prefer using an
`Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer,
**kwargs)
```



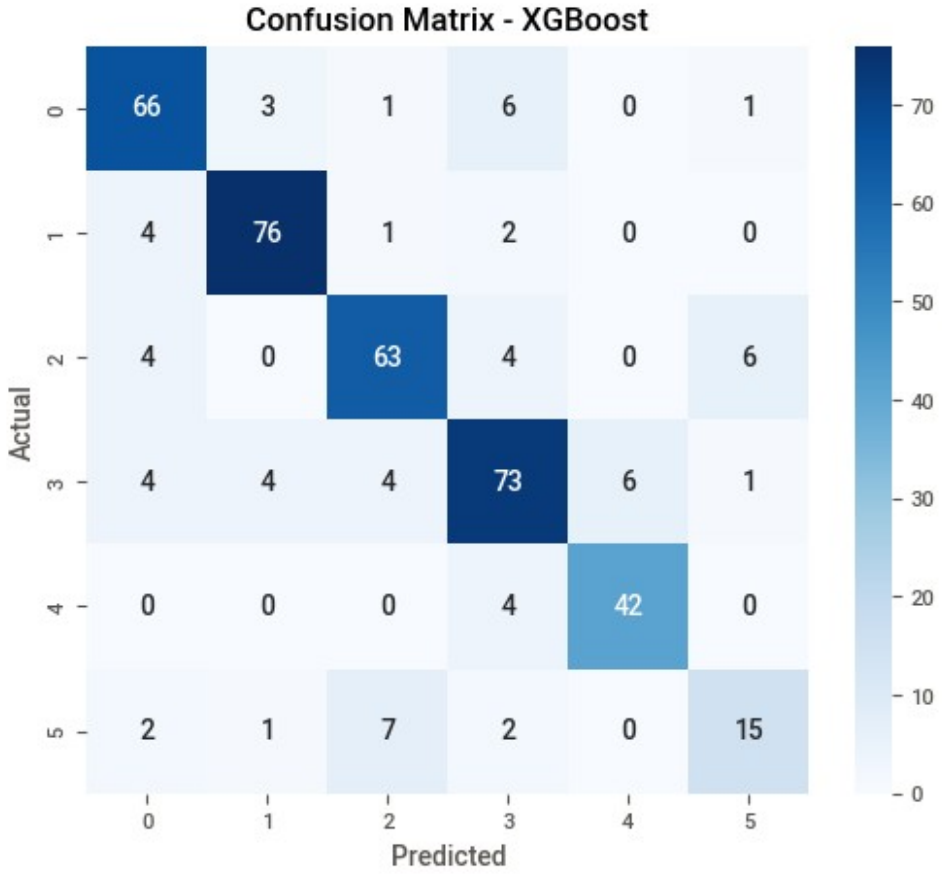

```
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step

Neural Network - X_full Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.86      0.83        77
           1       0.88      0.86      0.87        83
           2       0.83      0.81      0.82        77
           3       0.78      0.79      0.79        92
           4       0.84      0.93      0.89        46
           5       0.60      0.44      0.51        27

    accuracy                           0.81       402
   macro avg       0.79      0.78      0.78       402
weighted avg       0.81      0.81      0.81       402


Random Forest Accuracy: 0.8035
              precision    recall  f1-score   support

          C1       0.80      0.84      0.82        77
          C2       0.84      0.92      0.88        83
          C3       0.78      0.77      0.77        77
          C4       0.81      0.74      0.77        92
          C5       0.90      0.93      0.91        46
```

```
              C6       0.52      0.44      0.48        27

    accuracy                           0.80       402
   macro avg       0.78      0.77      0.77       402
weighted avg       0.80      0.80      0.80       402
```

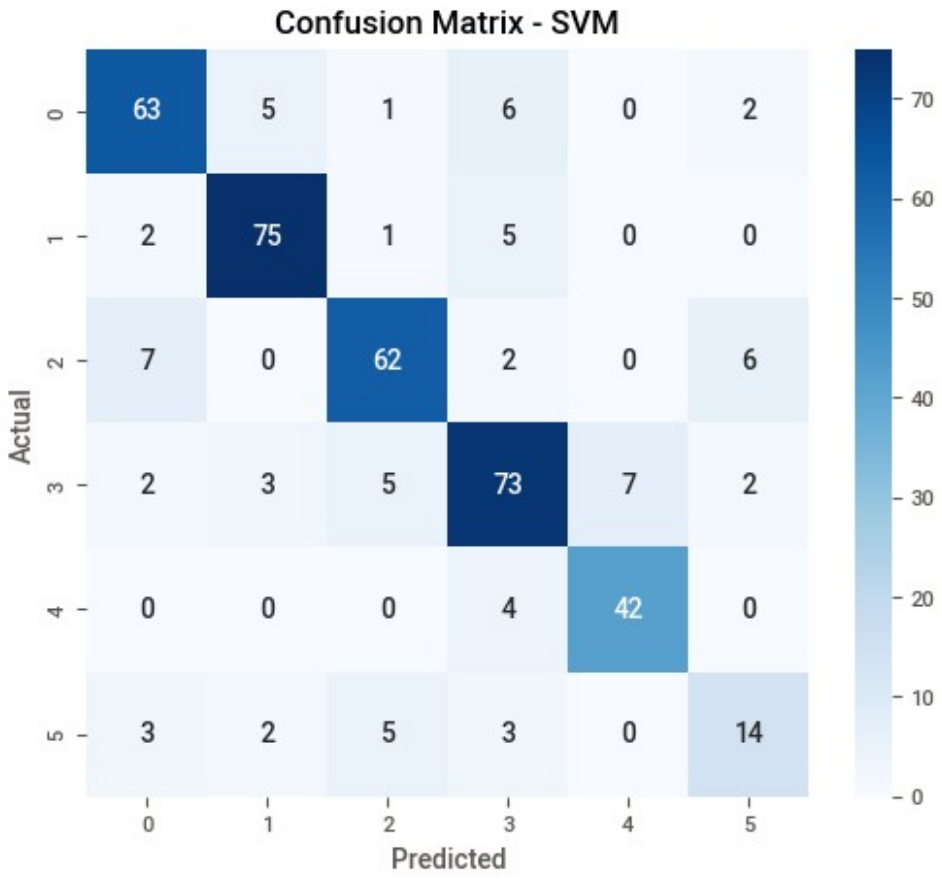



