

▼ Step 0: Loading the data and installing the libraries

```
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

!pip install scanpy==1.9.3

Collecting scanpy==1.9.3
  Downloading scanpy-1.9.3-py3-none-any.whl (2.0 MB)
    2.0/2.0 MB 19.0 MB/s eta 0:00:00
Collecting anndata>=0.7.4 (from scanpy==1.9.3)
  Downloading anndata-0.10.6-py3-none-any.whl (122 kB)
    122.1/122.1 kB 14.1 MB/s eta 0:00:00
Requirement already satisfied: numpy>=1.17.0 in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (1.25.2)
Requirement already satisfied: matplotlib>=3.4 in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (3.7.1)
Requirement already satisfied: pandas>=1.0 in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (1.5.3)
Requirement already satisfied: scipy>=1.4 in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (1.11.4)
Requirement already satisfied: seaborn in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (0.13.1)
Requirement already satisfied: h5py>=3 in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (3.9.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (4.66.2)
Requirement already satisfied: scikit-learn>=0.22 in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (1.2.2)
Requirement already satisfied: statsmodels>=0.10.0rc2 in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (0.14.
Requirement already satisfied: patsy in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (0.5.6)
Requirement already satisfied: networkx>=2.3 in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (3.2.1)
Requirement already satisfied: natsort in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (8.4.0)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (1.3.2)
Requirement already satisfied: numba>=0.41.0 in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (0.58.1)
Collecting umap-learn>=0.3.10 (from scanpy==1.9.3)
  Downloading umap-learn-0.5.5.tar.gz (90 kB)
    90.9/90.9 kB 10.2 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from scanpy==1.9.3) (23.2)
Collecting session-info (from scanpy==1.9.3)
  Downloading session_info-1.0.0.tar.gz (24 kB)
  Preparing metadata (setup.py) ... done
Collecting array-api-compat!=1.5,>1.4 (from anndata>=0.7.4->scanpy==1.9.3)
  Downloading array_api_compat-1.4.1-py3-none-any.whl (30 kB)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from anndata>=0.7.4->scanpy==1.9.3)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy==1.
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy==1.9.3)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy==1
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy==1
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy==1.9.3
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy==1
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.4->scanpy
Requirement already satisfied: llvmlite<0.42,>=0.41.0dev0 in /usr/local/lib/python3.10/dist-packages (from numba>=0.41.0->sc
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.0->scanpy==1.9.3) (20
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.22->sc
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy>scanpy==1.9.3) (1.16.0)
Collecting pynndescent>=0.5 (from umap-learn>=0.3.10->scanpy==1.9.3)
  Downloading pynndescent-0.5.11-py3-none-any.whl (55 kB)
    55.8/55.8 kB 5.4 MB/s eta 0:00:00
Collecting stdlib_list (from session-info>scanpy==1.9.3)
  Downloading stdlib_list-0.10.0-py3-none-any.whl (79 kB)
    79.8/79.8 kB 9.0 MB/s eta 0:00:00
Building wheels for collected packages: umap-learn, session-info
  Building wheel for umap-learn (setup.py) ... done
    Created wheel for umap-learn: filename=umap_learn-0.5.5-py3-none-any.whl size=86832 sha256=b5d4ce22ad5d374db55dba08ae8e13f
    Stored in directory: /root/.cache/pip/wheels/3a/70/07/428d2b58660a1a3b431db59b806a10da736612ebbc66c1bcc5
  Building wheel for session-info (setup.py) ... done
    Created wheel for session-info: filename=session_info-1.0.0-py3-none-any.whl size=8027 sha256=7aeaf2e87265623d15b72811e019
    Stored in directory: /root/.cache/pip/wheels/6a/aa/b9/eb5d4031476ec10802795b97ccf937b9bd998d68a9b268765a
Successfully built umap-learn session-info
Installing collected packages: stdlib_list, array-api-compat, session-info, pynndescent, anndata, umap-learn, scanpy
Successfully installed anndata-0.10.6 array-api-compat-1.4.1 pynndescent-0.5.11 scanpy-1.9.3 session-info-1.0.0 stdlib_list-0.10.0

!pip install leidenalg

Collecting leidenalg
  Downloading leidenalg-0.10.2-cp38-abi3-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.0 MB)
    2.0/2.0 MB 27.2 MB/s eta 0:00:00
Collecting igraph<0.12,>=0.10.0 (from leidenalg)
  Downloading igraph-0.11.4-cp39-abi3-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.3 MB)
    3.3/3.3 MB 49.6 MB/s eta 0:00:00
Collecting texttable>=1.6.2 (from igraph<0.12,>=0.10.0->leidenalg)
  Downloading texttable-1.7.0-py2.py3-none-any.whl (10 kB)
```

```
Installing collected packages: texttable, igraph, leidenalg
Successfully installed igraph-0.11.4 leidenalg-0.10.2 texttable-1.7.0
```

```
!pip install decoupler
```

```
Collecting decoupler
  Downloading decoupler-1.6.0-py3-none-any.whl (95 kB)
  ━━━━━━━━━━━━━━━━ 95.8/95.8 KB 1.7 MB/s eta 0:00:00
Requirement already satisfied: numba in /usr/local/lib/python3.10/dist-packages (from decoupler) (0.58.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from decoupler) (4.66.2)
Requirement already satisfied: anndata in /usr/local/lib/python3.10/dist-packages (from decoupler) (0.10.6)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from decoupler) (4.10.0)
Requirement already satisfied: array-api-compat!=1.5,>1.4 in /usr/local/lib/python3.10/dist-packages (from anndata->decoupler) (1.2.0)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from anndata->decoupler) (1.2.0)
Requirement already satisfied: h5py>=3.1 in /usr/local/lib/python3.10/dist-packages (from anndata->decoupler) (3.9.0)
Requirement already satisfied: natsort in /usr/local/lib/python3.10/dist-packages (from anndata->decoupler) (8.4.0)
Requirement already satisfied: numpy>=1.23 in /usr/local/lib/python3.10/dist-packages (from anndata->decoupler) (1.25.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from anndata->decoupler) (23.2)
Requirement already satisfied: pandas!=2.1.0rc0,!>2.1.2,>=1.4 in /usr/local/lib/python3.10/dist-packages (from anndata->decoupler) (2.1.4)
Requirement already satisfied: scipy>1.8 in /usr/local/lib/python3.10/dist-packages (from anndata->decoupler) (1.11.4)
Requirement already satisfied: llvmlite<0.42,>=0.41.0dev0 in /usr/local/lib/python3.10/dist-packages (from numba->decoupler) (0.41.0)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=2.1.0rc0,!>2.1.2,>=1.4->decoupler) (2.31.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=2.1.0rc0,!>2.1.2,>=1.4->decoupler) (2023.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas!=2.1.0rc0,!>2.1.2,>=1.4->decoupler) (1.16.0)
Installing collected packages: decoupler
Successfully installed decoupler-1.6.0
```

```
!pip install omnipath
```

```
Collecting omnipath
  Downloading omnipath-1.0.8-py3-none-any.whl (68 kB)
  ━━━━━━━━━━━━━━━━ 68.9/68.9 KB 2.6 MB/s eta 0:00:00
Requirement already satisfied: attrs>=20.2.0 in /usr/local/lib/python3.10/dist-packages (from omnipath) (23.2.0)
Collecting docrep>=0.3.1 (from omnipath)
  Downloading docrep-0.3.2.tar.gz (33 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: inflect>=4.1.0 in /usr/local/lib/python3.10/dist-packages (from omnipath) (7.0.0)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from omnipath) (23.2)
Requirement already satisfied: pandas>=1.2.0 in /usr/local/lib/python3.10/dist-packages (from omnipath) (1.5.3)
Requirement already satisfied: requests>=2.24.0 in /usr/local/lib/python3.10/dist-packages (from omnipath) (2.31.0)
Requirement already satisfied: tqdm>=4.51.0 in /usr/local/lib/python3.10/dist-packages (from omnipath) (4.66.2)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from omnipath) (4.10.6)
Requirement already satisfied: urllib3>=1.26.0 in /usr/local/lib/python3.10/dist-packages (from omnipath) (2.0.7)
Requirement already satisfied: wrapt>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from omnipath) (1.14.1)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from docrep>=0.3.1->omnipath) (1.16.0)
Requirement already satisfied: pydantic>=1.9.1 in /usr/local/lib/python3.10/dist-packages (from inflect>=4.1.0->omnipath) (1.10.0)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.2.0->omnipath) (2.31.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.2.0->omnipath) (2023.1)
Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.2.0->omnipath) (1.25.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.24.0->omnipath) (3.0.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.24.0->omnipath) (3.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.24.0->omnipath) (2023.1)
Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic>=1.9.1->inflect) (0.4.0)
Requirement already satisfied: pydantic-core==2.16.3 in /usr/local/lib/python3.10/dist-packages (from pydantic>=1.9.1->inflect) (2.16.3)
Building wheels for collected packages: docrep
  Building wheel for docrep (setup.py) ... done
  Created wheel for docrep: filename=docrep-0.3.2-py3-none-any.whl size=19878 sha256=7e17049d78f60d945071795bb27bafe9c7a01fc
  Stored in directory: /root/.cache/pip/wheels/c3/64/48/03c38d8d906159eaa210b3c548fdb590eb3e2a4a5745ae2172
Successfully built docrep
Installing collected packages: docrep, omnipath
Successfully installed docrep-0.3.2 omnipath-1.0.8
```

```
import scanpy as sc
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

Reading visium data

