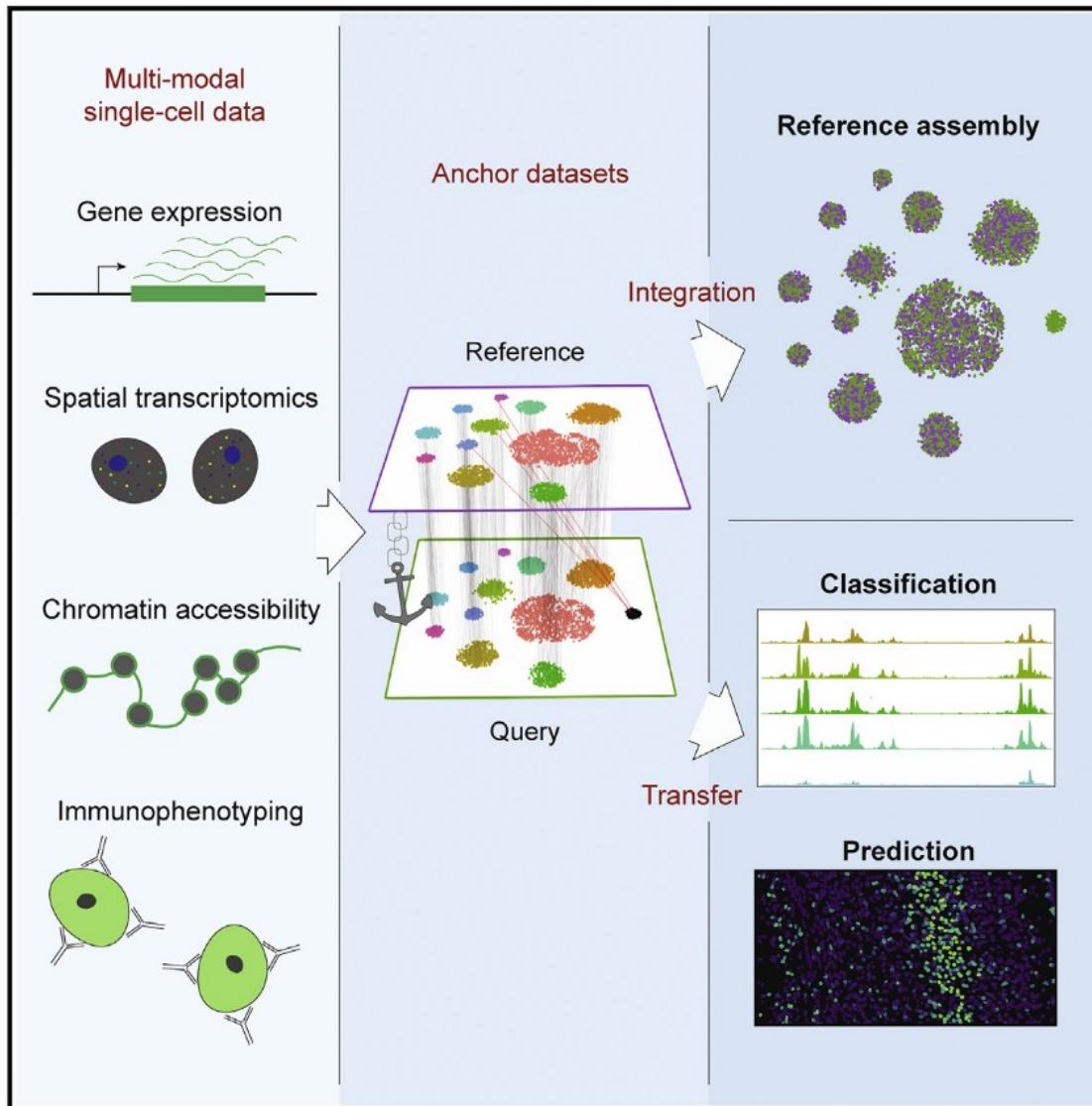


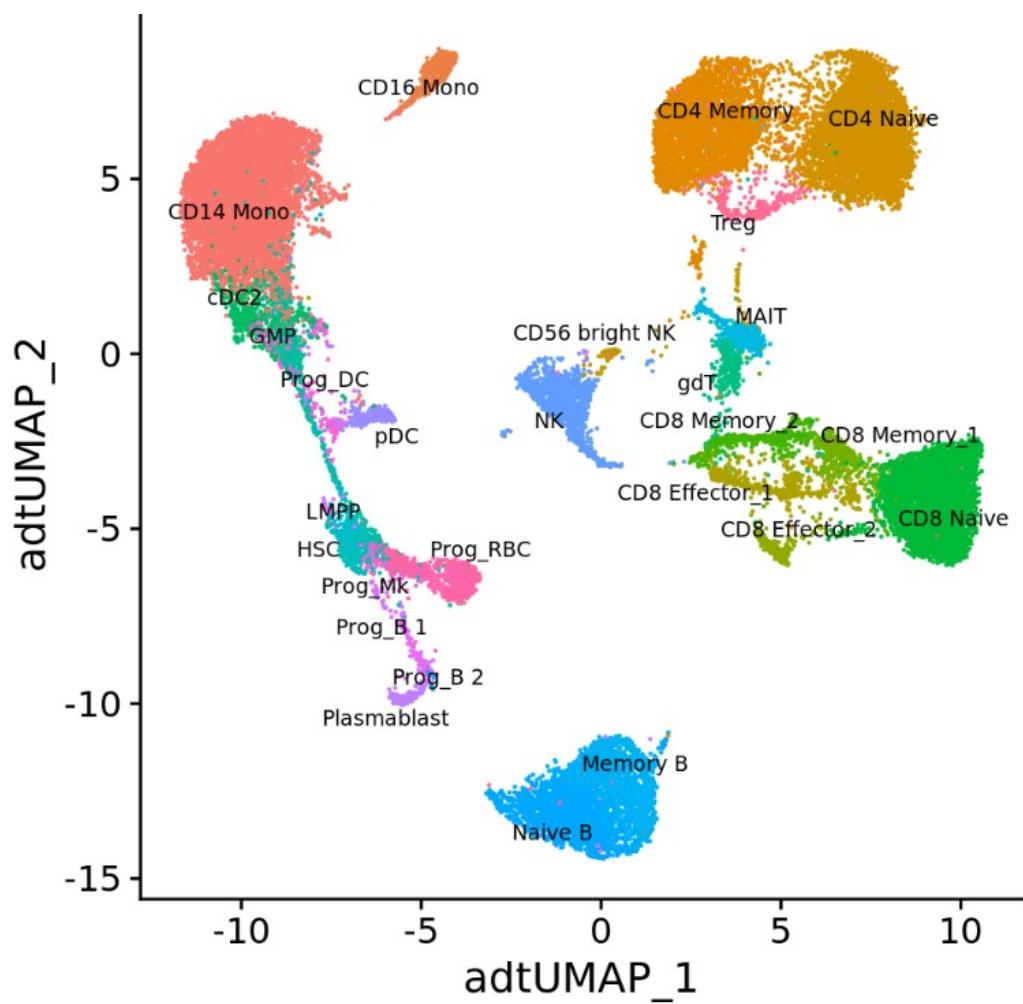
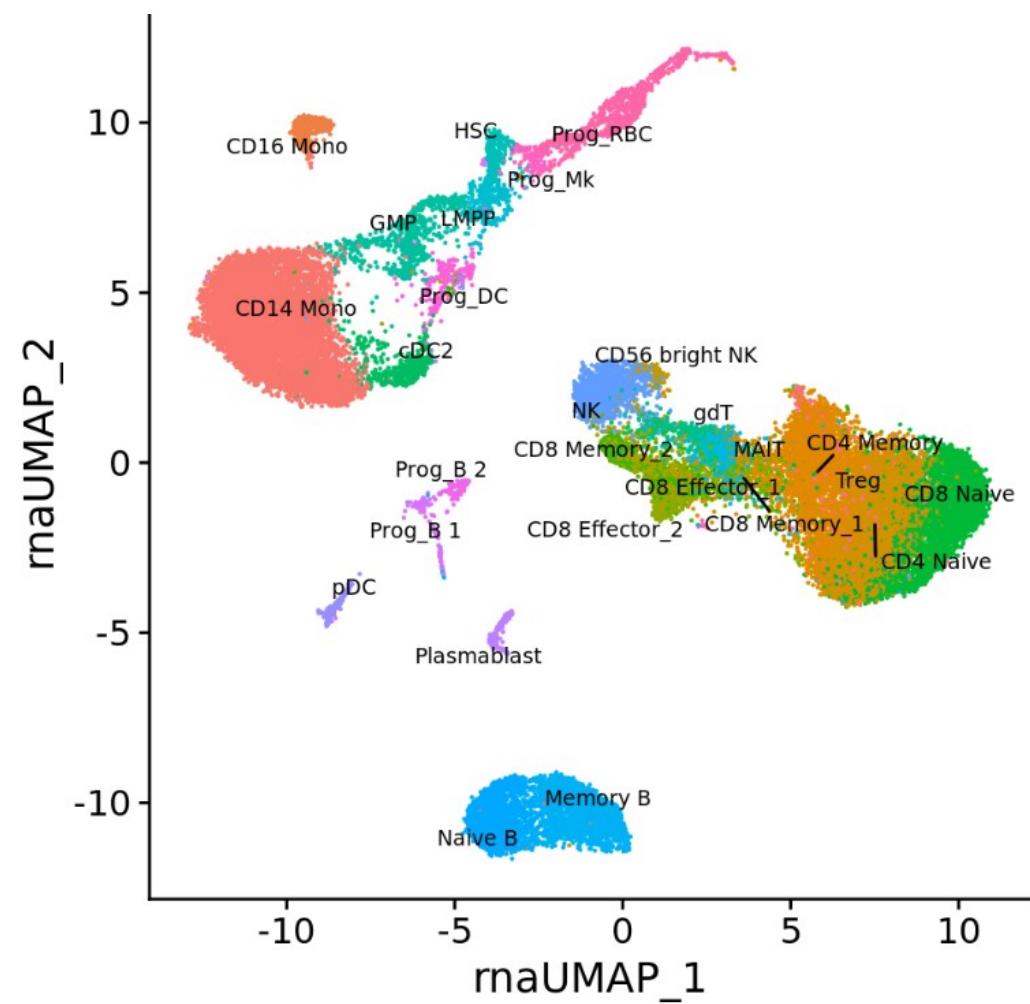
Multimodal Single Cell OMICs Integration

