

concat_1_EX05

August 15, 2022

```
[1]: import scarf
      scarf.__version__
```

```
[1]: '0.20.0'
```

```
[2]: reader = scarf.CSVReader('concat_1_EX05.csv', skip_rows=341,
      ↪ skip_cols=['Time'], cell_data_cols=['SampleID'])
      reader.nCells, reader.nFeatures
```

```
Performing CSV file consistency check: 0%|
```

```
↪
```

```
...
```

```
[2]: (500000, 28)
```

```
[3]: reader.feature_ids()
```

```
[3]: array(['FSC-A', 'FSC-H', 'FSC-W', 'SSC-A', 'SSC-H', 'SSC-W',
          'Comp-APC-A :: EPCR', 'Comp-APC-Cy7-A :: B220',
          'Comp-Alexa Fluor 700-A :: CD48', 'Comp-BUV395-A :: CD19',
          'Comp-BUV496-A :: CD43', 'Comp-BUV563-A :: CD105',
          'Comp-BUV615-A :: CD11B', 'Comp-BUV661-A :: CD93',
          'Comp-BUV737-A :: CD5', 'Comp-BUV805-A :: CD24',
          'Comp-BV421-A :: FLT3', 'Comp-BV510-A :: CD45',
          'Comp-BV605-A :: CD150', 'Comp-BV650-A :: BP1',
          'Comp-BV711-A :: CD41', 'Comp-BV786-A :: KIT',
          'Comp-FITC-A :: IGM', 'Comp-PE-A :: IL7R', 'Comp-PE-Cy5-A :: LIN',
          'Comp-PE-Cy7-A :: SCA', 'Comp-PE-Texas Red-A :: CD25',
          'Comp-PerCP-Cy5-5-A :: IGD'], dtype=object)
```

```
[4]: writer = scarf.CSVtoZarr(reader, zarr_fn='concat_1_EX05.zarr',
      ↪ assay_name='ADT', chunk_size=(10000, 100))
```

```
[5]: writer.dump()
```

```
0%|
```

```
↪
```

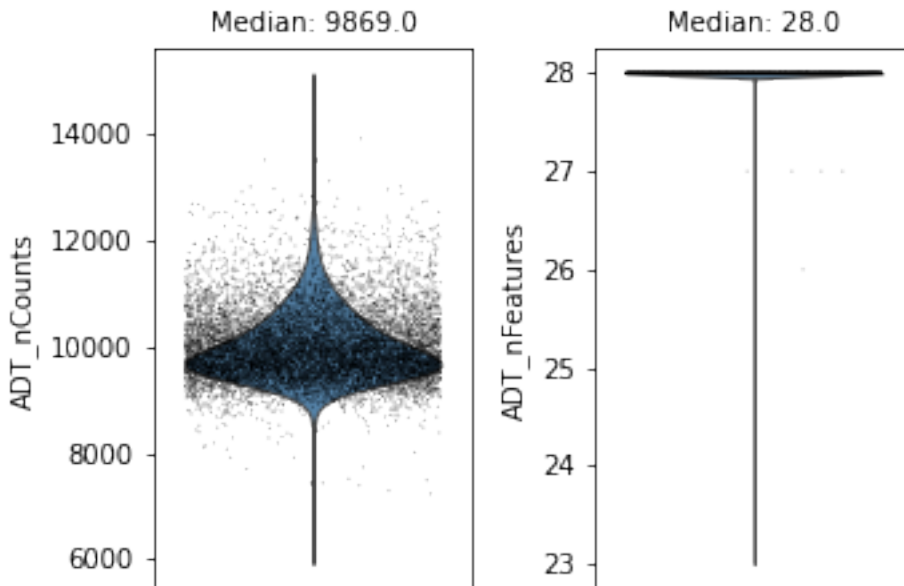
```
...
```

```
[7]: ds = scarf.DataStore('concat_1_EX05.zarr', min_features_per_cell=0, nthreads=8)
      ds
```

```
(ADT) Computing nCells and dropOuts: 0%|
↪
...
(ADT) Computing nCounts: 0%|
↪
...
(ADT) Computing nFeatures: 0%|
↪
...
```

```
[7]: DataStore has 500000 (500000) cells with 1 assays: ADT
      Cell metadata:
          'I', 'ids', 'names', 'ADT_nCounts', 'ADT_nFeatures',
          'SampleID'
      ADT assay has 28 (28) features and following metadata:
          'I', 'ids', 'names', 'dropOuts', 'nCells',
```

```
[8]: ds.plot_cells_dists()
```



```
[9]: ds.filter_cells(
      attrs=['ADT_nCounts', 'ADT_nFeatures'],
      lows=[8500, 27],
      highs=[12000, 29], reset_previous=True
    )
```

```
INFO: 7182 cells flagged for filtering out using attribute ADT_nCounts
INFO: 271 cells flagged for filtering out using attribute ADT_nFeatures
```

```
[10]: ds.make_graph(feats_key='I', k=11, dims=0)
```

```
Writing data to normed__I__I/data: 0%|
↪
Calculating mean of norm. data: 0%|
↪
Calculating std. dev. of norm. data: 0%|
↪
Fitting ANN: 0%|
↪
Fitting kmeans: 0%|
↪
Estimating seed partitions: 0%|
↪
Saving KNN graph: 0%|
↪
Smoothing KNN distances: 0%|
↪
INFO: ANN recall: 99.74%
```

```
[12]: ds.run_leiden_clustering(resolution=0.6)
```