```
Best XGBoost Parameters for X_full: {'learning_rate': 0.1,
'max_depth': 2, 'n_estimators': 200}

XGBoost Accuracy: 0.8333
              precision    recall  f1-score   support

          C1       0.82      0.86      0.84        77
          C2       0.90      0.92      0.91        83
          C3       0.83      0.82      0.82        77
          C4       0.80      0.79      0.80        92
          C5       0.88      0.91      0.89        46
          C6       0.65      0.56      0.60        27

    accuracy                           0.83       402
   macro avg       0.81      0.81      0.81       402
```

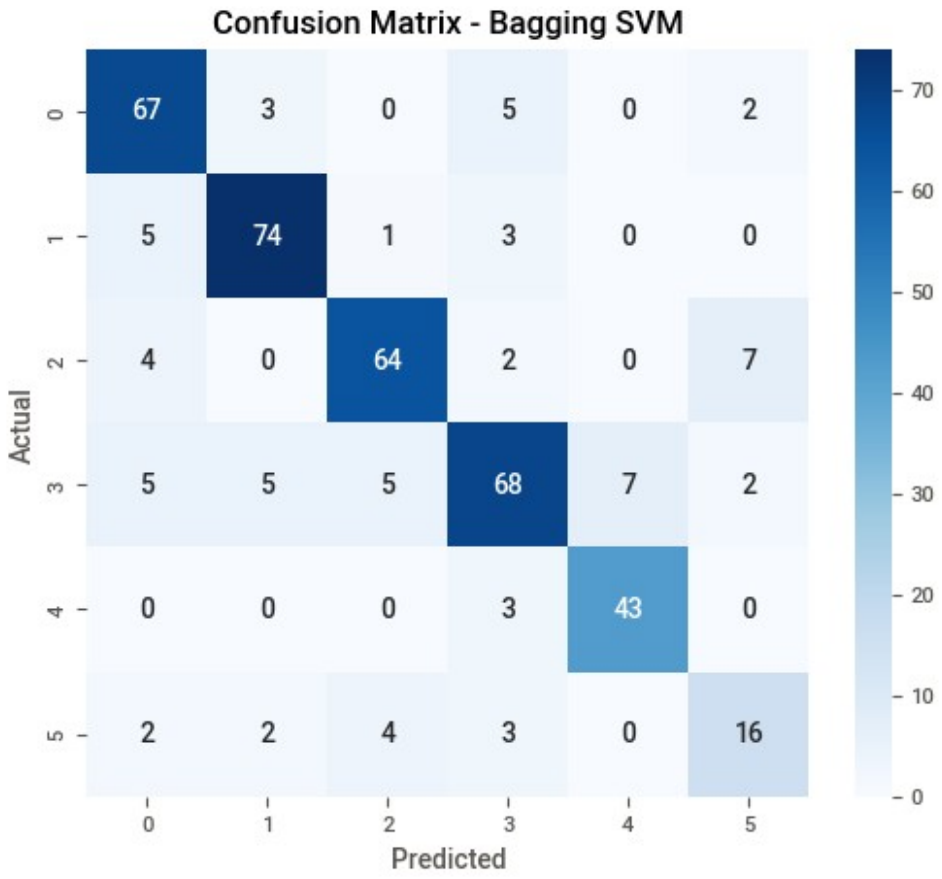
```
weighted avg       0.83      0.83      0.83       402
```




```
Fitting 3 folds for each of 150 candidates, totalling 450 fits
Best SVM Parameters for X_full: {'C': 10, 'class_weight': 'balanced',
'gamma': 'auto', 'kernel': 'rbf'}

SVM Accuracy: 0.8184
              precision    recall  f1-score   support

          C1       0.82      0.82      0.82        77
          C2       0.88      0.90      0.89        83
          C3       0.84      0.81      0.82        77
          C4       0.78      0.79      0.79        92
          C5       0.86      0.91      0.88        46
          C6       0.58      0.52      0.55        27