Single Cell Omics Course 2021

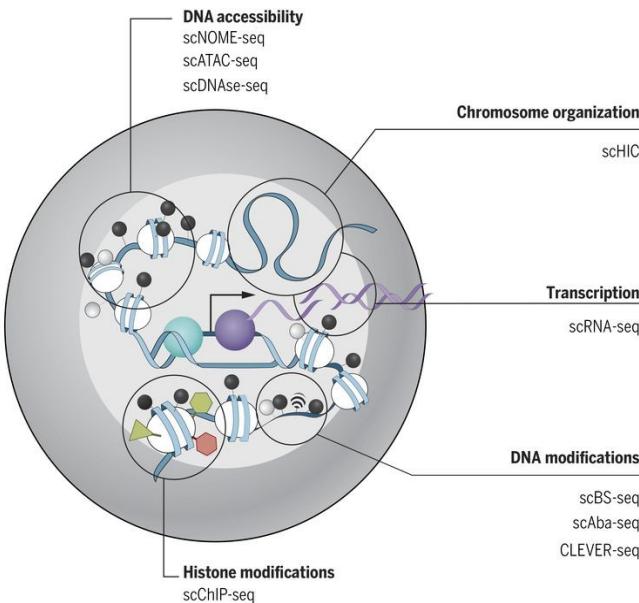
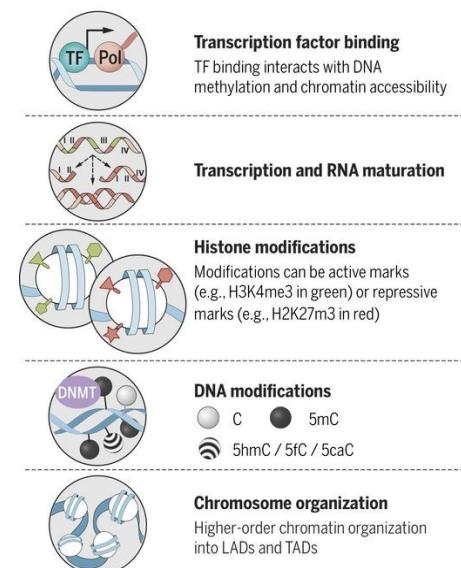
Nikolay Oskolkov, NBIS SciLifeLab

Lund, 29.01.2021

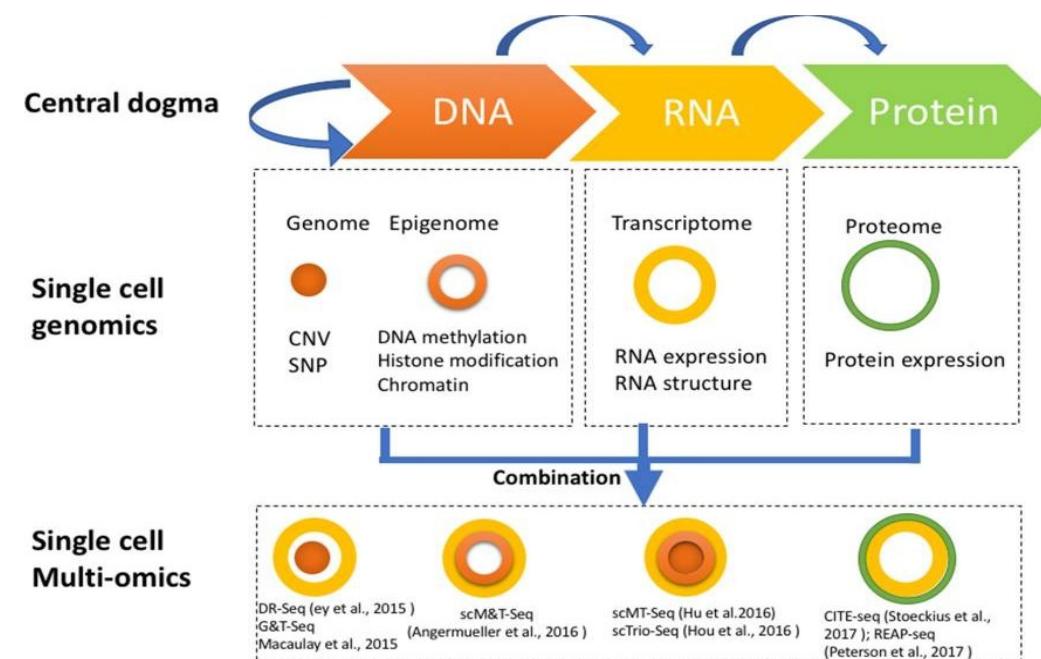




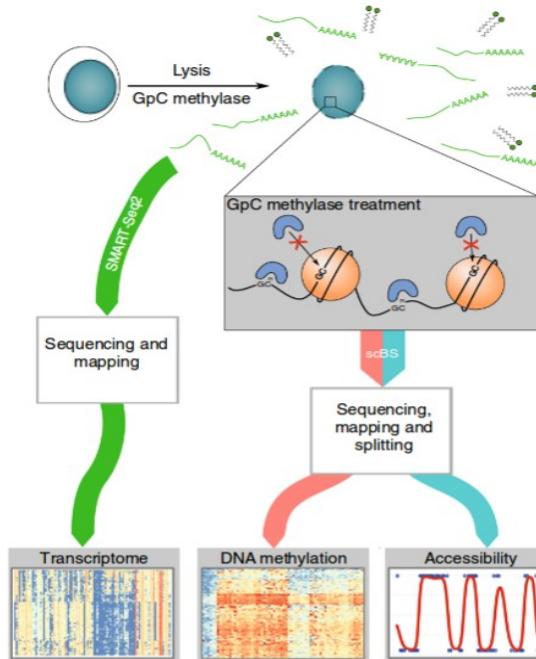
“Despite tremendous functional diversity, distinct populations of T cells such as effector, regulatory, $\gamma\delta$, and mucosal associated invariant T (MAIT), often cannot be effectively separated by scRNA-seq alone, even when using the most sensitive and cutting-edge technologies”



