▾ TABULA SAPIENS - HEART

```
#Trying with smaller dataset
#Around 400 MB
```

▾ Trying with smaller dataset

Around 400 MB

```
!pip install anndata
```

```
Collecting anndata
  Downloading anndata-0.9.2-py3-none-any.whl (104 kB)
                                    ━━━━━━━━━━━━━━━━━━━ 104.2/104.2 kB 1.4 MB/s eta 0:00:00
Requirement already satisfied: pandas!=2.0.1,>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from anndata) (1.5.3)
Requirement already satisfied: numpy>=1.16.5 in /usr/local/lib/python3.10/dist-packages (from anndata) (1.23.5)
Requirement already satisfied: scipy>1.4 in /usr/local/lib/python3.10/dist-packages (from anndata) (1.11.2)
Requirement already satisfied: h5py>=3 in /usr/local/lib/python3.10/dist-packages (from anndata) (3.9.0)
Requirement already satisfied: natsort in /usr/local/lib/python3.10/dist-packages (from anndata) (8.4.0)
Requirement already satisfied: packaging>=20 in /usr/local/lib/python3.10/dist-packages (from anndata) (23.1)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=2.0.1,>=1.1.1->anndata) (
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=2.0.1,>=1.1.1->anndata) (2023.3.pos
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas!=2.0.1,>=1.1.1->
Installing collected packages: anndata
Successfully installed anndata-0.9.2
```

```
import numpy as np # linear algebra
import pandas as pd
import os

import time
import matplotlib.pyplot as plt
import seaborn as sns
```

```
#1 Code to read file into colaboratory:
!pip install -U -q PyDrive
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials

# 1. Authenticate and create the PyDrive client.
auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)
```

```
-------------------------------------------------------------------------
#Tabula Sapiens Heart Cells ------>

# https://drive.google.com/file/d/1oMOjGyf14G4dcyqrQChlD4f7TK0YV2Vf/view?usp=sharing - link to dataset

#2. Get the file
#make sure you upload all your data files to your Google drive and change share->Advanced->change->anyone with the link can view
downloaded = drive.CreateFile({'id':'1oMOjGyf14G4dcyqrQChlD4f7TK0YV2Vf'}) # replace the id with id of file you want to access
downloaded.GetContentFile('TabulaSapiens_Heart_Dataset.h5ad')
```

/usr/local/lib/python3.10/dist-packages/google/colab/ message.py in read reply from input(message id, timeout sec)

```
print(downloaded)
```

GoogleDriveFile({'id': '1oMOjGyf14G4dcyqrQChlD4f7TK0YV2Vf', 'kind': 'drive#file', 'userPermission': {'id': 'me', 'type': 'user', 'role':

```
!pip install scanpy
```

```
Requirement already satisfied: scanpy in /usr/local/lib/python3.10/dist-packages (1.9.5)
Requirement already satisfied: anndata>=0.7.4 in /usr/local/lib/python3.10/dist-packages (from scanpy) (0.9.2)
Requirement already satisfied: numpy>=1.17.0 in /usr/local/lib/python3.10/dist-packages (from scanpy) (1.23.5)
Requirement already satisfied: matplotlib>=3.4 in /usr/local/lib/python3.10/dist-packages (from scanpy) (3.7.1)
Requirement already satisfied: pandas>=1.0 in /usr/local/lib/python3.10/dist-packages (from scanpy) (1.5.3)
Requirement already satisfied: scipy>=1.4 in /usr/local/lib/python3.10/dist-packages (from scanpy) (1.11.2)
Requirement already satisfied: seaborn in /usr/local/lib/python3.10/dist-packages (from scanpy) (0.12.2)
Requirement already satisfied: h5py>=3 in /usr/local/lib/python3.10/dist-packages (from scanpy) (3.9.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from scanpy) (4.66.1)
Requirement already satisfied: scikit-learn>=0.24 in /usr/local/lib/python3.10/dist-packages (from scanpy) (1.2.2)
Requirement already satisfied: statsmodels>=0.10.0rc2 in /usr/local/lib/python3.10/dist-packages (from scanpy) (0.14.0)
Requirement already satisfied: patsy in /usr/local/lib/python3.10/dist-packages (from scanpy) (0.5.3)
Requirement already satisfied: networkx>=2.3 in /usr/local/lib/python3.10/dist-packages (from scanpy) (3.1)
Requirement already satisfied: natsort in /usr/local/lib/python3.10/dist-packages (from scanpy) (8.4.0)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from scanpy) (1.3.2)
Requirement already satisfied: numba>=0.41.0 in /usr/local/lib/python3.10/dist-packages (from scanpy) (0.56.4)
Requirement already satisfied: umap-learn>=0.3.10 in /usr/local/lib/python3.10/dist-packages (from scanpy) (0.5.3)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from scanpy) (23.1)
Requirement already satisfied: session-info in /usr/local/lib/python3.10/dist-packages (from scanpy) (1.0.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy) (1.1.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy) (4.42.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy) (1.4.5)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy) (2.8.2)
Requirement already satisfied: llvmlite<0.40,>=0.39.0dev0 in /usr/local/lib/python3.10/dist-packages (from numba>=0.41.0->scanpy) (0.39.