```
adata = sc.read_visium('/content/drive/MyDrive/Dataset')
```

```
adata.var_names_make_unique()
```

```
/usr/local/lib/python3.10/dist-packages/anndata/_core/anndata.py:1820: UserWarning: Variable names are not unique. To make them unique, use utils.warn_names_duplicates("var")
/usr/local/lib/python3.10/dist-packages/anndata/_core/anndata.py:1820: UserWarning: Variable names are not unique. To make them unique, use utils.warn_names_duplicates("var")
```

The obs are "spots", not "cells". So kinda low-res bulk RNA

```
adata #2518 spots X 17943 genes
```

```
AnnData object with n_obs × n_vars = 2518 × 17943
obs: 'in_tissue', 'array_row', 'array_col'
var: 'gene_ids', 'feature_types', 'genome'
uns: 'spatial'
obsm: 'spatial'
```

Image data (coordinates) are stored under the following adata slot:

```
adata.uns['spatial']['Visium_FFPE_Human_Breast_Cancer']['images']
```

```
{'hires': array([[ [1.          , 1.          , 1.          ],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99607843, 0.99215686],
   ...,
   [0.9843137 , 0.99607843, 0.99215686],
   [0.9882353 , 0.99607843, 0.99215686],
   [0.99607843, 0.99607843, 0.99607843]],

  [[1.          , 1.          , 1.          ],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99607843, 0.99215686],
   ...,
   [0.9843137 , 0.99215686, 0.9882353 ],
   [0.9882353 , 0.99215686, 0.99215686],
   [0.99607843, 0.99607843, 0.99607843]],

  [[1.          , 1.          , 1.          ],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99607843, 0.99215686],
   ...,
   [0.9843137 , 0.99215686, 0.9882353 ],
   [0.9843137 , 0.99215686, 0.9882353 ],
   [0.99215686, 0.99607843, 0.99607843]],

  ...,
  [[0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   ...,
   [0.99607843, 1.          , 0.99607843],
   [1.          , 1.          , 1.          ],
   [0.99607843, 0.99607843, 0.99607843]],

  [[0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   ...,
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99215686, 0.99215686]],

  [[0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   ...,
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99215686, 0.99215686]],

  [[0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   ...,
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99215686, 0.99215686]],

  [[0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   ...,
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99215686, 0.99215686]],

  [[0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   ...,
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99215686, 0.99215686]],

  [[0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   ...,
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99215686, 0.99215686]],

  [[0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   ...,
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99215686, 0.99215686]],

  [[0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   ...,
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99215686, 0.99215686, 0.99215686]]], dtype=float32),
 'lowres': array([[ [0.99607843, 0.99607843, 0.99215686],
   [0.99215686, 0.99215686, 0.9882353 ],
   [0.99215686, 0.99215686, 0.99215686]],

  [[0.9843137 , 0.99215686, 0.9882353 ],
   [0.9843137 , 0.99215686, 0.9882353 ],
   [0.9882353 , 0.99215686, 0.99215686]],

  [[0.99607843, 0.99607843, 0.99215686],
   [0.99607843, 0.99607843, 0.99607843],
   [0.99607843, 0.99607843, 0.99607843]]],
```