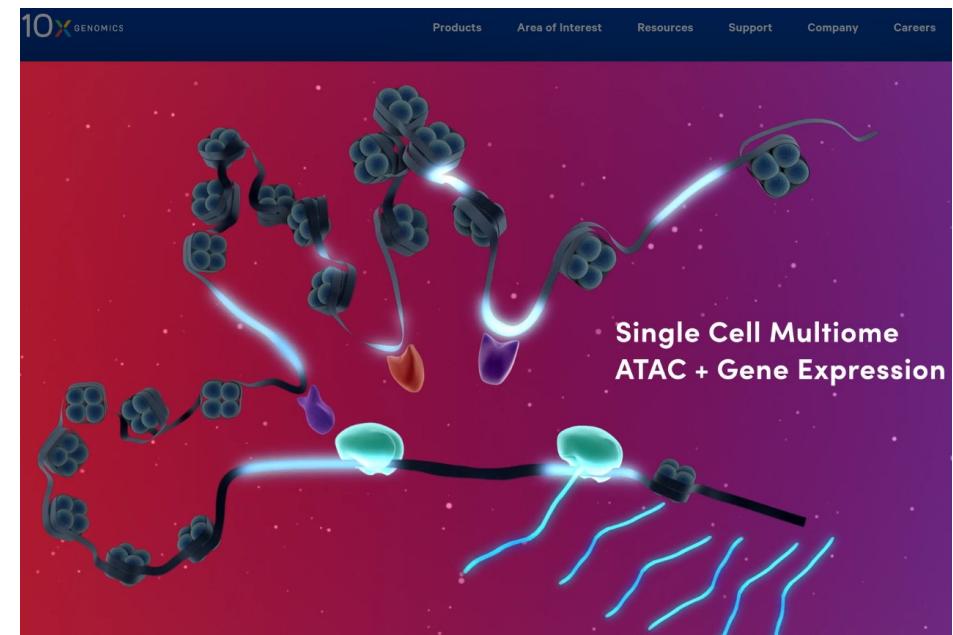
Kelsey et al., 2017, Science 358, 69-75



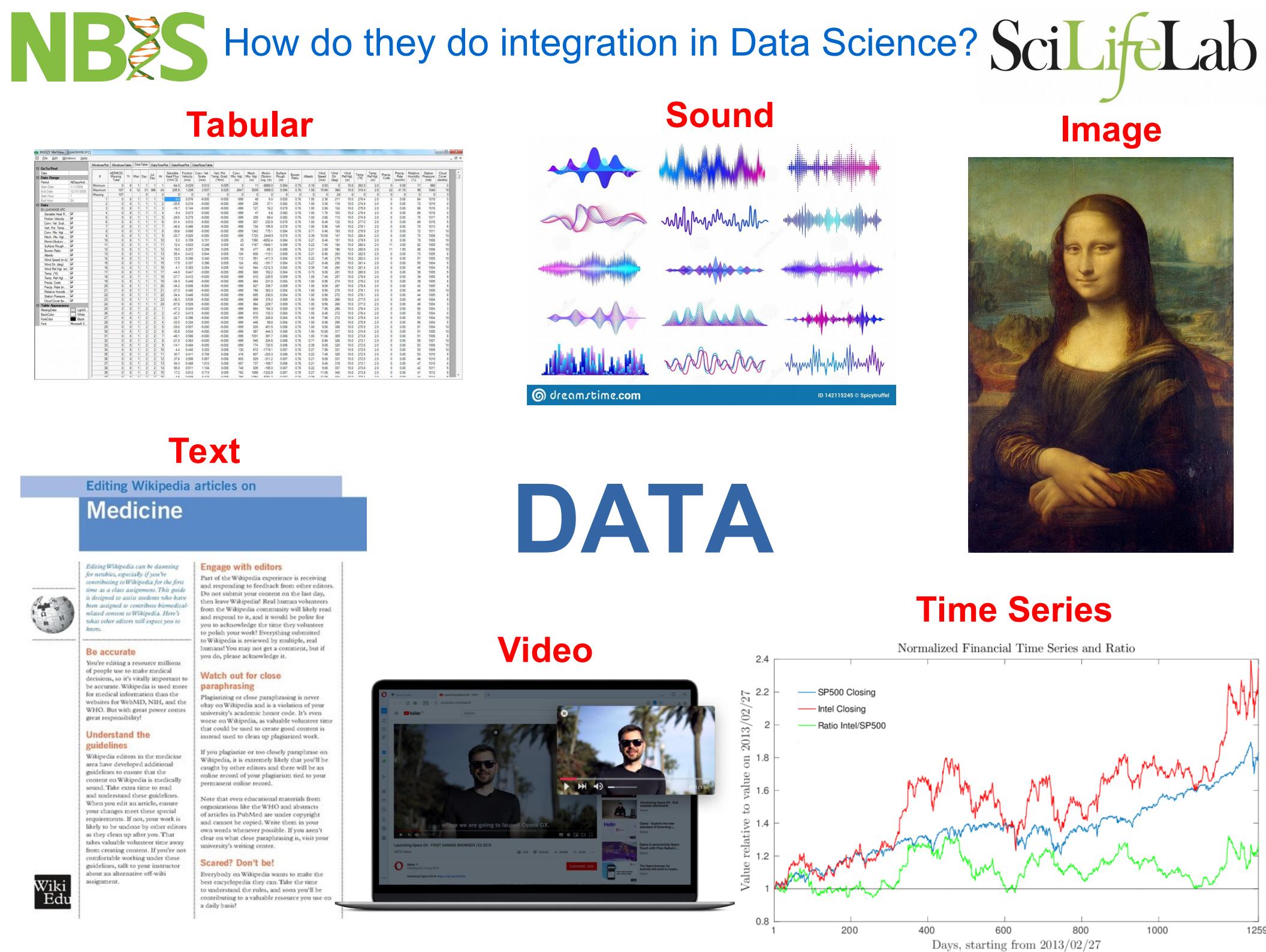
Hu et al., 2018, Frontier in Cell and Developmental Biology 6, 1-13