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from numba>=0.41.0->scanpy) (67.7.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.0->scanpy) (2023.3.post1)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.24->scanpy) (3.2.0)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy->scanpy) (1.16.0)
Requirement already satisfied: pynndescent>=0.5 in /usr/local/lib/python3.10/dist-packages (from umap-learn>=0.3.10->scanpy) (0.5.10)
Requirement already satisfied: stdlib-list in /usr/local/lib/python3.10/dist-packages (from session-info->scanpy) (0.9.0)
```

```
import scanpy as sc
import anndata
import importlib
from sklearn.decomposition import PCA

import matplotlib as mpl


adata = sc.read_h5ad('TabulaSapiens_Heart_Dataset.h5ad')



print(adata)
```

```
AnnData object with n_obs × n_vars = 11505 × 58604
    obs: 'assay_ontology_term_id', 'donor_id', 'anatomical_information', 'n_counts_UMIs', 'n_genes', 'cell_ontology_class', 'free_annota
    var: 'feature_type', 'highly_variable', 'means', 'dispersions', 'dispersions_norm', 'mean', 'std', 'ensembl_version', 'feature_is_fi
    uns: '_scvi', '_training_mode', 'assay_colors', 'cell_ontology_class_colors', 'dendrogram_cell_type_tissue', 'dendrogram_computation
    obsm: 'X_pca', 'X_scvi', 'X_scvi_umap', 'X_umap'
    obsp: 'connectivities', 'distances'
```

```python
# View basic statistics of the data
print(adata.var_names[:10])  # View the first 10 gene names
```

```
Index(['ENSG00000223972', 'ENSG00000227232', 'ENSG00000278267',
       'ENSG00000243485', 'ENSG00000284332', 'ENSG00000237613',
       'ENSG00000268020', 'ENSG00000240361', 'ENSG00000186092',
       'ENSG00000238009'],
      dtype='object', name='ensemblid')
```

```python
print(adata.obs.head())      # View the first few rows of the observation (cell) data
```

```
                                              assay_ontology_term_id donor_id  \
cell_id
AAACCCAAGAGCAAGA_TSP12_Heart_Atria_10X_1_1              EFO:0009922    TSP12
AAACCCAAGATGGCGT_TSP12_Heart_Atria_10X_1_1              EFO:0009922    TSP12