    accuracy                           0.82       402
   macro avg       0.79      0.79      0.79       402
weighted avg       0.82      0.82      0.82       402
```

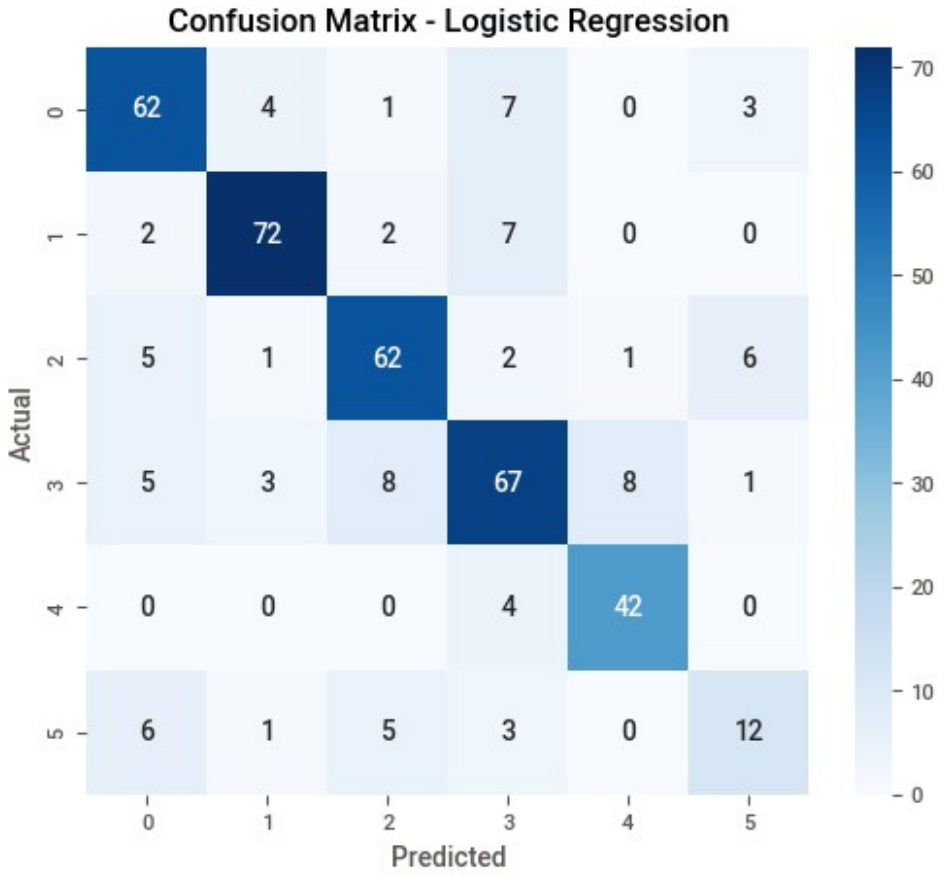



```
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_base.py:166:
FutureWarning: `base_estimator` was renamed to `estimator` in version
1.2 and will be removed in 1.4.
  warnings.warn(


Bagging SVM Accuracy: 0.8259
              precision    recall  f1-score   support

          C1       0.81      0.87      0.84        77
          C2       0.88      0.89      0.89        83
          C3       0.86      0.83      0.85        77
          C4       0.81      0.74      0.77        92
          C5       0.86      0.93      0.90        46
          C6       0.59      0.59      0.59        27

    accuracy                           0.83       402
   macro avg       0.80      0.81      0.81       402
weighted avg       0.83      0.83      0.82       402
```

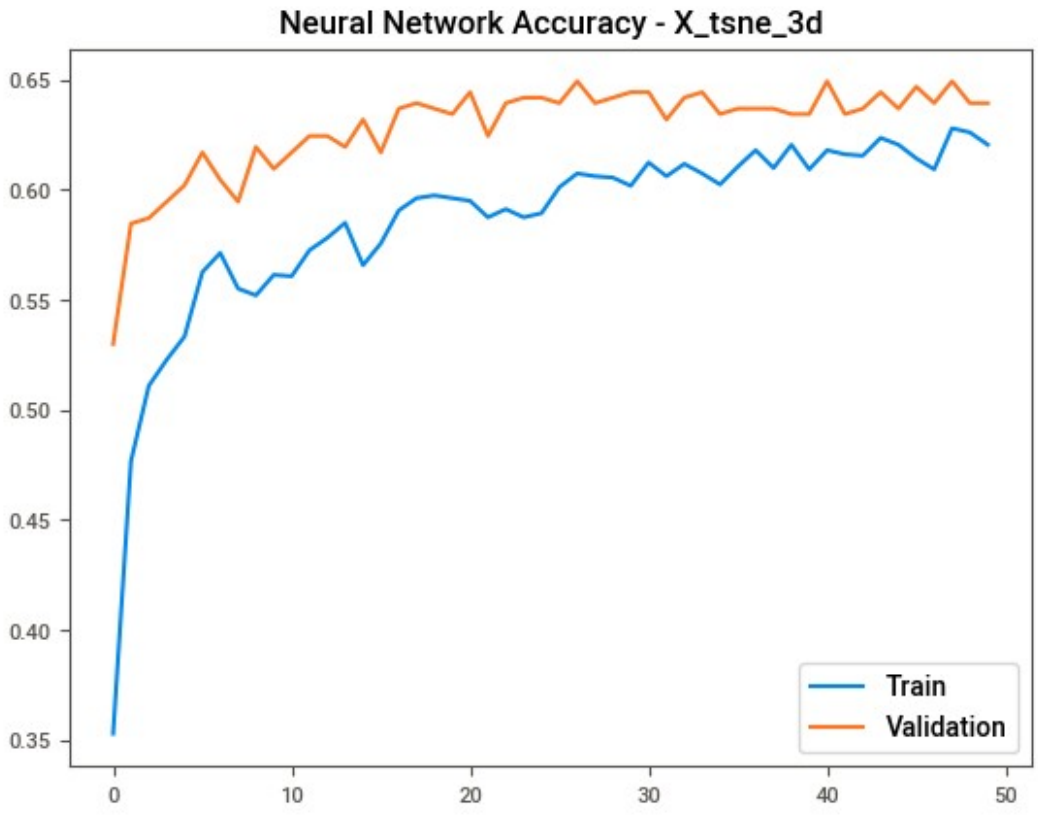



```
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/
_logistic.py:458: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(


Logistic Regression Accuracy: 0.7886
              precision    recall  f1-score   support

          C1       0.78      0.81      0.79        77
          C2       0.89      0.87      0.88        83
          C3       0.79      0.81      0.80        77
          C4       0.74      0.73      0.74        92
          C5       0.82      0.91      0.87        46
```