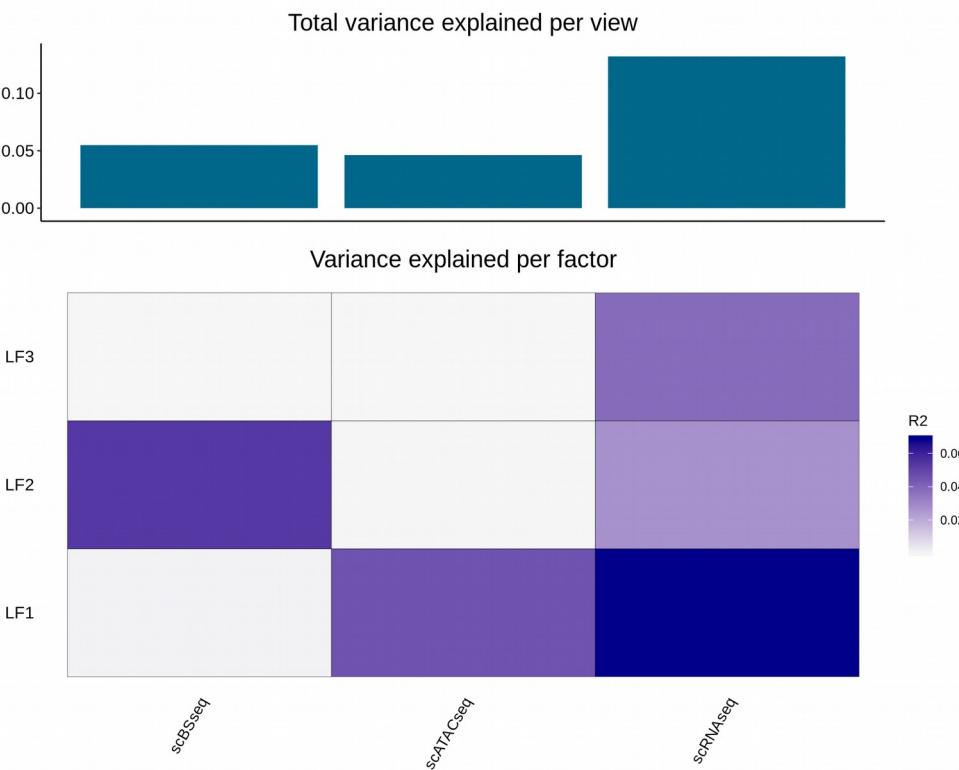
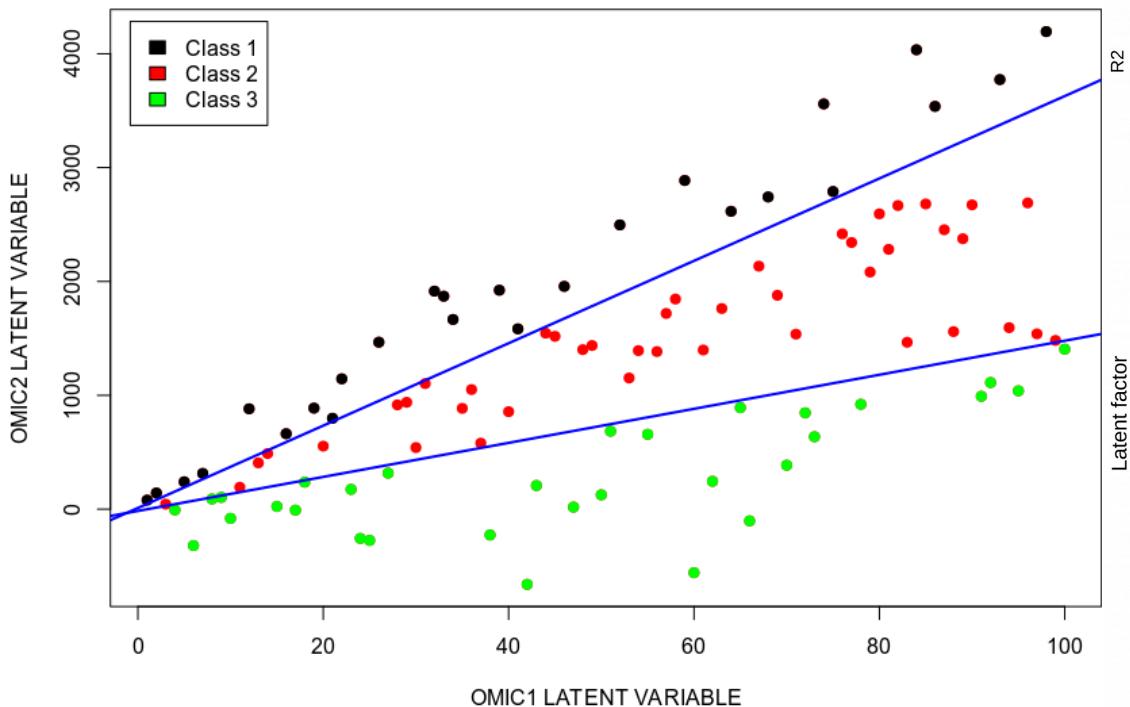
Clark et al., 2018, Nature Communications 9, 781



10X Genomics Multiome ATAC + Gene Expression



Idea Behind OMICs Integration:
See Patterns Hidden in Individual OMICS



“Most importantly, simultaneous analysis of multiple modalities should **improve** on the ability to discover and characterize cell states, compared to independent analyses of each modality when performed separately”

How I Evaluate OMICs Integration, Data Science: Boost in Prediction

TEXT (78%)

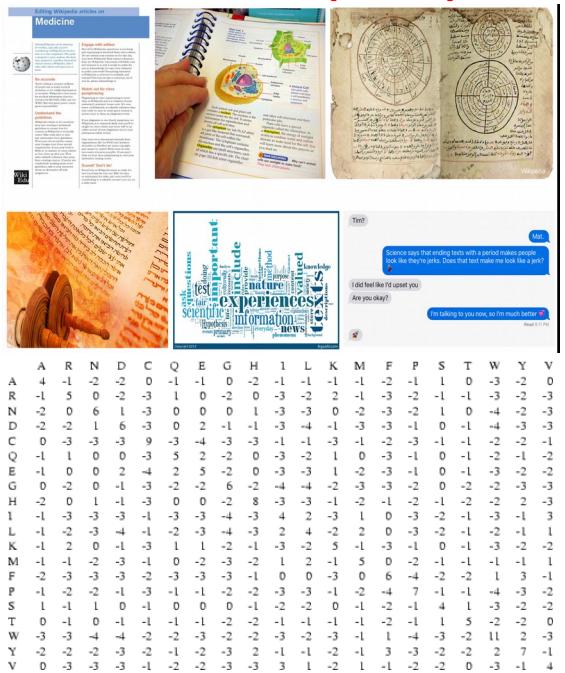
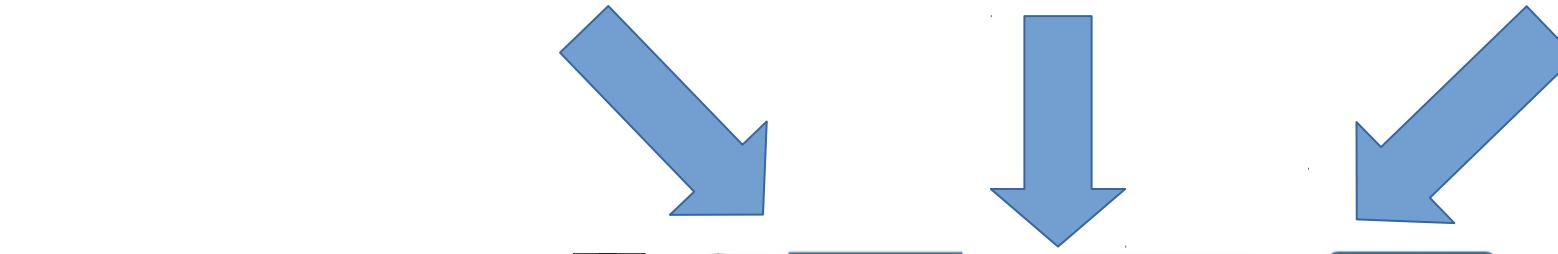
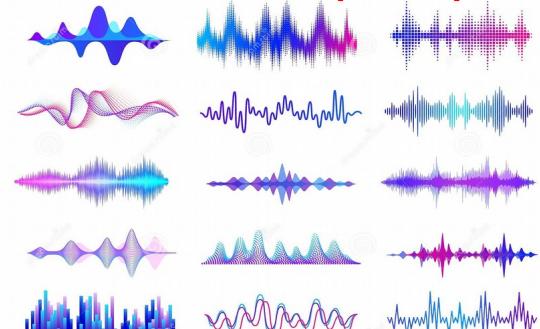


IMAGE (83%)