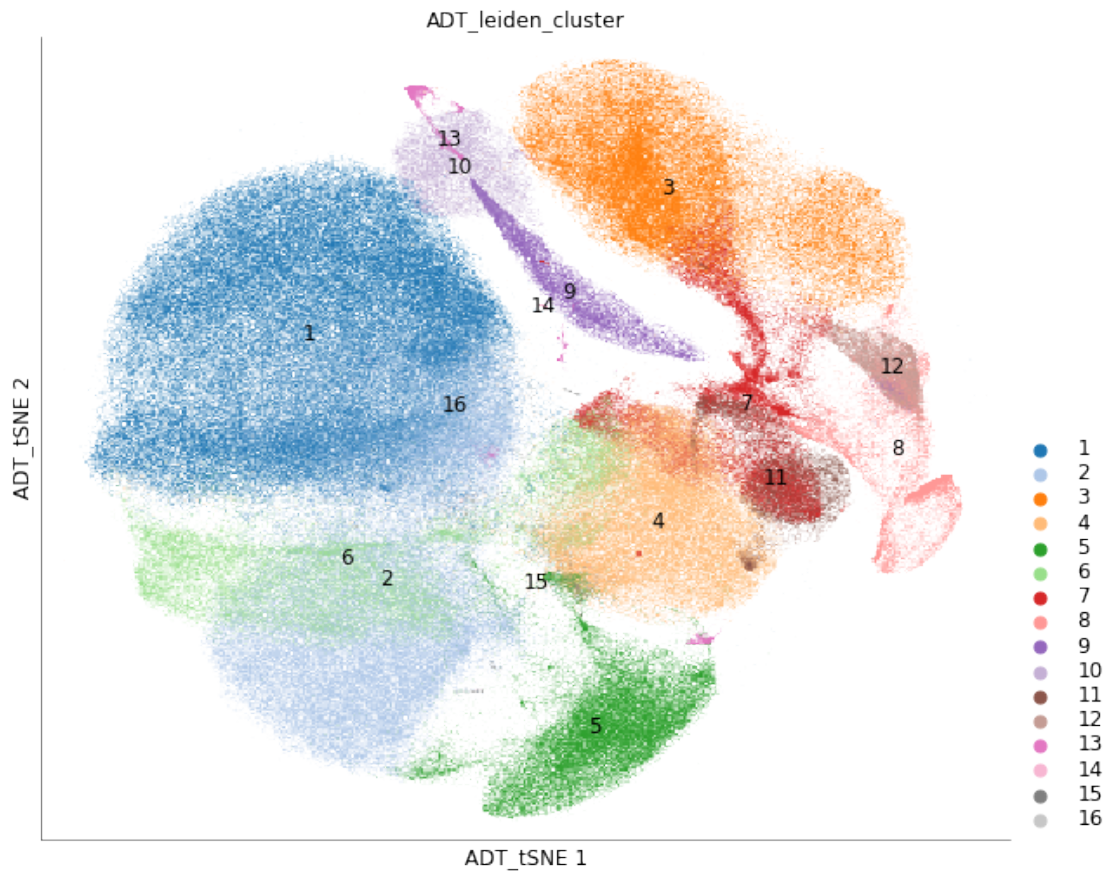
```
[13]: ds.run_tsne(
    alpha=10,
    box_h=0.7,
    early_iter=200,
    max_iter=1000,
    parallel=True,
    tsne_dims = 3
)
```

```
Saving KNN matrix in MTX format: 0%|
↪
INFO: b'Number of vertices: 492569'
INFO: b'Embedding dimensions: 3'
INFO: b'Rescaling parameter \xc3\xbb: 1'
INFO: b'Early exag. multiplier \xc3\xb1: 10'
INFO: b'Maximum iterations: 1000'
INFO: b'Early exag. iterations: 200'
INFO: b'Box side length h: 0.7'
```