```
          C6       0.55      0.44      0.49        27

    accuracy                           0.79       402
   macro avg       0.76      0.76      0.76       402
weighted avg       0.79      0.79      0.79       402
```

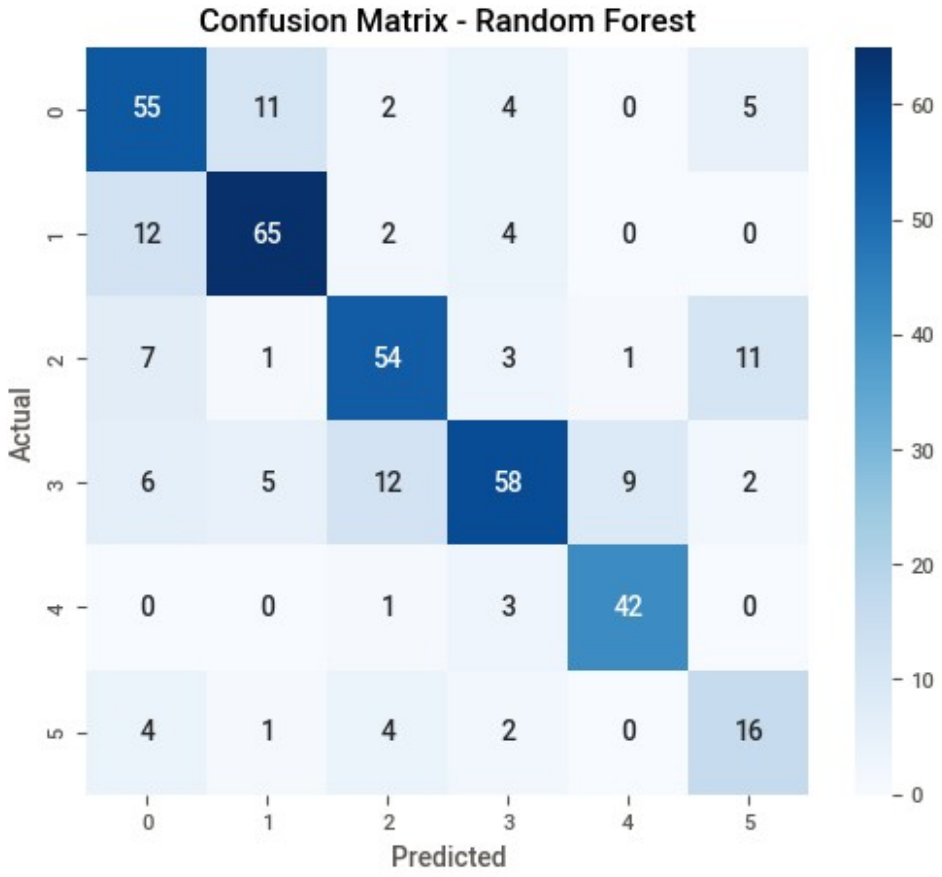



```
===== Training Models on X_tsne_3d Dataset =====
Training set: (1607, 3), Test set: (402, 3)

/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/
dense.py:87: UserWarning: Do not pass an `input_shape`/`input_dim`
argument to a layer. When using Sequential models, prefer using an
`Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer,
**kwargs)
```



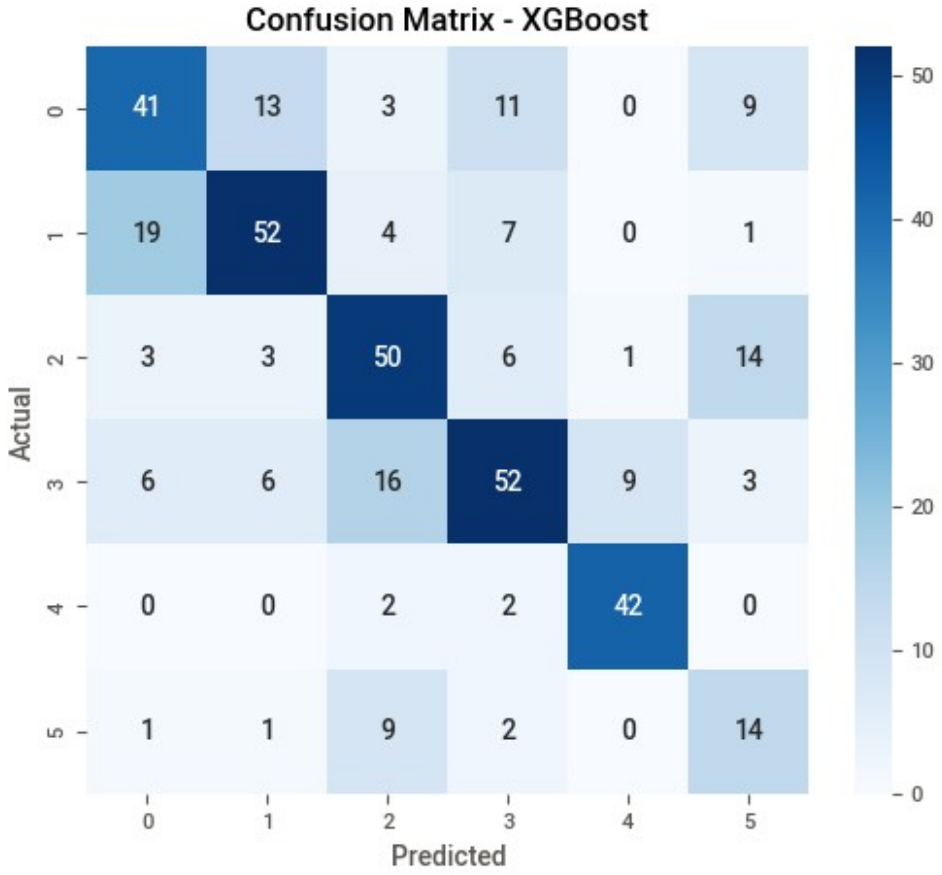

