cNMF Package Tutorial Walkthrough

**Introduction**

The purpose of this report is to run through the cNMF package (1) tutorial for running the cNMF python package through R. Primarily, this is to see if I can actually run the program, followed by comparing my analysis with the authors and trying to understand the logic behind conclusions.

**Note regarding installation:**

The analysis I went through was run in a mamba environment set up specifically for using the cNMF package. This ended up taking multiple requiring multiple attempts due to conflicts between package installations. The mamba environment was set up in Ubuntu using Windows Subsystem for Linux (WSL) with the following package versions:

|  |  |  |
| --- | --- | --- |
| **Package** | **Version** | **Command** |
| Python | 3.10.16 | mamba install -c conda-forge python = 3.10.16 |
| Scikit-learn | 1.3.2 | mamba install -c conda-forge scikit-learn=1.3 |
| scanpy | 1.9.8 | mamba install -c bioconda scanpy=1.9 |
| anndata | 11.3 | Installed with on the prior packages |
| R | 4.4.2 | mamba install -c conda-forge r-base=4.4.2 |
| Seurat | 5.1.0 | mamba install -c conda-forge r-seurat=5.1.0 |
| cnmf | 1.6.0 | mamba install bioconda::cnmf |
| tidyverse | 2.0.0 | mamba install conda-forge::r-tidyverse |
| openxlsx | 4.2.8 | mamba install conda-forge::r-openxlsx |

The data for the tutorial is the PBMC 3K dataset from 10X genomics, which can be downloaded using the following link: <https://cf.10xgenomics.com/samples/cell/pbmc3k_filtered_gene_bc_matrices.tar.gz>

In addition to running through the cNMF analysis tutorial I also analyzed the data set using Seurat. I’ll compare my conclusions with what was previously described by the authors to run through my rationale.