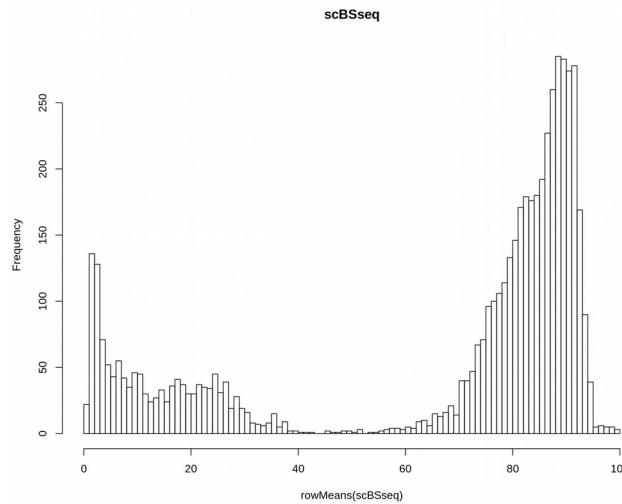
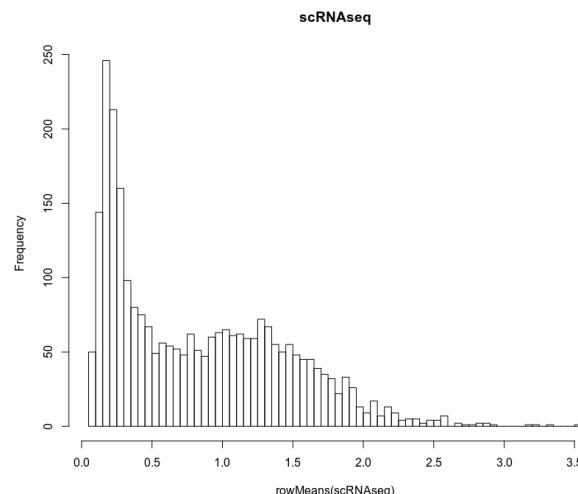
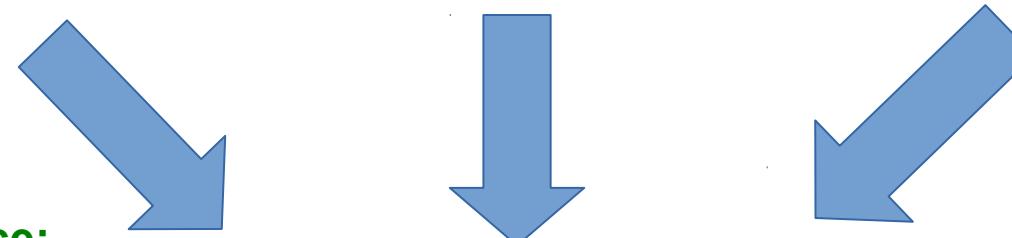
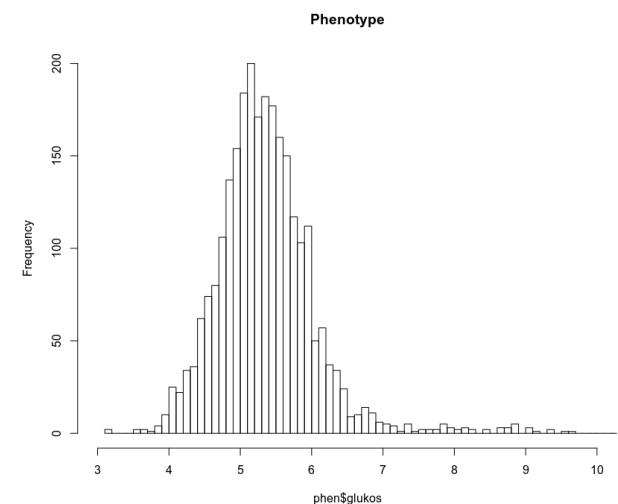


SOUND (75%)



Predict Facebook user interests

Data Integration Accuracy: 96%

Methylation (78%)**Gene Expression (83%)****Phenotype (75%)****1) Convert to common space:**

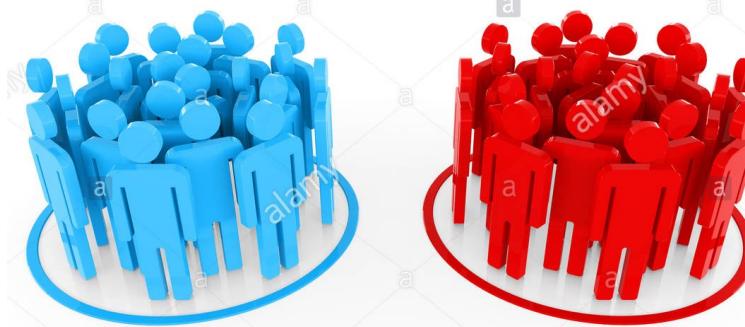
Neural Networks, SNF, UMAP

2) Explicitly model distributions:

MOFA, Bayesian Networks

3) Extract common variation:

PLS, CCA, Factor Analysis

**HEALTHY****SICK****Data Integration
Accuracy: 96%**

Early Works (in scOMICs)

Mapping the Human Body at the Cellular Level

Community generated, multi-omic,
open data processed by standardized pipelines

 4.5M
CELLS 33
ORGANS 289
DONORS 28
PROJECTS 81
LABS[FIND PROJECTS](#)

Filter projects by attribute e.g. organ, project title.