```
adata.obs
```

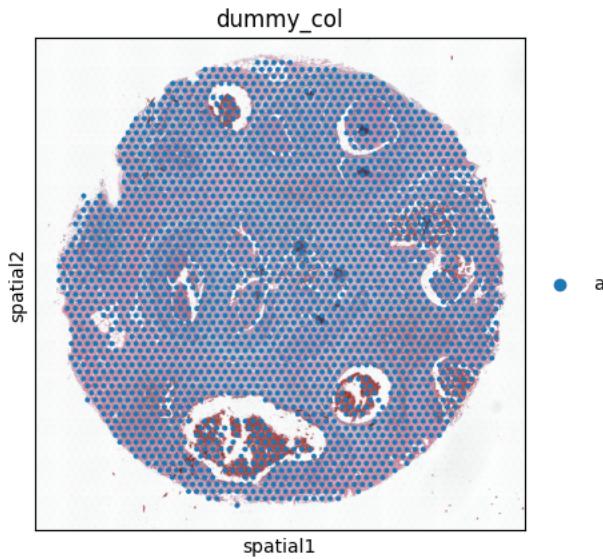
	in_tissue	array_row	array_col
AAACAAGTATCTCCA-1	1	50	102
AAACACCAATAACTGC-1	1	59	19
AAACAGAGCGACTCCT-1	1	14	94
AAACAGCTTCAGAAG-1	1	43	9
AAACAGGGTCTATATT-1	1	47	13
...
TTGTTCA GTGTGCTAC-1	1	24	64
TTGTTGTGTCAAGA-1	1	31	77
TTGTTTCACATCCAGG-1	1	58	42
TTGTTTCATTAGTCTA-1	1	60	30
TTGTTTCCATACAAC-1	1	45	27

2518 rows × 3 columns

Adding a dummy "constant" col to adata.obs so that we can see the spots on spatial map

```
adata.obs['dummy_col'] = 'a'
```

```
# plt.rcParams["figure.figsize"] = (8, 8)
sc.pl.spatial(adata, color = 'dummy_col')
```



▼ Step 1: QC Filtering

```
adata.var["mt"] = adata.var_names.str.startswith("MT-")
sc.pp.calculate_qc_metrics(adata, qc_vars=["mt"], inplace=True)
```

```
adata.obs
```

	in_tissue	array_row	array_col	dummy_col	n_genes_by_coun
AAACAAGTATCTCCCA-1	1	50	102	a	63
AAACACCAATAACTGC-1	1	59	19	a	31
AAACAGAGCGACTCCT-1	1	14	94	a	47
AAACAGCTTCAGAAG-1	1	43	9	a	29
AAACAGGGTCTATATT-1	1	47	13	a	25
...
TTGTTCAGTGTGCTAC-1	1	24	64	a	40
TTGTTGTGTCAAGA-1	1	31	77	a	56
TTGTTCACATCCAGG-1	1	58	42	a	19
TTGTTTCATTAGTCTA-1	1	60	30	a	38
TTGTTCCATACAAC-1	1	45	27	a	26

2518 rows × 15 columns

```

fig, axs = plt.subplots(1, 2, figsize=(15, 4))
sns.distplot(adata.obs["total_counts"], kde=False, ax=axs[0])
sns.distplot(adata.obs["n_genes_by_counts"], kde=False, bins=60, ax=axs[1])

plt.show()

```

```
<ipython-input-15-18138c9bfa86>:2: UserWarning:  
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.  
Please adapt your code to use either `displot` (a figure-level function with  
similar flexibility) or `histplot` (an axes-level function for histograms).
```

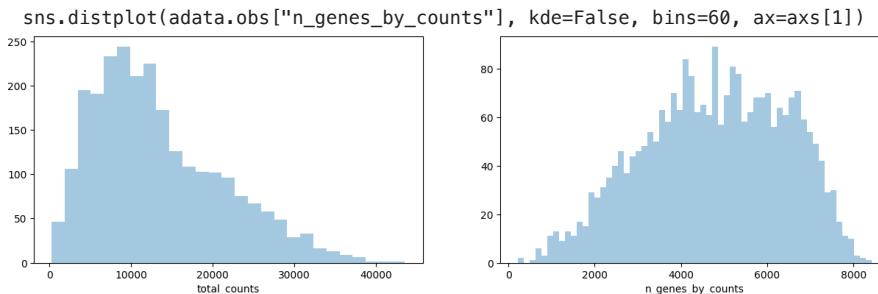
For a guide to updating your code to use the new functions, please see
<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(adata.obs["total_counts"], kde=False, ax=axs[0])  
<ipython-input-15-18138c9bfa86>:3: UserWarning:
```

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>



```
adata.obs['total_counts'] = adata.X.sum(axis=1)  
  
percentile_98 = np.percentile(adata.obs['total_counts'], 98)  
percentile_2 = np.percentile(adata.obs['total_counts'], 2)  
  
print(f"The 98th percentile of total counts is: {percentile_98}")  
print(f"The 2nd percentile of total counts is: {percentile_2}")
```

The 98th percentile of total counts is: 31979.659999999993
The 2nd percentile of total counts is: 1993.38

```
sc.pp.filter_cells(adata, min_counts = percentile_2)
```

```
sc.pp.filter_cells(adata, max_counts=percentile_98)
```

```
fig, axs = plt.subplots(1, 2, figsize=(15, 4))  
sns.distplot(adata.obs["total_counts"], kde=False, ax=axs[0])  
sns.distplot(adata.obs["n_genes_by_counts"], kde=False, bins=60, ax=axs[1])  
  
plt.show()
```

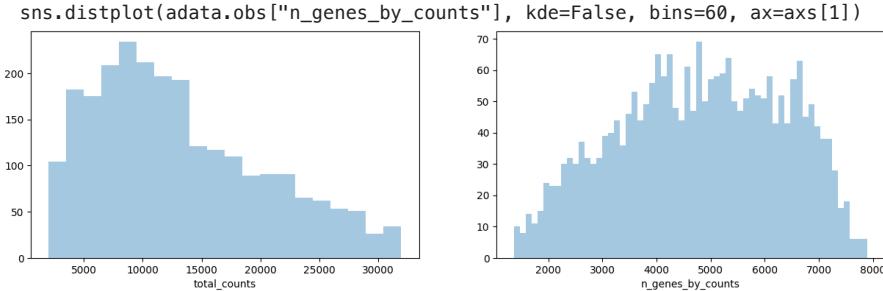
```
<ipython-input-19-18138c9bfa86>:2: UserWarning:  
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.  
Please adapt your code to use either `displot` (a figure-level function with  
similar flexibility) or `histplot` (an axes-level function for histograms).
```