INFO: b'Drop edges originating from leaf nodes? 0'
INFO: b'Number of processes: 8'
INFO: b'3031 out of 492569 nodes already stochastic'
INFO: b'Skipping \xce\xbb rescaling...'
INFO: b'Nested dissection permutation...Permuting matrixDONE'
INFO: b'm = 492569| n = 492569| nnnz = 8688570'
INFO: b'Working with double precision'
INFO: b'Iteration 1: error is 138.076'
INFO: b'Iteration 50: error is 136.664 (50 iterations in 8.74478 seconds)'
INFO: b'Iteration 100: error is 133.174 (50 iterations in 8.5332 seconds)'
INFO: b'Iteration 150: error is 127.238 (50 iterations in 8.32837 seconds)'
INFO: b'Iteration 200: error is 10.1641 (50 iterations in 8.42361 seconds)'
INFO: b'Iteration 250: error is 9.97632 (50 iterations in 8.52773 seconds)'
INFO: b'Iteration 300: error is 9.49745 (50 iterations in 8.30929 seconds)'
INFO: b'Iteration 350: error is 9.06293 (50 iterations in 8.31169 seconds)'
INFO: b'Iteration 400: error is 8.73115 (50 iterations in 8.31135 seconds)'
INFO: b'Iteration 450: error is 8.48473 (50 iterations in 8.36582 seconds)'
INFO: b'Iteration 500: error is 8.29158 (50 iterations in 8.61829 seconds)'
INFO: b'Iteration 550: error is 8.13362 (50 iterations in 14.5324 seconds)'
INFO: b'Iteration 600: error is 7.99967 (50 iterations in 14.5796 seconds)'
INFO: b'Iteration 650: error is 7.88376 (50 iterations in 17.1728 seconds)'
INFO: b'Iteration 700: error is 7.78136 (50 iterations in 15.0299 seconds)'
INFO: b'Iteration 750: error is 7.68964 (50 iterations in 14.6829 seconds)'
INFO: b'Iteration 800: error is 7.60671 (50 iterations in 19.9858 seconds)'
INFO: b'Iteration 850: error is 7.53113 (50 iterations in 15.871 seconds)'
INFO: b'Iteration 900: error is 7.46164 (50 iterations in 25.6615 seconds)'
INFO: b'Iteration 950: error is 7.3975 (50 iterations in 27.4658 seconds)'
INFO: b'Iteration 999: error is 7.33898 (50 iterations in 16.6433 seconds)'

```
INFO: b'--- Time spent in each module ---'  
INFO: b''  
INFO: b'Attractive forces: 22.5259 sec [8.86344%] | Repulsive forces:  
231.618 sec [91.1366%]'  
INFO: b'Saving embedding to:  
/projects/fs1/stefanl/FACS/b2077755-ca2f-4c3b-9f33-6a67736e686d_output.txt'
```

```
[14]: ds.plot_layout(  
    layout_key='ADT_tsNE',  
    color_by='ADT_leiden_cluster',  
    do_shading=True,  
    width=10,  
    height=10,  
    shade_npixels=400  
)
```



```
[15]: ds.run_tsne(  
    alpha=10,  
    box_h=0.7,
```