```
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step

Neural Network - X_tsne_3d Classification Report:
              precision    recall  f1-score   support

           0       0.61      0.60      0.61        77
           1       0.73      0.71      0.72        83
           2       0.57      0.61      0.59        77
           3       0.65      0.57      0.60        92
           4       0.84      0.89      0.86        46
           5       0.34      0.44      0.39        27

    accuracy                           0.64       402
   macro avg       0.62      0.64      0.63       402
weighted avg       0.65      0.64      0.64       402


Random Forest Accuracy: 0.7214
              precision    recall  f1-score   support

          C1       0.65      0.71      0.68        77
          C2       0.78      0.78      0.78        83
          C3       0.72      0.70      0.71        77
          C4       0.78      0.63      0.70        92
          C5       0.81      0.91      0.86        46
```

```
              C6       0.47      0.59      0.52        27

    accuracy                           0.72       402
   macro avg       0.70      0.72      0.71       402
weighted avg       0.73      0.72      0.72       402
```

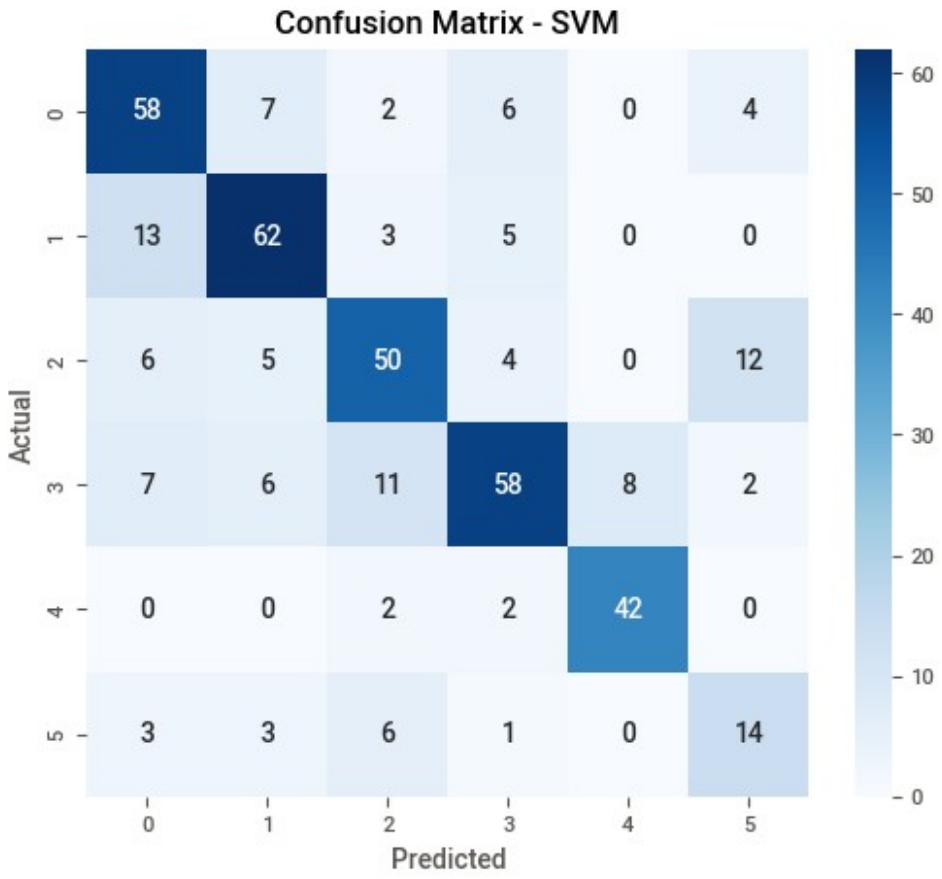



```
Best XGBoost Parameters for X_tsne_3d: {'learning_rate': 0.1,
'max_depth': 2, 'n_estimators': 200}

XGBoost Accuracy: 0.6244
              precision    recall  f1-score   support

          C1       0.59      0.53      0.56        77
          C2       0.69      0.63      0.66        83
          C3       0.60      0.65      0.62        77
          C4       0.65      0.57      0.60        92
          C5       0.81      0.91      0.86        46
          C6       0.34      0.52      0.41        27