For a guide to updating your code to use the new functions, please see
<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(adata.obs["total_counts"], kde=False, ax=axs[0])  
<ipython-input-19-18138c9bfa86>:3: UserWarning:  
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>



```
n_genes_by_cell = np.sum(adata.X > 0, axis=1)  
  
# Assign the calculated n_genes to the AnnData observations (cells)  
adata.obs['n_genes'] = n_genes_by_cell  
  
# Now calculate the 98th and 2nd percentiles for n_genes  
percentile_98_n_genes = np.percentile(adata.obs['n_genes'], 98)  
percentile_2_n_genes = np.percentile(adata.obs['n_genes'], 2)  
  
print(f"The 98th percentile of n_genes is: {percentile_98_n_genes}")  
print(f"The 2nd percentile of n_genes is: {percentile_2_n_genes}")  
  
The 98th percentile of n_genes is: 7354.4  
The 2nd percentile of n_genes is: 1849.0  
  
sc.pp.filter_cells(adata, min_genes = percentile_2_n_genes)  
  
sc.pp.filter_cells(adata, max_genes = percentile_98_n_genes)  
  
fig, axs = plt.subplots(1, 2, figsize=(15, 4))  
sns.distplot(adata.obs["total_counts"], kde=False, ax=axs[0])  
sns.distplot(adata.obs["n_genes_by_counts"], kde=False, bins=60, ax=axs[1])  
plt.show()
```

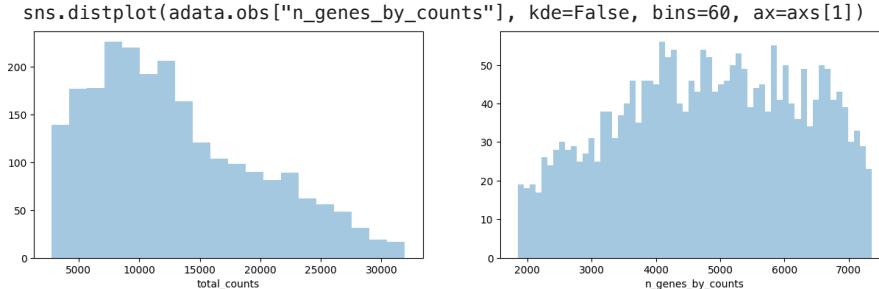
```
<ipython-input-23-18138c9bfa86>:2: UserWarning:  
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.  
Please adapt your code to use either `displot` (a figure-level function with  
similar flexibility) or `histplot` (an axes-level function for histograms).
```

For a guide to updating your code to use the new functions, please see
<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

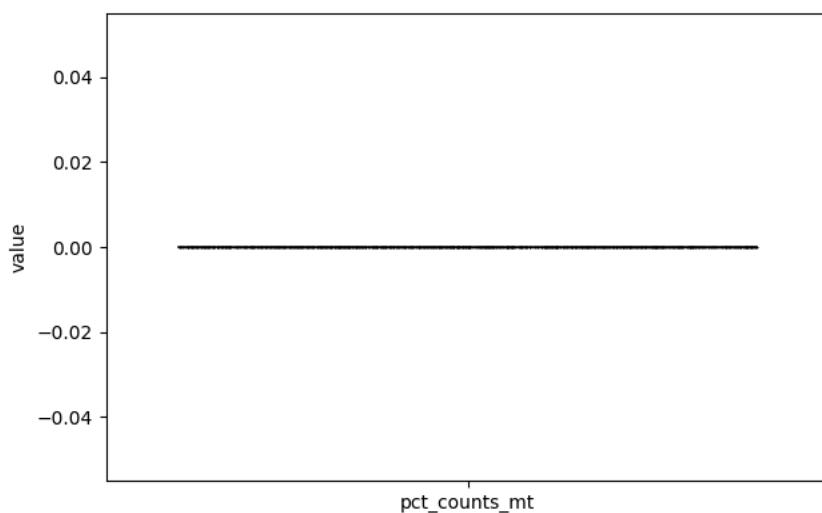
```
sns.distplot(adata.obs["total_counts"], kde=False, ax=axs[0])  
<ipython-input-23-18138c9bfa86>:3: UserWarning:  
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>



```
sc.pl.violin(adata, ['pct_counts_mt'], jitter=0.4)  
  
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning:  
The `scale` parameter has been renamed and will be removed in v0.15.0. Pass `den  
ax = sns.violinplot()
```



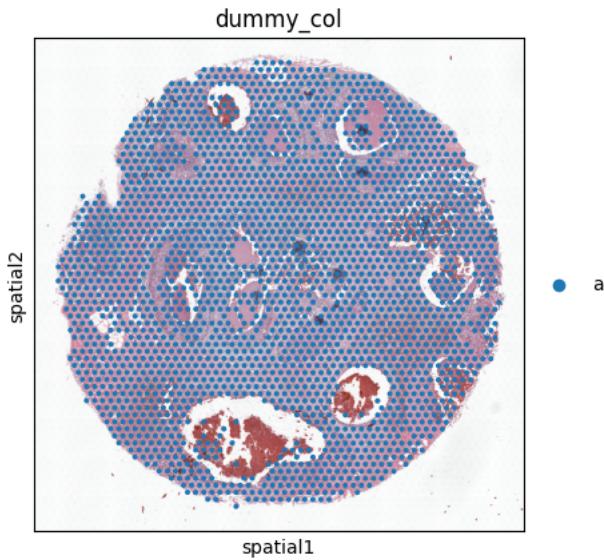
```
sc.pp.filter_genes(adata, min_cells=3)
```

```
adata
```

```
AnnData object with n_obs × n_vars = 2318 × 16672  
obs: 'in_tissue', 'array_row', 'array_col', 'dummy_col', 'n_genes_by_counts', 'log1p_n_genes_by_counts',
```

```
'total_counts', 'log1p_total_counts', 'pct_counts_in_top_50_genes', 'pct_counts_in_top_100_genes',
'pct_counts_in_top_200_genes', 'pct_counts_in_top_500_genes', 'total_counts_mt', 'log1p_total_counts_mt', 'pct_counts_mt',
'n_counts', 'n_genes'
var: 'gene_ids', 'feature_types', 'genome', 'mt', 'n_cells_by_counts', 'mean_counts', 'log1p_mean_counts',
'pct_dropout_by_counts', 'total_counts', 'log1p_total_counts', 'n_cells'
uns: 'spatial', 'dummy_col_colors'
obsm: 'spatial'
```

```
sc.pl.spatial(adata, color = 'dummy_col')
```



▼ Step 2: Normalize and scale data

```
sc.pp.normalize_total(adata, inplace=True)
sc.pp.log1p(adata)
```

▼ Step 3 and 4: PCA & Leiden