AAACCCAAGGGTTAAT_TSP12_Heart_Atria_10X_1_1              EFO:0009922    TSP12
AAACCCAAGTATGCAA_TSP12_Heart_Atria_10X_1_1              EFO:0009922    TSP12
AAACCCAAGTCGTTAC_TSP12_Heart_Atria_10X_1_1              EFO:0009922    TSP12

                                           anatomical_information  \
cell_id
AAACCCAAGAGCAAGA_TSP12_Heart_Atria_10X_1_1                  Atria
AAACCCAAGATGGCGT_TSP12_Heart_Atria_10X_1_1                  Atria
AAACCCAAGGGTTAAT_TSP12_Heart_Atria_10X_1_1                  Atria
AAACCCAAGTATGCAA_TSP12_Heart_Atria_10X_1_1                  Atria
AAACCCAAGTCGTTAC_TSP12_Heart_Atria_10X_1_1                  Atria

                                           n_counts_UMIs  n_genes  \
cell_id
AAACCCAAGAGCAAGA_TSP12_Heart_Atria_10X_1_1        2864.0     1375
AAACCCAAGATGGCGT_TSP12_Heart_Atria_10X_1_1        5179.0     2333
AAACCCAAGGGTTAAT_TSP12_Heart_Atria_10X_1_1       36202.0     6249
AAACCCAAGTATGCAA_TSP12_Heart_Atria_10X_1_1        7949.0     2377
AAACCCAAGTCGTTAC_TSP12_Heart_Atria_10X_1_1       29970.0     6492

                                                cell_ontology_class  \
cell_id
AAACCCAAGAGCAAGA_TSP12_Heart_Atria_10X_1_1  cardiac endothelial cell
AAACCCAAGATGGCGT_TSP12_Heart_Atria_10X_1_1  cardiac endothelial cell
AAACCCAAGGGTTAAT_TSP12_Heart_Atria_10X_1_1        cardiac muscle cell
AAACCCAAGTATGCAA_TSP12_Heart_Atria_10X_1_1        cardiac muscle cell
AAACCCAAGTCGTTAC_TSP12_Heart_Atria_10X_1_1        cardiac muscle cell

                                               free_annotation  \
cell_id
AAACCCAAGAGCAAGA_TSP12_Heart_Atria_10X_1_1     Endothelial Cells
AAACCCAAGATGGCGT_TSP12_Heart_Atria_10X_1_1     Endothelial Cells
AAACCCAAGGGTTAAT_TSP12_Heart_Atria_10X_1_1  Cardiac Muscle Cells
AAACCCAAGTATGCAA_TSP12_Heart_Atria_10X_1_1  Cardiac Muscle Cells
AAACCCAAGTCGTTAC_TSP12_Heart_Atria_10X_1_1  Cardiac Muscle Cells

                                           manually_annotated  compartment  \
cell_id
AAACCCAAGAGCAAGA_TSP12_Heart_Atria_10X_1_1               True  endothelial
AAACCCAAGATGGCGT_TSP12_Heart_Atria_10X_1_1               True  endothelial