    accuracy                           0.62       402
   macro avg       0.61      0.63      0.62       402
```

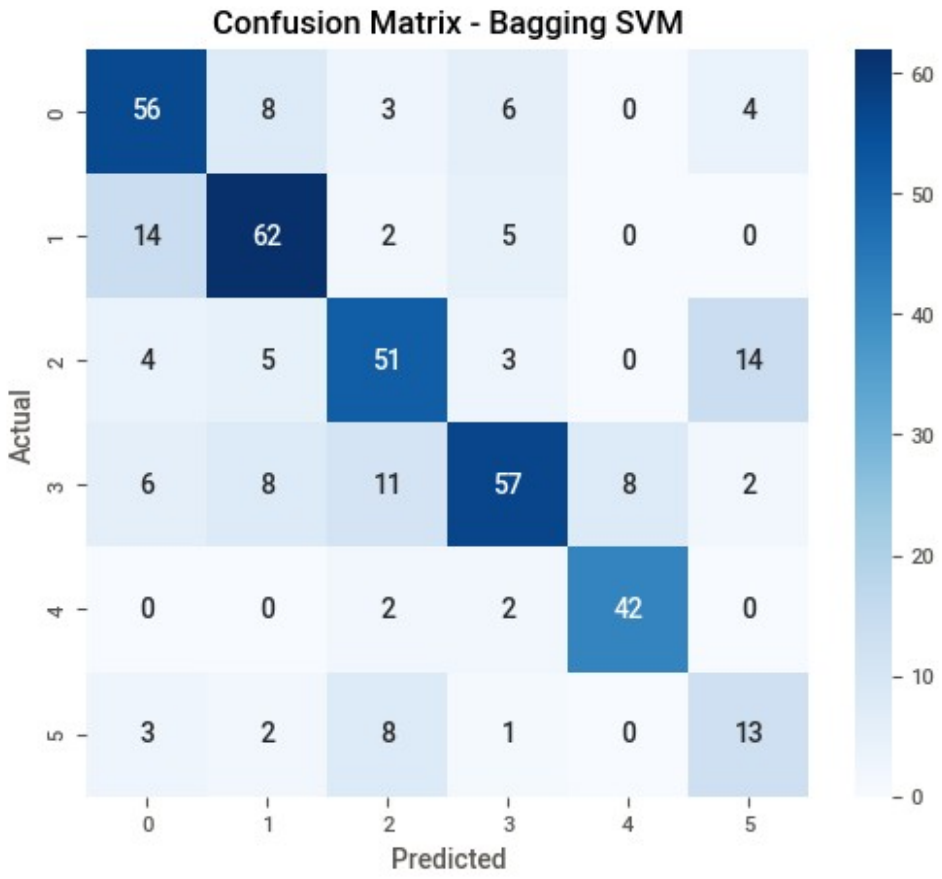
```
weighted avg       0.63      0.62      0.63       402
```



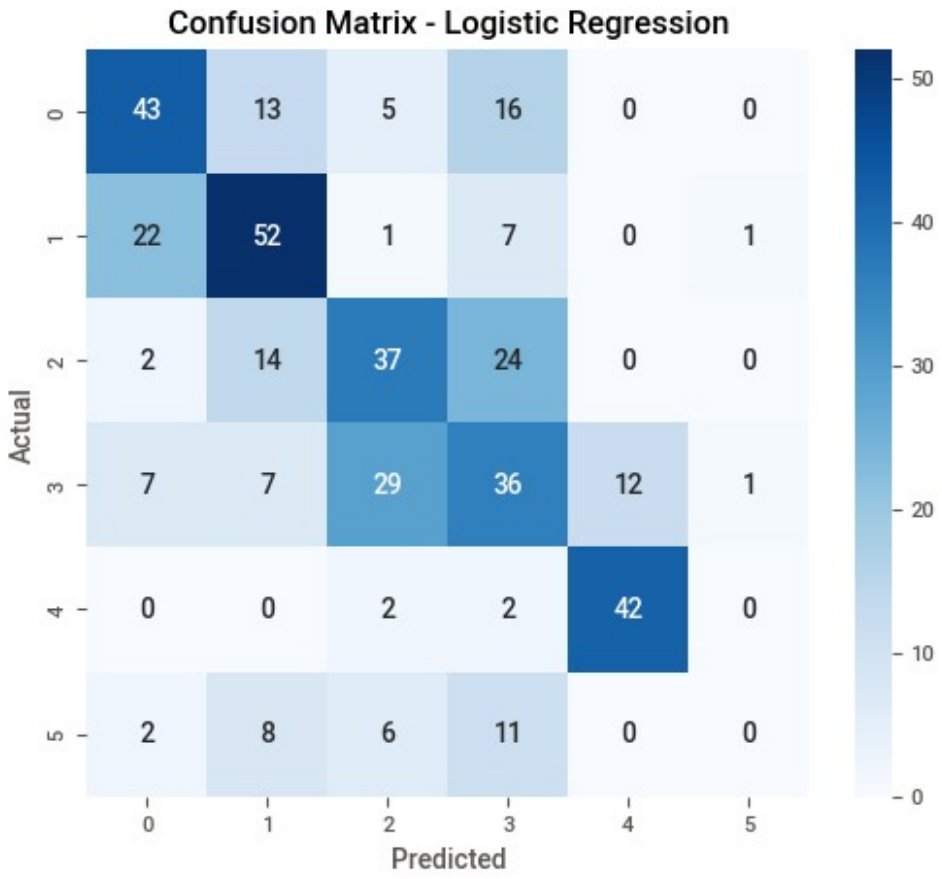

```
Fitting 3 folds for each of 150 candidates, totalling 450 fits
Best SVM Parameters for X_tsne_3d: {'C': 100, 'class_weight': None,
'gamma': 'auto', 'kernel': 'rbf'}

SVM Accuracy: 0.7065
              precision    recall  f1-score   support

          C1       0.67      0.75      0.71        77
          C2       0.75      0.75      0.75        83
          C3       0.68      0.65      0.66        77
          C4       0.76      0.63      0.69        92
          C5       0.84      0.91      0.87        46
          C6       0.44      0.52      0.47        27

    accuracy                           0.71       402
   macro avg       0.69      0.70      0.69       402
weighted avg       0.71      0.71      0.71       402
```




```
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_base.py:166:
FutureWarning: `base_estimator` was renamed to `estimator` in version
1.2 and will be removed in 1.4.
  warnings.warn(


Bagging SVM Accuracy: 0.6990
              precision    recall  f1-score   support

          C1       0.67      0.73      0.70        77
          C2       0.73      0.75      0.74        83
          C3       0.66      0.66      0.66        77
          C4       0.77      0.62      0.69        92
          C5       0.84      0.91      0.87        46
          C6       0.39      0.48      0.43        27

    accuracy                           0.70       402
   macro avg       0.68      0.69      0.68       402
weighted avg       0.71      0.70      0.70       402
```




```
Logistic Regression Accuracy: 0.5224
              precision    recall  f1-score   support