```
adata.raw = adata #save normalized data to use later in lianapy
```

```
sc.pp.highly_variable_genes(adata, flavor="seurat", n_top_genes=2000)
```

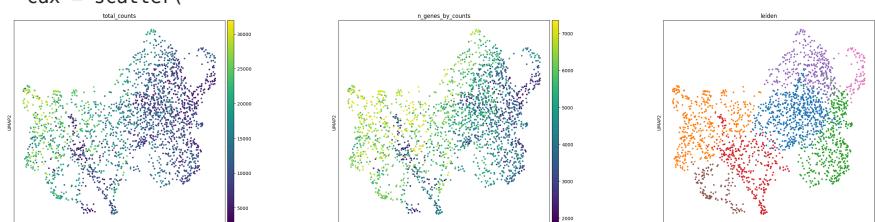
```
sc.pp.pca(adata)
sc.pp.neighbors(adata)
sc.tl.umap(adata)
sc.tl.leiden(adata, resolution=0.5)
```

```
adata
```

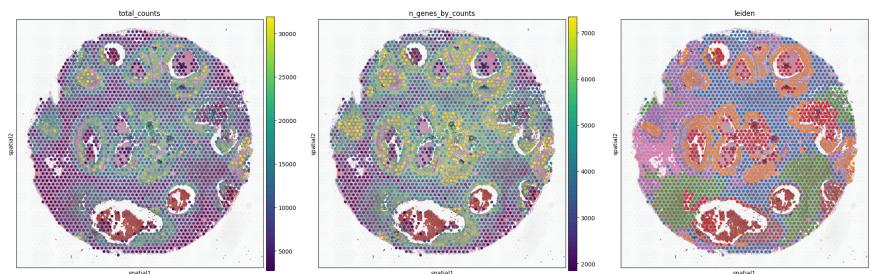
```
AnnData object with n_obs × n_vars = 2318 × 16672
  obs: 'in_tissue', 'array_row', 'array_col', 'dummy_col', 'n_genes_by_counts', 'log1p_n_genes_by_counts',
  'total_counts', 'log1p_total_counts', 'pct_counts_in_top_50_genes', 'pct_counts_in_top_100_genes',
  'pct_counts_in_top_200_genes', 'pct_counts_in_top_500_genes', 'total_counts_mt', 'log1p_total_counts_mt', 'pct_counts_mt',
  'n_counts', 'n_genes', 'leiden'
  var: 'gene_ids', 'feature_types', 'genome', 'mt', 'n_cells_by_counts', 'mean_counts', 'log1p_mean_counts',
  'pct_dropout_by_counts', 'total_counts', 'log1p_total_counts', 'n_cells', 'highly_variable', 'means', 'dispersions',
  'dispersions_norm'
  uns: 'spatial', 'dummy_col_colors', 'log1p', 'hvg', 'pca', 'neighbors', 'umap', 'leiden'
  obsm: 'spatial', 'X_pca', 'X_umap'
  varm: 'PCs'
  obsp: 'distances', 'connectivities'
```

```
plt.rcParams["figure.figsize"] = (8, 8)
sc.pl.umap(adata, color=["total_counts", "n_genes_by_counts", "leiden"], wspace=0.4)
```

```
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_tools/scatterplots.py:3
cax = scatter'
```



```
plt.rcParams["figure.figsize"] = (8, 8)
sc.pl.spatial(adata, img_key="hires", color=["total_counts", "n_genes_by_counts", "leiden"])
```



▼ Step 5: Finding Markers for Clusters and Cell-type annotation

▼ Manual Annotation

```
sc.tl.rank_genes_groups(adata, 'leiden', method='wilcoxon')

#convert to dataframe
results = adata.uns['rank_genes_groups']
('0', '1', '2', '3', '4')

out = np.array([[0,0,0,0,0]])
for group in results['names'].dtype.names:
    out = np.vstack((out, np.vstack((results['names'][group],
                                    results['scores'][group],
                                    results['pvals_adj'][group],
                                    results['logfoldchanges'][group],
                                    np.array([group] * len(results['names'][group])).astype('object'))).T))

markers = pd.DataFrame(out[1:], columns = ['Gene', 'scores', 'pval_adj', 'lfc', 'cluster'])
markers = markers[(markers.pval_adj < 0.05) & (abs(markers.lfc) > 1)]
```

markers

	Gene	scores	pval_adj	lfc	cluster
0	AEBP1	21.732389	0.0	1.514522	0
1	CCN2	20.906864	0.0	1.428344	0
2	SFRP4	20.101196	0.0	1.72928	0
3	SFRP2	20.093746	0.0	1.469333	0
4	CXCL14	19.795189	0.0	1.74175	0
...
116699	AZGP1	-11.100204	0.0	-1.935027	6
116700	SPINT2	-11.196445	0.0	-2.399995	6
116701	ERBB2	-11.421288	0.0	-2.523322	6
116702	XBP1	-11.455423	0.0	-1.486808	6
116703	FTH1	-12.169497	0.0	-1.14535	6

8982 rows × 5 columns

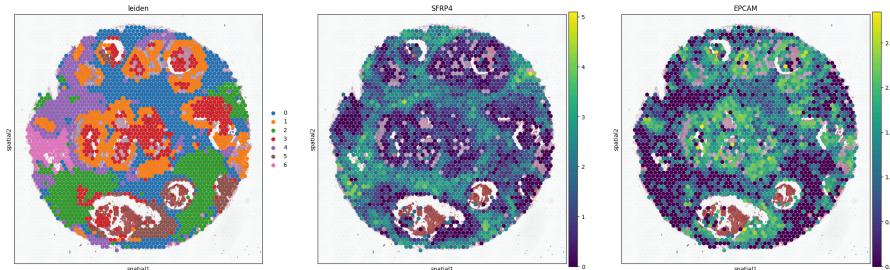
▼ Annotating Cell Type for Cluster 0

markers[markers.cluster == '0']

	Gene	scores	pval_adj	lfc	cluster
0	AEBP1	21.732389	0.0	1.514522	0
1	CCN2	20.906864	0.0	1.428344	0
2	SFRP4	20.101196	0.0	1.72928	0
3	SFRP2	20.093746	0.0	1.469333	0
4	CXCL14	19.795189	0.0	1.74175	0
...
16471	APOC2	-5.272581	0.000004	-1.275777	0
16505	ADGRF1	-5.527629	0.000001	-1.024545	0
16618	AQP9	-7.021473	0.0	-1.800967	0
16631	ANGPTL4	-7.293883	0.0	-1.069606	0
16655	C15orf48	-8.41431	0.0	-1.059094	0

208 rows × 5 columns

sc.pl.spatial(adata, img_key="hires", color=["leiden", "SFRP4", "EPCAM"], size=1.5)



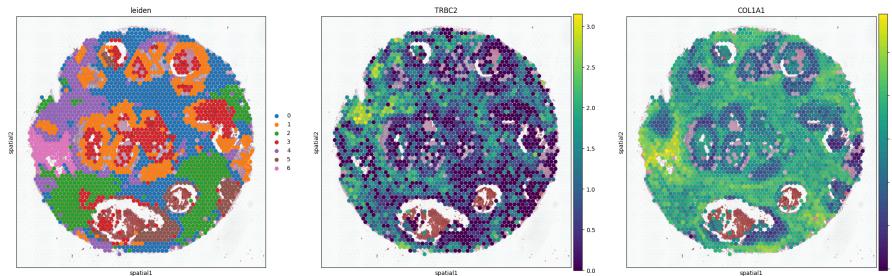
❖ Annotating Cell Type for Cluster 3