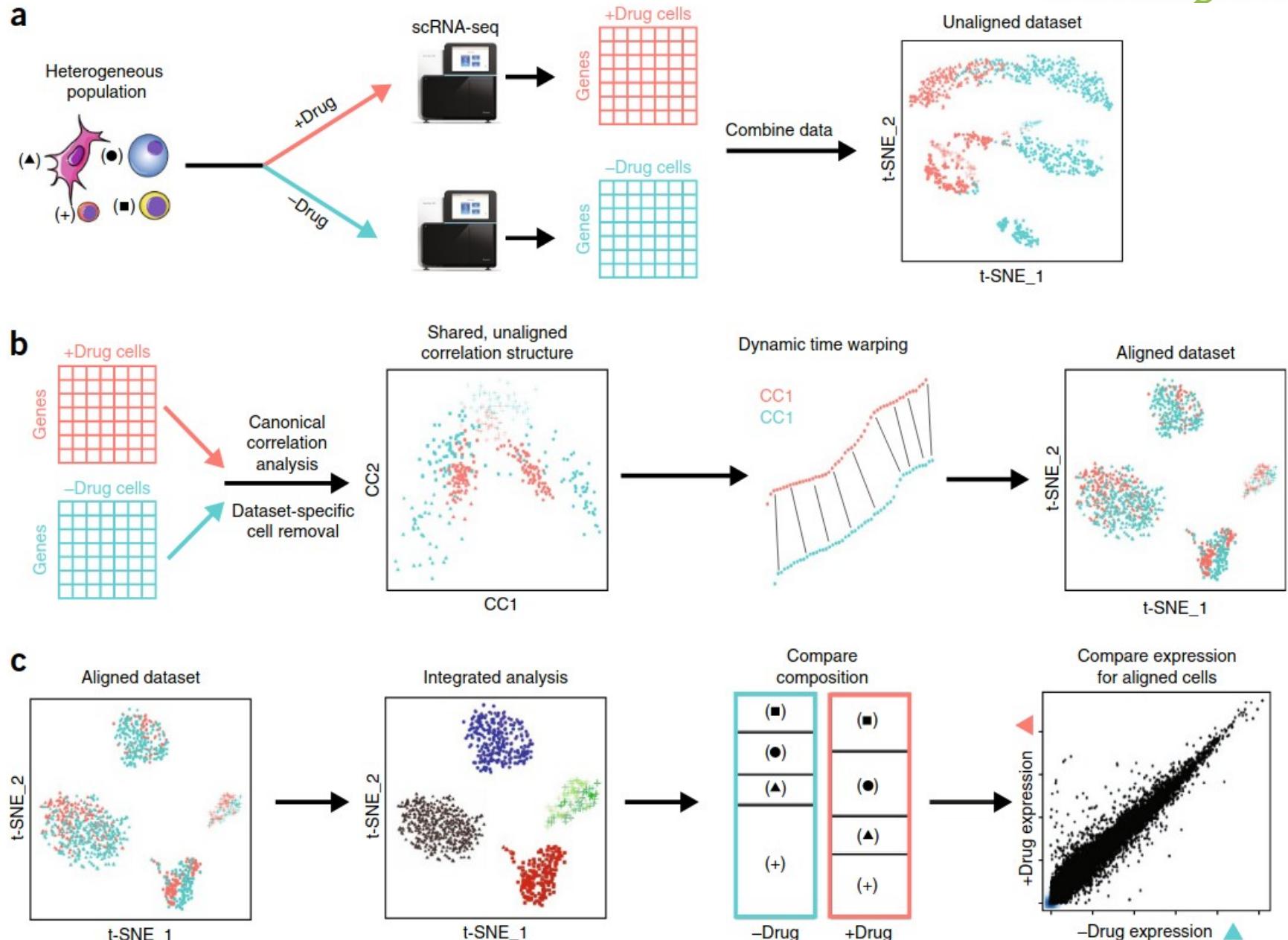
[GO](#)Feedback & Support 

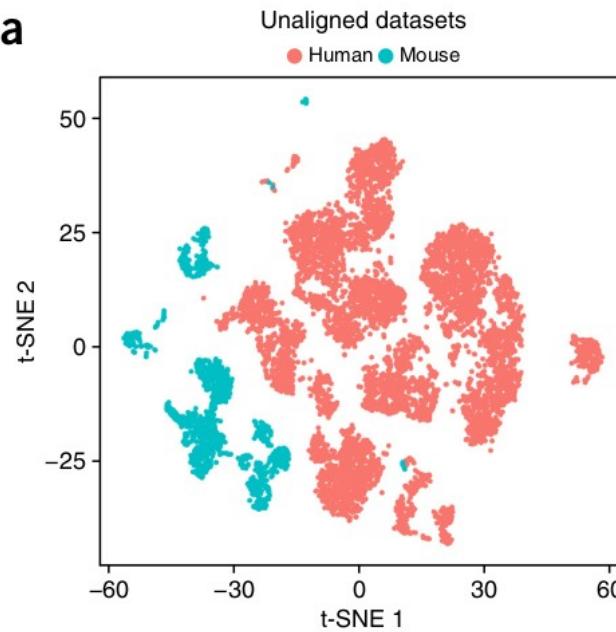
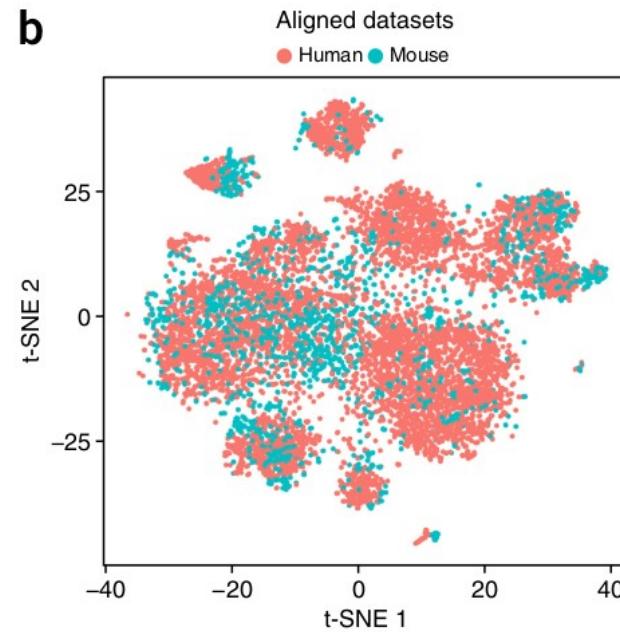
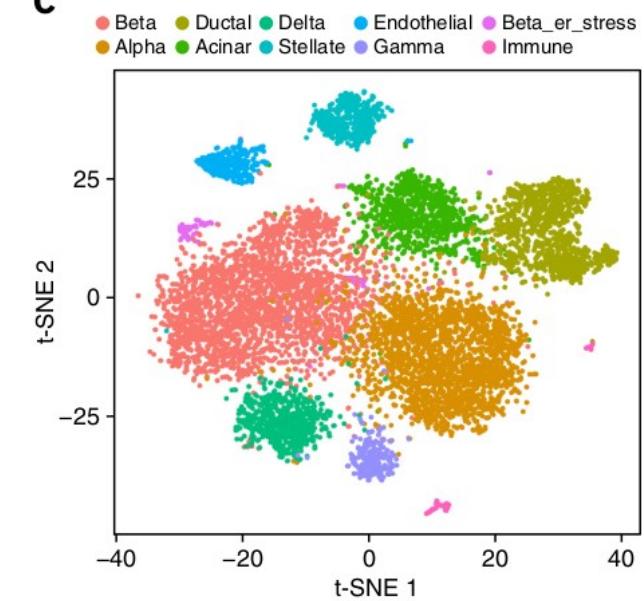
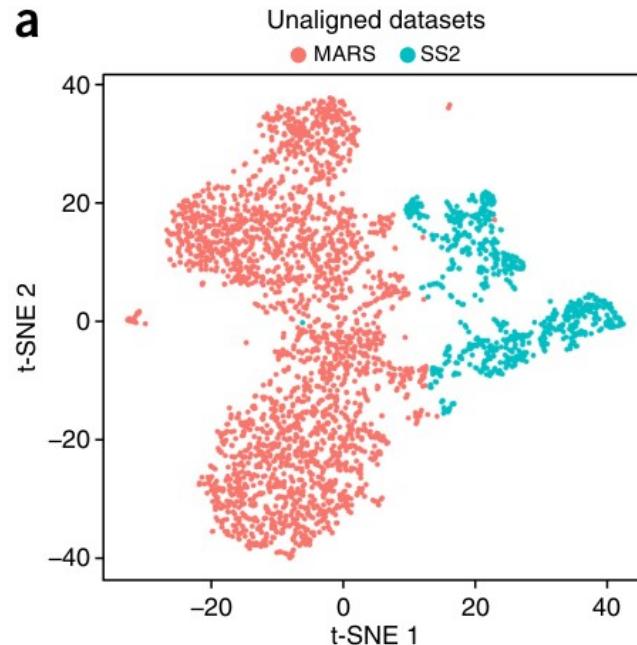
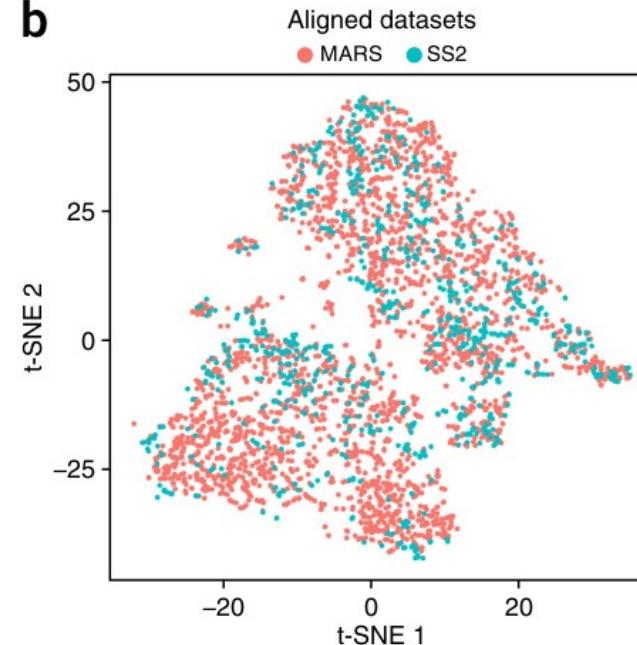
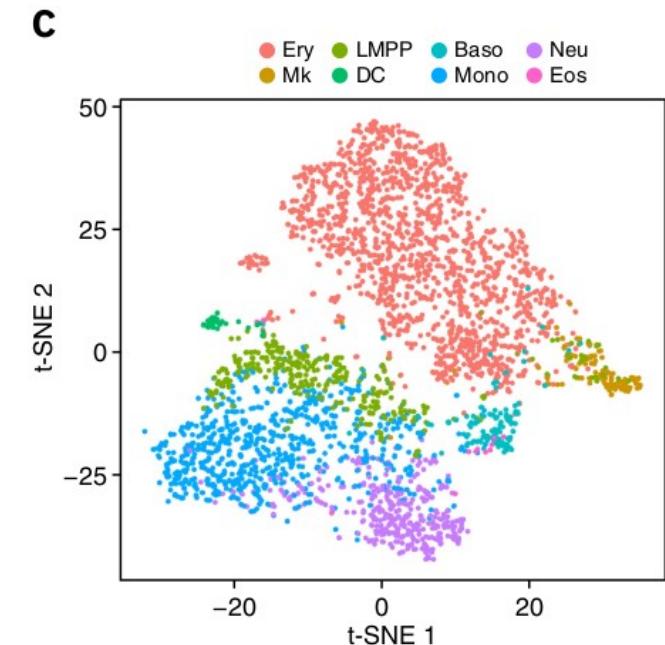
4.5M Cells

