Notes about velocity workshop in Python (from Jihe)

Plan:

* Teach velocity (in Python) as a 3 h module
* **Add required knowledge to the current basic Python module**
  + **How to import library, how to get help**
  + **How to read in files**

Velocity analysis workflow (we don’t need to teach all steps; we could supply files from step1 and 2, and focus on step 3)

1. Shell - Get input matrices (unspliced counts, spliced counts) with **velocyto**
   1. Provide script for how to generate the matrices
2. R/Python - Get your clusters/reducedDimensions etc. either from your favorite R or Python workflow
   1. Convert Rds to python version h5ad (scanpy already use h5ad)
3. Python - Merge the information for **scVelo** velocity analysis
   1. **Teach velocity basic, but also mention dynamic modeling (longer to run, more accurate method)**

Dataset:

* **Example dataset from scVelo**
  + [**https://scvelo.readthedocs.io/vignettes/**](https://scvelo.readthedocs.io/vignettes/) **(perfect data example)**
* **PBMC data from the scRNAseq workshop?**
* Tseng (Mary)
  + Pro: clear biology; published
  + Con: velocity not done yet; need to retrieve data somewhere
* Mark Feinberg (Preeti/Mary)
  + Pro: velocity already done (Mary’s notebook demo)
  + Con: not yet published? Not clear cell type identification

Current materials:

* <https://github.com/hbc/tutorials/blob/93ae2a8e82be516ddd5b5311bda1c3bd8ae91f65/scRNAseq/scRNAseq_analysis_tutorial/lessons/velocity.md>
* <https://smorabit.github.io/tutorials/8_velocyto/>

No need to teach these concepts in our current basic Python module:

* Pandas (AnnData is based on pandas, but we don’t use it)
* Matplotlib
* Numpy

Improvement of the basic Python module (based on qualtric survey feedback):

* Demonstrate application in biology - something that cannot be done on excel
  + Dataset from R
  + Plotting