```
markers[markers.cluster == '3']
```

	Gene	scores	pval_adj	lfc	cluster
50016	SPP1	22.649845	0.0	3.187531	3
50017	CTSL	16.755529	0.0	1.593313	3
50018	SLC11A1	16.501286	0.0	2.263902	3
50019	PLIN2	16.346254	0.0	1.63611	3
50024	IFI30	14.857883	0.0	1.248809	3
...
66683	C3	-13.393376	0.0	-1.21542	3
66684	SFRP2	-13.449463	0.0	-1.402559	3
66685	DCN	-13.564318	0.0	-1.178585	3
66686	C1R	-14.785619	0.0	-1.210216	3
66687	C1S	-15.075749	0.0	-1.462258	3

285 rows × 5 columns

```
sc.pl.spatial(adata, img_key="hires", color=["leiden", "TRBC2", "COL1A1"], size=1.5,)
```



❖ DecoupleR for cell type annotation

```
# Cell-type annotation using DecoupleR
```

```
import decoupler as dc
```

```
# Query Omnipath and get PanglaoDB
markers = dc.get_resource('PanglaoDB')
markers
```

56.5k/? [00:00<00:00, 1.88MB/s]

5.91M/? [00:00<00:00, 15.0MB/s]

	genesymbol	canonical_marker	cell_type	germ_layer	human	human_sensitivity
0	CTRB1	False	Enterocytes	Endoderm	True	0.0000
1	CTRB1	True	Acinar cells	Endoderm	True	1.0000
2	KLK1	True	Endothelial cells	Mesoderm	True	0.0000
3	KLK1	False	Goblet cells	Endoderm	True	0.5882
4	KLK1	False	Epithelial cells	Mesoderm	True	0.0000
...
8456	SLC14A1	True	Urothelial cells	Mesoderm	True	0.0000
8457	UPK3A	True	Urothelial cells	Mesoderm	True	0.0000
8458	UPK1A	True	Urothelial cells	Mesoderm	True	0.0000
8459	UPK2	True	Urothelial cells	Mesoderm	True	0.0000
8460	UPK3B	True	Urothelial cells	Mesoderm	True	0.0000

8461 rows × 13 columns

```
markers = markers[markers['human'] & markers['canonical_marker']]
markers = markers[~markers.duplicated(['cell_type', 'genesymbol'])]
markers
```

	genesymbol	canonical_marker	cell_type	germ_layer	human	human_sensitivity
1	CTRB1	True	Acinar cells	Endoderm	True	1.000
2	KLK1	True	Endothelial cells	Mesoderm	True	0.000
5	KLK1	True	Principal cells	Mesoderm	True	0.000
6	KLK1	True	Acinar cells	Endoderm	True	0.833
7	KLK1	True	Plasmacytoid dendritic cells	Mesoderm	True	0.000
...
8456	SLC14A1	True	Urothelial cells	Mesoderm	True	0.000
8457	UPK3A	True	Urothelial cells	Mesoderm	True	0.000
8458	UPK1A	True	Urothelial cells	Mesoderm	True	0.000
8459	UPK2	True	Urothelial cells	Mesoderm	True	0.000
8460	UPK3B	True	Urothelial cells	Mesoderm	True	0.000

5180 rows × 13 columns

```
dc.run_ora(
    mat=adata,
    net=markers,
    source='cell_type',
    target='genesymbol',
    min_n=3,
    verbose=True,
    use_raw=False
)
```

Running ora on mat with 2318 samples and 16672 targets for 147 sources.
100%|██████████| 2318/2318 [00:11<00:00, 201.63it/s]

adata

```
AnnData object with n_obs × n_vars = 2318 × 16672
  obs: 'in_tissue', 'array_row', 'array_col', 'dummy_col', 'n_genes_by_counts', 'log1p_n_genes_by_counts',
  'total_counts', 'log1p_total_counts', 'pct_counts_in_top_50_genes', 'pct_counts_in_top_100_genes',
  'pct_counts_in_top_200_genes', 'pct_counts_in_top_500_genes', 'total_counts_mt', 'log1p_total_counts_mt',
  'pct_counts_mt', 'n_counts', 'n_genes', 'leiden'
  var: 'gene_ids', 'feature_types', 'genome', 'mt', 'n_cells_by_counts', 'mean_counts', 'log1p_mean_counts',
  'pct_dropout_by_counts', 'total_counts', 'log1p_total_counts', 'n_cells', 'highly_variable', 'means', 'dispersions',
  'dispersions_norm'
  uns: 'spatial', 'dummy_col_colors', 'log1p', 'hvg', 'pca', 'neighbors', 'umap', 'leiden', 'leiden_colors',
  'rank_genes_groups'
  obsm: 'spatial', 'X_pca', 'X_umap', 'ora_estimate', 'ora_pvals'
  varm: 'PCs'
  obsp: 'distances', 'connectivities'
```

adata.obsm["ora_estimate"]

source	Acinar cells	Endothelial cells	Principal cells	Plasmacytoid dendritic cells	Paneth cells	Eps c
AAACAAGTATCTCCA-1	0.861578	0.332016	-0.000000	0.489842	0.141690	0.29
AAACACCAATAACTGC-1	0.428465	5.279730	-0.000000	7.281368	0.892899	0.06
AAACAGAGCGACTCCT-1	0.428465	1.763840	-0.000000	5.467467	-0.000000	1.07
AAACAGCTTCAGAAG-1	2.065867	4.253485	-0.000000	2.450819	0.892899	0.06
AAACAGGGTCTATATT-1	0.134749	0.677480	-0.000000	3.120091	0.141690	0.63
...
TTGTTCAGTGTGCTAC-1	0.861578	1.157248	0.287868	1.846029	0.892899	0.29
TTGTTGTGTCAAGA-1	0.428465	4.253485	-0.000000	1.846029	0.892899	0.29
TTGTTCACATCCAGG-1	0.861578	2.487946	0.287868	0.060043	0.892899	0.06
TTGTTTCATTAGTCTA-1	0.861578	7.588376	-0.000000	0.856887	1.458790	0.63
TTGTTCCATACAAC-1	0.861578	5.825885	-0.000000	1.312103	0.141690	1.07

2318 rows × 147 columns

acts = dc.get_acts(adata, obsm_key='ora_estimate')