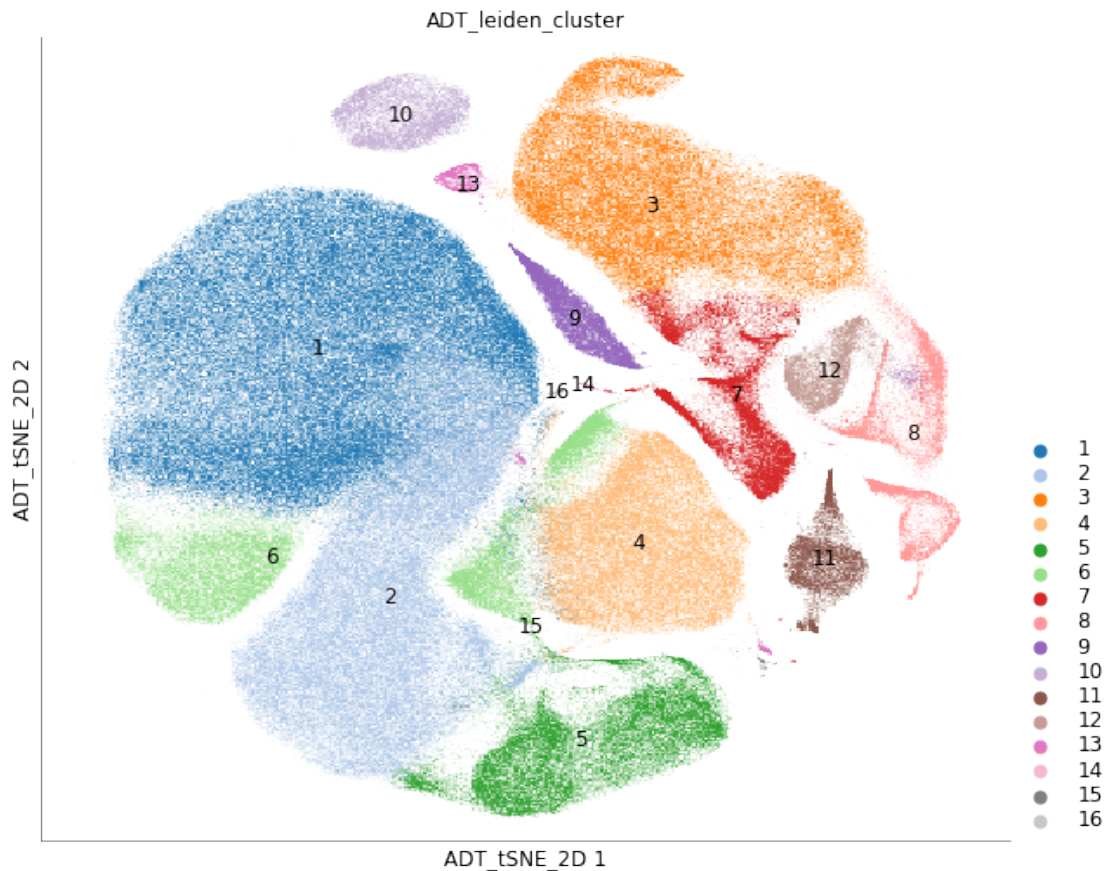
```
early_iter=200,  
max_iter=1000,  
parallel=True,  
tsne_dims = 2,  
label="tSNE_2D"  
)
```