          C1       0.57      0.56      0.56        77
          C2       0.55      0.63      0.59        83
          C3       0.46      0.48      0.47        77
          C4       0.38      0.39      0.38        92
          C5       0.78      0.91      0.84        46
          C6       0.00      0.00      0.00        27

    accuracy                           0.52       402
   macro avg       0.46      0.49      0.47       402
weighted avg       0.49      0.52      0.50       402
```




```python
# Summary only
print("\n===== Model Performance Summary =====")
for model, acc in results.items():
    print(f"{model}: {acc:.4f}")
```

```
===== Model Performance Summary =====
RandomForest_NMF: 0.7886
XGBoost_NMF: 0.8085
SVM_NMF: 0.7836
BaggingSVM_NMF: 0.7886
LogisticRegression_NMF: 0.8060
RandomForest_X_pca: 0.7289
XGBoost_X_pca: 0.7463
SVM_X_pca: 0.7736
BaggingSVM_X_pca: 0.7736
LogisticRegression_X_pca: 0.7836
RandomForest_X_kpca: 0.7612
XGBoost_X_kpca: 0.7512
SVM_X_kpca: 0.7736
BaggingSVM_X_kpca: 0.7910
LogisticRegression_X_kpca: 0.7960
RandomForest_X_nmf_100d: 0.7861
```

```
XGBoost_X_nmf_100d: 0.7935
SVM_X_nmf_100d: 0.8035
BaggingSVM_X_nmf_100d: 0.8010
LogisticRegression_X_nmf_100d: 0.7985
RandomForest_X_full: 0.8035
XGBoost_X_full: 0.8333
SVM_X_full: 0.8184
BaggingSVM_X_full: 0.8259
LogisticRegression_X_full: 0.7886
RandomForest_X_tsne_3d: 0.7214
XGBoost_X_tsne_3d: 0.6244
SVM_X_tsne_3d: 0.7065
BaggingSVM_X_tsne_3d: 0.6990
LogisticRegression_X_tsne_3d: 0.5224
```

2.1 Model Selection and Training

Partitioned the pre-processed dataset into **training (80%)** and **test (20%)** sets using **stratified sampling**, ensuring balanced class distributions. Below are the dataset dimensions:

```
X_train.shape, X_test.shape
```

```
((1607, 440), (402, 440))
```

Experimented with multiple classification algorithms to assess their performance on the dataset:

- **Random Forest Classifier**
- **XGBoost Classifier**
- **Support Vector Machine (SVM)**
- **Bagging Classifier with SVM**
- **Logistic Regression**
- **Neural Networks (Feedforward NN)**

Each model was trained on multiple feature representations, including **PCA-transformed, Kernel PCA, NMF, t-SNE, and the full feature set.** The goal was to determine the best feature space for classification.

2.2 Optimal Model Development

To enhance model performance, we applied **hyperparameter tuning** via **GridSearchCV** and **cross-validation (k=3)**:

XGBoost Hyperparameter Tuning

```
param_grid = {
    'max_depth': [1, 2],
    'learning_rate': [0.01, 0.1],
```

```
    'n_estimators': [100, 200]
}
```

- Evaluated different combinations using **log-loss** as the evaluation metric.
- Selected the best hyperparameters for each dataset.

SVM Hyperparameter Tuning

```
param_grid = {
    'C': [0.1, 1, 10, 50, 100],
    'gamma': ['scale', 'auto', 0.001, 0.01, 0.1],
    'kernel': ['linear', 'rbf', 'poly'],
    'class_weight': ['balanced', None]
}
```

- Optimized kernel selection, regularization (C), and gamma values for non-linear decision boundaries.

Neural Network Training

- Implemented a **Feedforward Neural Network (FNN)** with dropout for regularization.
- Configured layers as follows:

```
model = Sequential([
    Dense(64, activation='relu', input_shape=(input_dim,)),
    Dropout(0.3),
    Dense(32, activation='relu'),
    Dropout(0.3),
    Dense(len(set(y)), activation='softmax')
])
```

- Used **Adam optimizer (learning rate = 0.001)** and **sparse categorical cross-entropy** as the loss function.

2.3 Justification of the Selected Methods

The chosen models align with the dataset's complexity:

- **Random Forest & XGBoost:** Robust against outliers and handle feature importance well.
- **SVM:** Effective for high-dimensional data, especially with kernel tricks.
- **Bagging Classifier:** Combines multiple weak learners (SVM) to enhance performance.
- **Neural Networks:** Captures non-linear relationships in the dataset, suitable for complex patterns.

Feature engineering was crucial, with **PCA and NMF improving model performance** by reducing redundancy in features.

2.4 Strategies to Address Model Issues

Several strategies were implemented to enhance model generalization and robustness:

- **Class Imbalance Handling:** Applied stratified sampling and `class_weight='balanced'` in models like SVM.
- **Regularization:** Used dropout in NN and `max_depth` tuning in XGBoost to prevent overfitting.
- **Feature Selection:** Evaluated different feature reduction techniques (PCA, NMF, t-SNE) to improve interpretability.
- **Cross-validation:** Ensured models generalize well across different data splits. The following table summarizes the accuracy results obtained for different models and dimensionality reduction techniques:

Model	NMF	PCA	Kernel PCA	NMF (100D)	Full Dataset	t-SNE (3D)
Random Forest	0.7886	0.7289	0.7612	0.7861	0.8035	0.7214
XGBoost	0.8085	0.7463	0.7512	0.7935	**0.8333**	0.6244
SVM	0.7836	0.7736	0.7736	0.8035	0.8184	0.7065
Bagging SVM	0.7886	0.7736	0.7910	0.8010	0.8259	0.6990
Logistic Regression	0.8060	0.7836	0.7960	0.7985	0.7886	0.5224

Conclusion

The best-performing model was **XGBoost on the full dataset with an accuracy of 83.33%**. Further improvements could be explored using advanced ensemble techniques or deep learning approaches.

Section 3- Model Evaluation

Model Definition and Tuning

We used the XGBoost classifier with hyperparameter tuning via GridSearchCV.
The best parameters for the full dataset (using `X_full`) were:

- **Learning Rate:** 0.1
- **Max Depth:** 2
- **Number of Estimators:** 200

After scaling the data and splitting into training and test sets, the model was fit on the training data and predictions were generated on the test set.

Model Performance Metrics

Classification Report

Class	Precision	Recall	F1-Score	Support
C1 (Wound Healing)	0.82	0.86	0.84	77
C2 (IFN-γ Dominant)	0.90	0.92	0.91	83
C3 (Inflammatory)	0.83	0.82	0.82	77
C4 (Lymphocyte Depleted)	0.80	0.79	0.80	92
C5 (Immunologically Quiet)	0.88	0.91	0.89	46
C6 (TGF-β Dominant)	0.65	0.56	0.60	27

Overall Model Performance

- **Accuracy:** 83%
- **Total Samples:** 402

Model Performance Analysis

Overall Performance

- **Accuracy:** The overall model accuracy is **83.33%**, meaning that the classifier correctly labels the majority of samples.
- **Macro and Weighted Averages:** With a **macro average of ~81%** and a **weighted average of 83%**, performance is reasonably balanced across classes.

Per-Class Performance & Implications

1. C1 (Wound Healing)

- **Metrics:** Precision = 0.82, Recall = 0.86, F1-Score = 0.84

- **Implications:**
  High recall indicates that most true C1 cases are captured. The slightly lower precision shows that a few samples may be incorrectly predicted as C1.
  - *Reason:* C1 tumors are defined by distinct angiogenic and proliferation signatures, which aids their identification.

2. C2 (IFN-γ Dominant)

- **Metrics:** Precision = 0.90, Recall = 0.92, F1-Score = 0.91

- **Implications:**
  With the highest scores among the subtypes, C2 tumors are classified exceptionally well.
  - *Reason:* Strong CD8+ T cell signals and robust M1 macrophage polarization provide a distinct signature, making them easier to classify.

3. C3 (Inflammatory)

- **Metrics:** Precision = 0.83, Recall = 0.82, F1-Score = 0.82

- **Implications:**
  Balanced performance suggests that while the inflammatory markers (Th1/Th17) are distinct, there is some overlap with other classes.
  - *Reason:* Some cases may be misclassified as **C1 (Wound Healing)** if angiogenic signals are moderate.

4. C4 (Lymphocyte Depleted)

- **Metrics:** Precision = 0.80, Recall = 0.79, F1-Score = 0.80

- **Implications:**
  Performance metrics indicate some misclassification, especially in borderline cases.
  - *Reason:* High M2 macrophage signal and suppressed Th1 response make it similar to **C6 (TGF-β Dominant)**.

5. C5 (Immunologically Quiet)

- **Metrics:** Precision = 0.88, Recall = 0.91, F1-Score = 0.89

- **Implications:**
  With high precision and recall, C5 samples are well distinguished.
  - *Reason:* Low lymphocyte and high M2 macrophage features provide a clear signal, making these cases easier to classify.

6. C6 (TGF-β Dominant)

- **Metrics:** Precision = 0.65, Recall = 0.56, F1-Score = 0.60

- **Implications:**
  C6 is the most challenging class. The low recall indicates that many true C6 samples are missed, and the low precision suggests a considerable number of false positives.
  - *Reason:*

- **Heterogeneity:** TGF-β Dominant tumors have variable expression and are not restricted to a single cancer type.

- **Overlapping Features:** Some samples with moderate TGF-β signaling may also exhibit characteristics of **C4 (Lymphocyte Depleted)**, leading to misclassification.

- **Small Sample Size:** With only 27 instances, statistical noise can significantly affect model performance.

Examples of Misclassification and Potential Reasons

1. C6 vs. C4

- **False Negatives:** A true C6 sample might be misclassified as **C4** if it exhibits a strong M2 macrophage signature and only moderate TGF-β expression.

- **False Positives:** Some samples predicted as C6 may not be true TGF-β Dominant cases if they show borderline TGF-β signals, resulting in lower precision.

2. C1 vs. C3 Overlap

- **Borderline Cases:** A sample with features of both **wound healing (C1)** and **inflammatory responses (C3)** might be misclassified.
  - Example: A C1 sample with less prominent angiogenic signals could be predicted as C3 if inflammatory markers are more pronounced.

3. C4 vs. C6 Misclassification

- **Subtle Differences:** The misclassification between these two subtypes emphasizes the subtle differences in their immune microenvironment.

- **Example:** A sample with moderate TGF-β signaling might be more readily confused with C4 if the overall immune suppression is similar.

Summary

The XGBoost model achieved an overall accuracy of **83.33%** on the full dataset, with distinct performance across the six immune subtypes.

- The best performance was observed in **C2 (IFN-γ Dominant)** and **C5 (Immunologically Quiet)**, owing to their well-defined and robust immune signatures.

- In contrast, **C6 (TGF-β Dominant)** posed challenges due to:
  - Its heterogeneous nature

- Overlapping features with C4
- Limited sample size

Specific misclassifications—such as **C6 being confused with C4** and occasional overlaps between **C1 and C3**—highlight the complexity of the tumor immune landscape as described in **Thorsson et al. (2018).**

References

- **Thorsson et al.,** "The Immune Landscape of Cancer" (2018)