```
# We need to remove inf and set them to the maximum value observed for pvals=0
acts_v = acts.X.ravel()
max_e = np.nanmax(acts_v[np.isfinite(acts_v)])
acts.X[~np.isfinite(acts.X)] = max_e
```

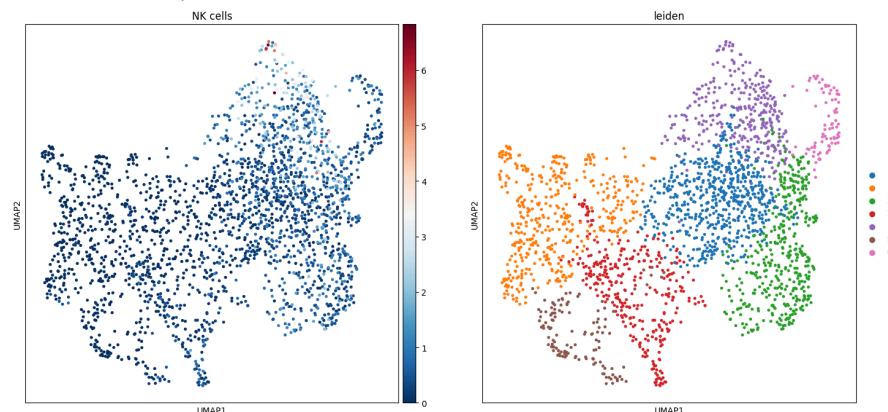
acts

```
/usr/local/lib/python3.10/dist-packages/anndata/_core/aligned_df.py:67: ImplicitModificationWarning: Transforming to str inc
  warnings.warn("Transforming to str index.", ImplicitModificationWarning)
AnnData object with n_obs × n_vars = 2318 × 147
  obs: 'in_tissue', 'array_row', 'array_col', 'dummy_col', 'n_genes_by_counts', 'log1p_n_genes_by_counts',
  'total_counts', 'log1p_total_counts', 'pct_counts_in_top_50_genes', 'pct_counts_in_top_100_genes',
```

```
'pct_counts_in_top_200_genes', 'pct_counts_in_top_500_genes', 'total_counts_mt', 'log1p_total_counts_mt', 'pct_counts_mt',
'n_counts', 'n_genes', 'leiden'
    uns: 'spatial', 'dummy_col_colors', 'log1p', 'hvg', 'pca', 'neighbors', 'umap', 'leiden', 'leiden_colors',
'rank_genes_groups'
    obsm: 'spatial', 'X_pca', 'X_umap', 'ora_estimate', 'ora_pvals'
```

```
sc.pl.umap(acts, color=['NK cells', 'leiden'], cmap='RdBu_r')
sc.pl.violin(acts, keys=['NK cells'], groupby='leiden')
```

/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_tools/scatterplots.py:3
cax = scatter'



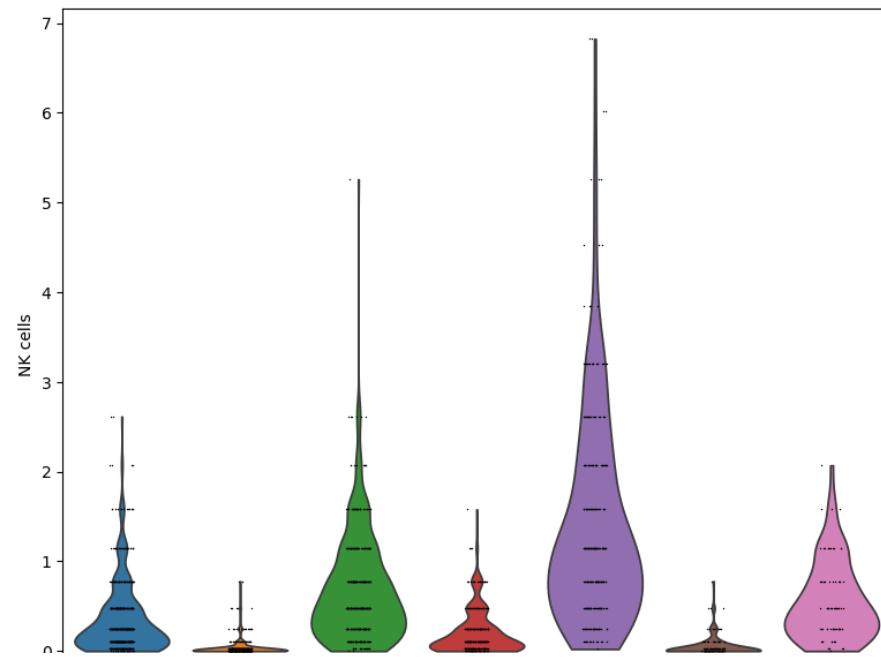
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v

```
ax = sns.violinplot(
```

/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning:

The `scale` parameter has been renamed and will be removed in v0.15.0. Pass `den
ax = sns.violinplot(



Annotation