Saving KNN matrix in MTX format: 0%| □
↪ ...

```
INFO: b'Number of vertices: 492569'  
INFO: b'Embedding dimensions: 2'  
INFO: b'Rescaling parameter \xce\xbb: 1'  
INFO: b'Early exag. multiplier \xce\xbb: 10'  
INFO: b'Maximum iterations: 1000'  
INFO: b'Early exag. iterations: 200'  
INFO: b'Box side length h: 0.7'  
INFO: b'Drop edges originating from leaf nodes? 0'  
INFO: b'Number of processes: 8'  
INFO: b'3031 out of 492569 nodes already stochastic'  
INFO: b'Skipping \xce\xbb rescaling...'  
INFO: b'Nested dissection permutation...Permuting matrixDONE'  
INFO: b'm = 492569| n = 492569| nnnz = 8688570'  
INFO: b'Working with double precision'  
INFO: b'Iteration 1: error is 138.443'  
INFO: b'Iteration 50: error is 136.848 (50 iterations in 3.64743  
seconds)'  
INFO: b'Iteration 100: error is 133.301 (50 iterations in 3.4809  
seconds)'  
INFO: b'Iteration 150: error is 127.186 (50 iterations in 3.4924  
seconds)'  
INFO: b'Iteration 200: error is 10.1991 (50 iterations in 3.3722  
seconds)'  
INFO: b'Iteration 250: error is 10.0042 (50 iterations in 3.26535  
seconds)'  
INFO: b'Iteration 300: error is 9.54615 (50 iterations in 3.30691  
seconds)'  
INFO: b'Iteration 350: error is 9.193 (50 iterations in 3.28112  
seconds)'  
INFO: b'Iteration 400: error is 8.92784 (50 iterations in 3.2837  
seconds)'  
INFO: b'Iteration 450: error is 8.72196 (50 iterations in 3.35932  
seconds)'  
INFO: b'Iteration 500: error is 8.55457 (50 iterations in 3.68299  
seconds)'  
INFO: b'Iteration 550: error is 8.41368 (50 iterations in 4.86626  
seconds)'  
INFO: b'Iteration 600: error is 8.292 (50 iterations in 5.28054
```

```
seconds)'
INFO: b'Iteration 650: error is 8.18466 (50 iterations in 5.32493
seconds)'
INFO: b'Iteration 700: error is 8.08894 (50 iterations in 4.48396
seconds)'
INFO: b'Iteration 750: error is 8.00234 (50 iterations in 4.66571
seconds)'
INFO: b'Iteration 800: error is 7.92335 (50 iterations in 4.80354
seconds)'
INFO: b'Iteration 850: error is 7.85081 (50 iterations in 5.08401
seconds)'
INFO: b'Iteration 900: error is 7.78432 (50 iterations in 4.8928
seconds)'
INFO: b'Iteration 950: error is 7.72278 (50 iterations in 4.50569
seconds)'
INFO: b'Iteration 999: error is 7.66693 (50 iterations in 4.33196
seconds)'
INFO: b'--- Time spent in each module ---'
INFO: b''
INFO: b'Attractive forces: 11.4426 sec [15.3167%] | Repulsive forces:
63.2642 sec [84.6833%]'
INFO: b'Saving embedding to:
/projects/fs1/stefanl/FACS/c02177b5-6ff4-41ca-9ba3-ddd9a4e1d82d_output.txt'
```

```
[16]: ds.plot_layout(
    layout_key='ADT_tSNE_2D',
    color_by='ADT_leiden_cluster',
    do_shading=True,
    width=10,
    height=10,
    shade_npixels=400
)
```



```
[17]: ds
```

```
[17]: DataStore has 492569 (500000) cells with 1 assays: ADT
      Cell metadata:
          'I', 'ids', 'names', 'ADT_leiden_cluster', 'ADT_nCounts',
          'ADT_nFeatures', 'ADT_tSNE1', 'ADT_tSNE2', 'ADT_tSNE3',
'ADT_tSNE_2D1',
          'ADT_tSNE_2D2', 'SampleID'
      ADT assay has 28 (28) features and following metadata:
          'I', 'ids', 'names', 'dropOuts', 'nCells',
```

```
[18]: scarf.writers.to_h5ad(
      assay=ds.ADT,
      h5ad_filename='concat_1_EX05.h5ad'
    )
```

0%|



...




```
[19]: adata = ds.to_anndata()
```

```
INFO: Converting raw data from ADT assay into CSR format: 0%|
```

```
↪
```

```
...
```

```
[20]: adata
```

```
[20]: AnnData object with n_obs × n_vars = 492569 × 28  
      obs: 'I', 'names', 'ADT_leiden_cluster', 'ADT_nCounts', 'ADT_nFeatures',  
          'ADT_tSNE1', 'ADT_tSNE2', 'ADT_tSNE3', 'ADT_tSNE_2D1', 'ADT_tSNE_2D2',  
          'SampleID'  
      var: 'I', 'gene_ids', 'dropOuts', 'nCells'
```

```
[22]: import anndata
```

```
[23]: adata2 = anndata.read('concat_1_EX05.h5ad')
```

```
[25]: adata2
```

```
[25]: AnnData object with n_obs × n_vars = 500000 × 28  
      obs: '_index'  
      var: '_index', 'gene_short_name'
```

```
[ ]:
```