AAACCCAAGGGTTAAT_TSP12_Heart_Atria_10X_1_1               True      stromal
AAACCCAAGTATGCAA_TSP12_Heart_Atria_10X_1_1               True      stromal
AAACCCAAGTCGTTAC_TSP12_Heart_Atria_10X_1_1               True      stromal

                                           sex_ontology_term_id  ...  \
cell_id                                                          ...
AAACCCAAGAGCAAGA_TSP12_Heart_Atria_10X_1_1          PATO:0000384  ...
AAACCCAAGATGGCGT_TSP12_Heart_Atria_10X_1_1          PATO:0000384  ...
AAACCCAAGGGTTAAT_TSP12_Heart_Atria_10X_1_1          PATO:0000384  ...
AAACCCAAGTATGCAA_TSP12_Heart_Atria_10X_1_1          PATO:0000384  ...
AAACCCAAGTCGTTAC_TSP12_Heart_Atria_10X_1_1          PATO:0000384  ...

                                           development_stage_ontology_term_id  \
cell_id
```

```python
# Plot the distribution of the gene expression levels
adata.plot.density(x='rna')
```

```
        ------------------------------------------------------------------------
        AttributeError                              Traceback (most recent call last)
        <ipython-input-33-6e556056fcfb> in <cell line: 2>()
```
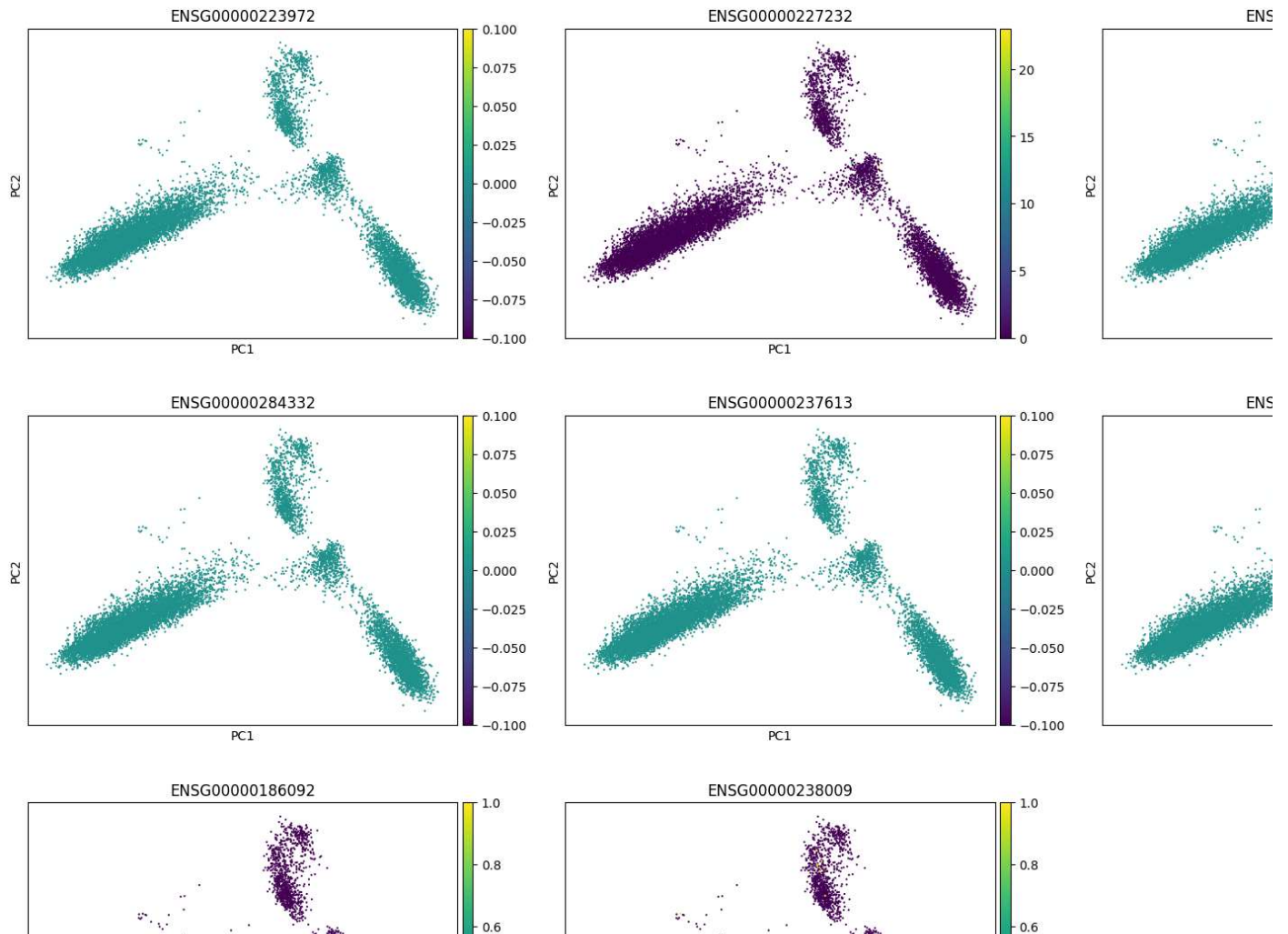
```
# Get the dimensions of the data
print("Number of Cells:", adata.n_obs)
print("Number of Genes:", adata.n_vars)
```

```
        Number of Cells: 11505
        Number of Genes: 58604
```
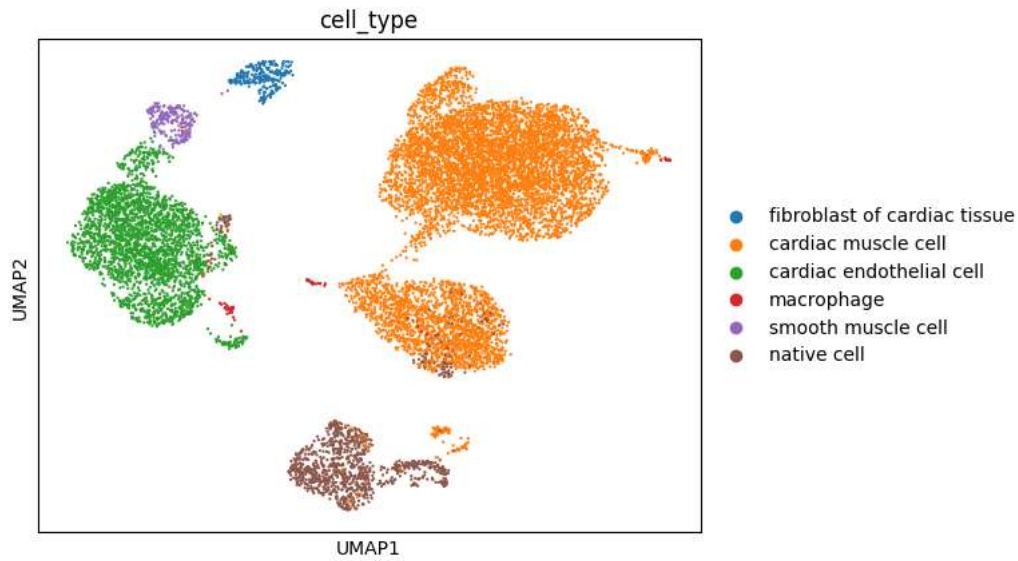
Dimensionality Reduction (e.g., PCA):

```
sc.tl.pca(adata)
sc.pl.pca(adata, color=['ENSG00000223972', 'ENSG00000227232' , 'ENSG00000278267', 'ENSG00000243485', 'ENSG00000284332', 'ENSG00000237613', 'E
```



```
# UMAP embedding
sc.pl.umap(adata, color='cell_type')
```

```
# PCA plot
sc.pl.pca(adata, color='cell_type')
```

```
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_tools/scatterplots.py:391: UserWarning: No data for colormapping provided via '
  cax = scatter(
```



```
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_tools/scatterplots.py:391: UserWarning: No data for colormapping provided via '
  cax = scatter(
```