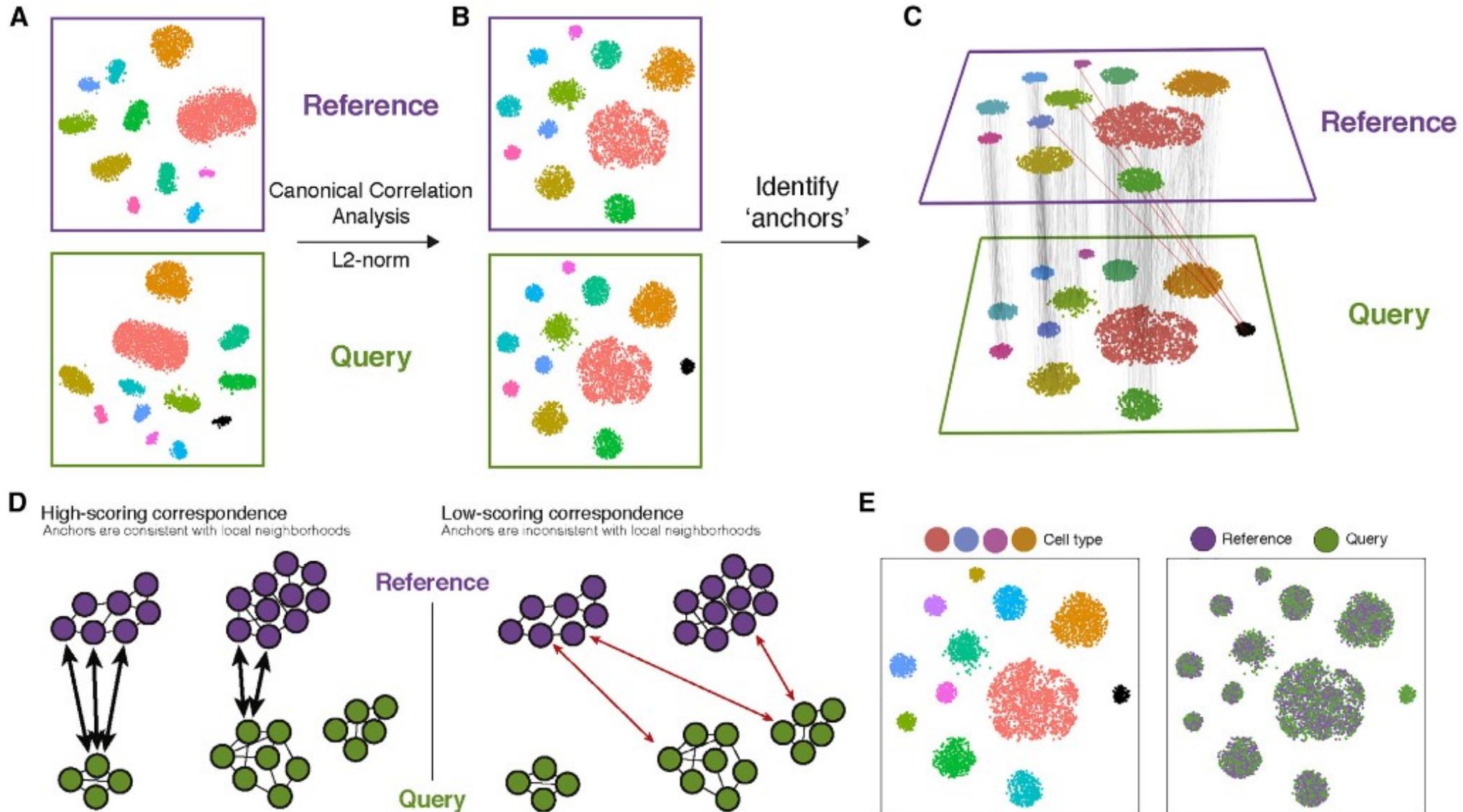
ALL CELLS

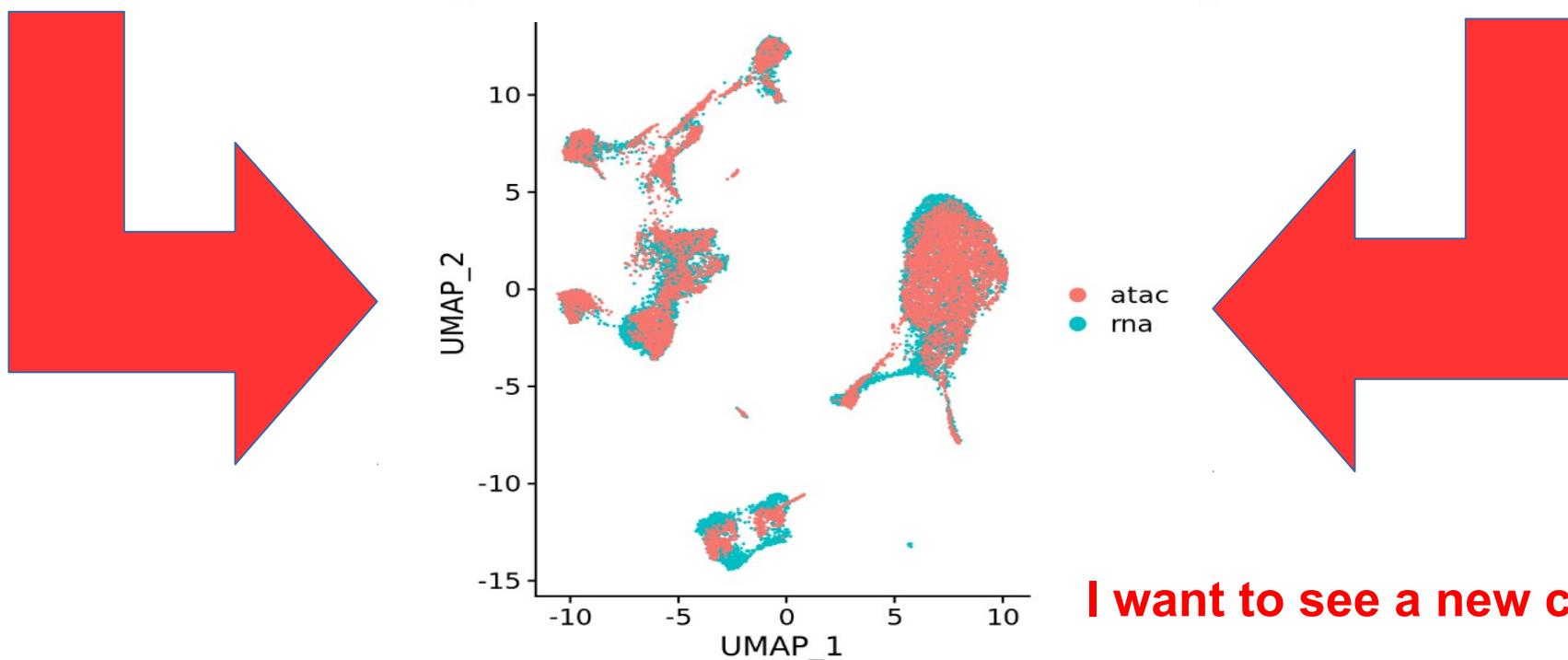
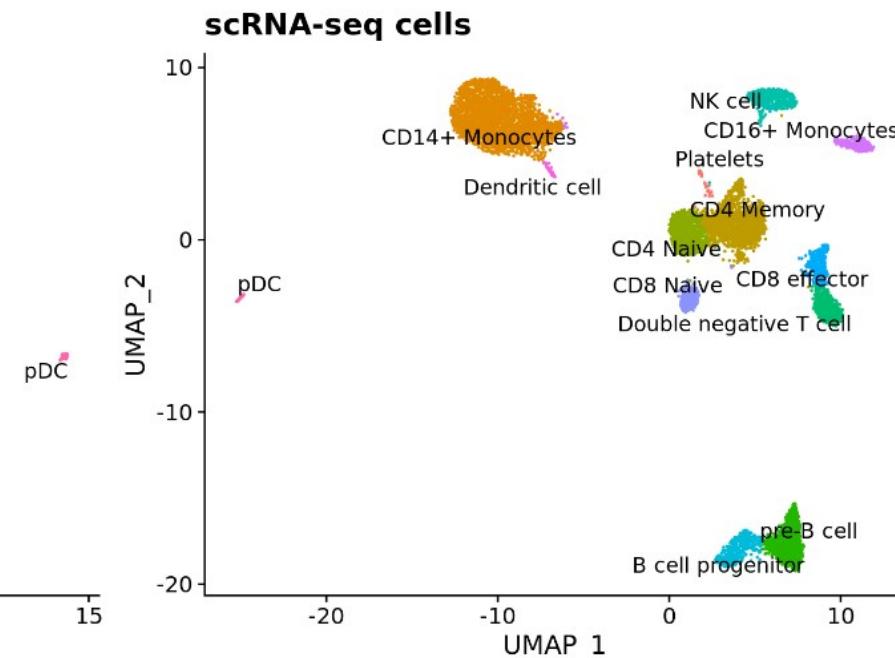
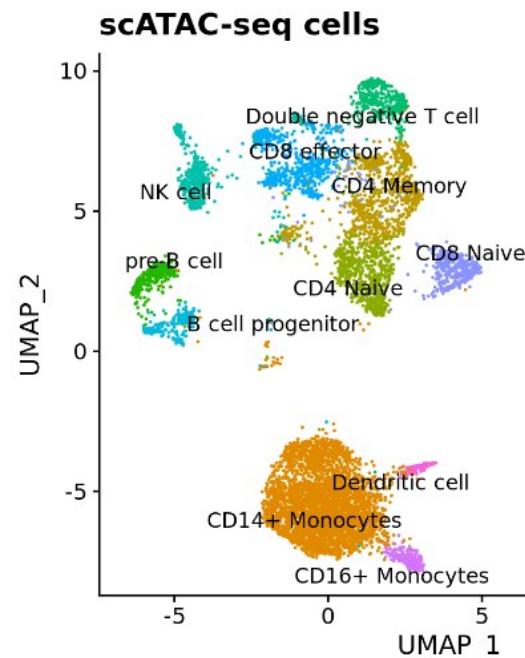
[Blood](#)[Kidney](#)