```
df = dc.rank_sources_groups(acts, groupby='leiden', reference='rest', method='t-test_overestim_var')
df
```

group	reference		names	statistic	meanchange	pvals	pvals_i
0	0	rest	Fibroblasts	25.208591	9.014468	3.552751e-109	5.22254
1	0	rest	Pancreatic stellate cells	21.964306	3.734860	4.610375e-88	3.388626e
2	0	rest	Hepatic stellate cells	18.988140	3.541342	1.773330e-68	8.689316e
3	0	rest	Tanocytes	15.949577	0.613102	4.185389e-52	1.538131e
4	0	rest	Myofibroblasts	15.677723	1.640234	1.434940e-50	4.218723e
...
1024	6	rest	Ductal cells	-9.287530	-2.544502	7.720115e-15	3.782856e
1025	6	rest	Cholangiocytes	-9.594819	-2.215967	2.971984e-15	2.184408e
1026	6	rest	Anterior pituitary gland cells	NaN	0.000000	1.000000e+00	1.000000e

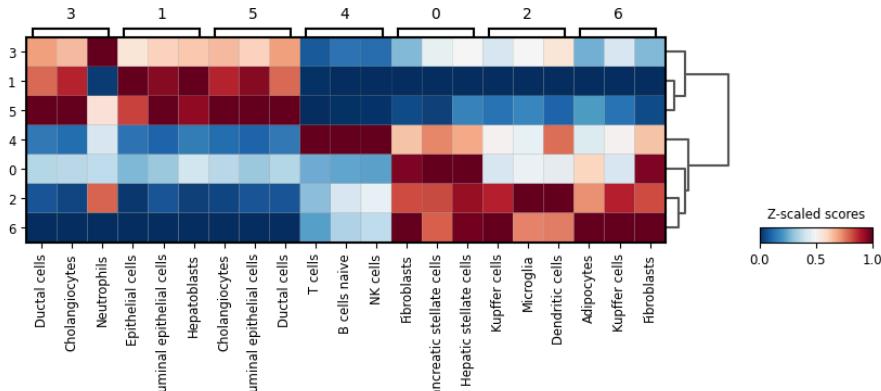
Burkinis

```
n_ctypes = 3
ctypes_dict = df.groupby('group').head(n_ctypes).groupby('group')['names'].apply(lambda x: list(x)).to_dict()
ctypes_dict

{'0': ['Fibroblasts', 'Pancreatic stellate cells', 'Hepatic stellate cells'],
 '1': ['Epithelial cells', 'Luminal epithelial cells', 'Hepatoblasts'],
 '2': ['Kupffer cells', 'Microglia', 'Dendritic cells'],
 '3': ['Ductal cells', 'Cholangiocytes', 'Neutrophils'],
 '4': ['T cells', 'B cells naive', 'NK cells'],
 '5': ['Cholangiocytes', 'Luminal epithelial cells', 'Ductal cells'],
 '6': ['Adipocytes', 'Kupffer cells', 'Fibroblasts']}
```

```
sc.pl.matrixplot(acts, ctypes_dict, 'leiden', dendrogram=True, standard_scale='var',
                 colorbar_title='Z-scaled scores', cmap='RdBu_r')
```

WARNING: dendrogram data not found (using key=dendrogram_leiden). Running `sc.tl



```
sc.pl.violin(acts, keys=['T cells', 'B cells', 'Platelets', 'Monocytes', 'NK cells'], groupby='leiden')
```

```
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in v
```

```
    ax = sns.violinplot()
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
The `scale` parameter has been renamed and will be removed in v0.15.0. Pass `den
    ax = sns.violinplot()
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in v
```

```
    ax = sns.violinplot()
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
The `scale` parameter has been renamed and will be removed in v0.15.0. Pass `den
    ax = sns.violinplot()
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in v
```

```
    ax = sns.violinplot()
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
The `scale` parameter has been renamed and will be removed in v0.15.0. Pass `den
    ax = sns.violinplot()
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in v
```

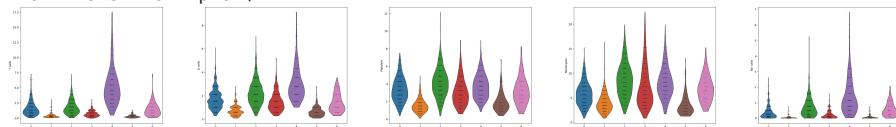
```
    ax = sns.violinplot()
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
The `scale` parameter has been renamed and will be removed in v0.15.0. Pass `den
    ax = sns.violinplot()
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in v
```

```
    ax = sns.violinplot()
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_anndata.py:843: FutureWarning
```

```
The `scale` parameter has been renamed and will be removed in v0.15.0. Pass `den
    ax = sns.violinplot()
```



```
annotation_dict = df.groupby('group').head(1).set_index('group')['names'].to_dict()
```

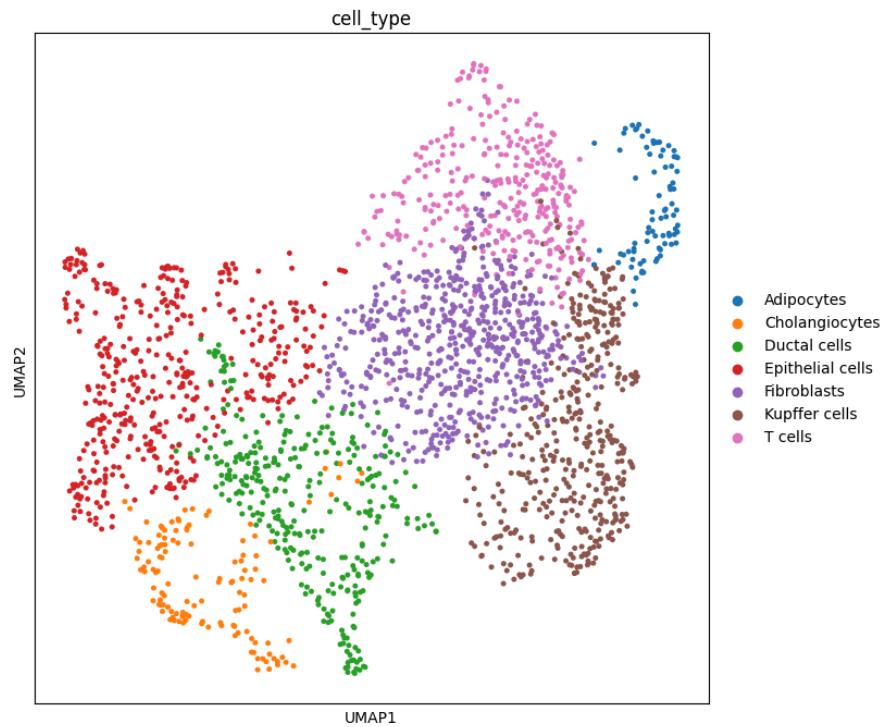
```
annotation_dict
```

```
{'0': 'Fibroblasts',
 '1': 'Epithelial cells',
 '2': 'Kupffer cells',
 '3': 'Ductal cells',
 '4': 'T cells',
 '5': 'Cholangiocytes',
 '6': 'Adipocytes'}
```

```
# Add cell type column based on annotation
adata.obs['cell_type'] = [annotation_dict[clust] for clust in adata.obs['leiden']]
```

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

```
/usr/local/lib/python3.10/dist-packages/scanpy/plotting/_tools/scatterplots.py:3  
    cax = scatter(
```



✓ Step 6: Finding Markers of Metastasis/Cancer

```
sc.pl.spatial(adata, img_key="hires", color=tumor_genes, size=1.5)
```