HCA ambition: create a comprehensive Atlas of human cells from all organs / tissues
Data harmonization / integration is one of major challenges of HCA

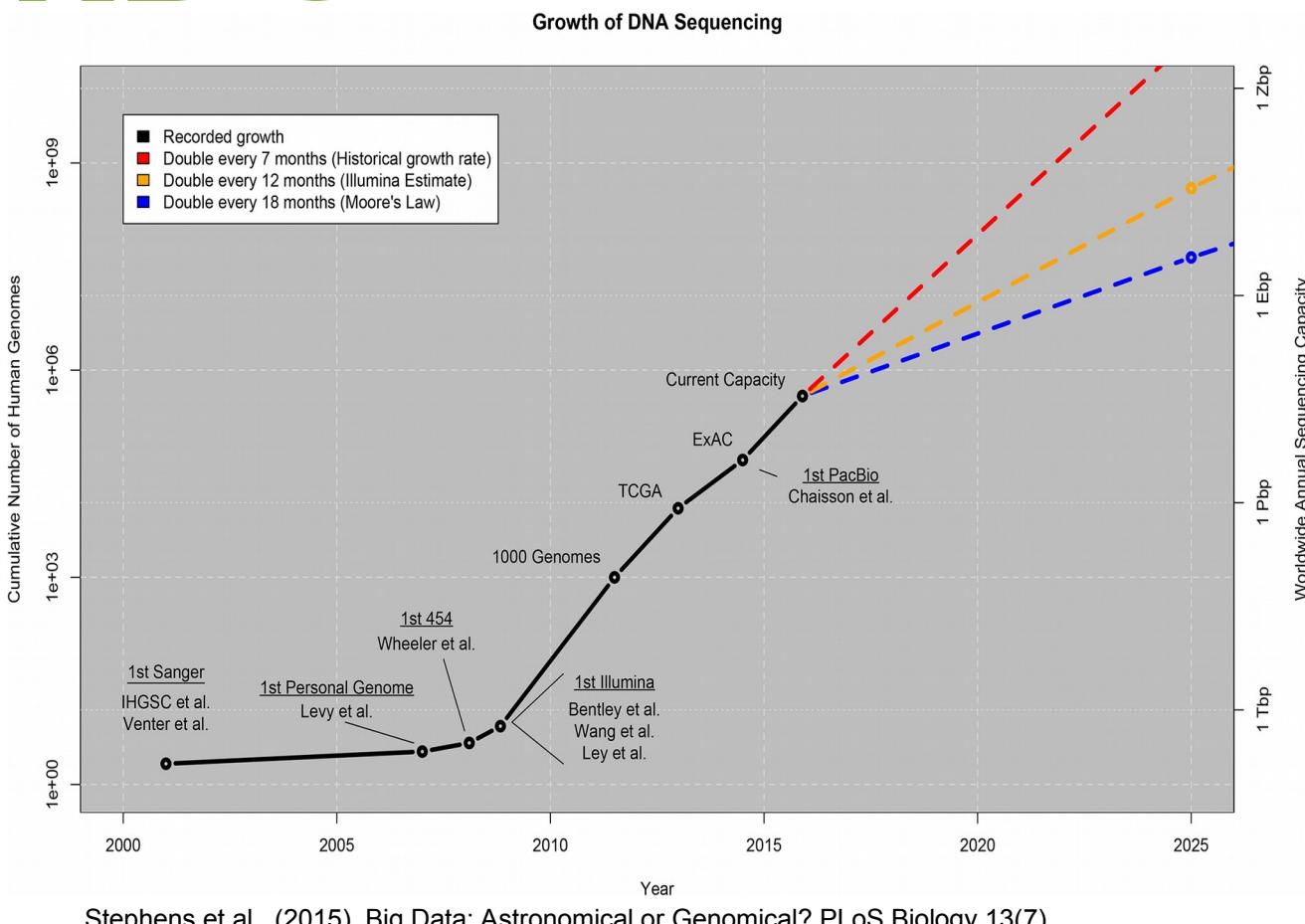


a**b****c****a****b****c**



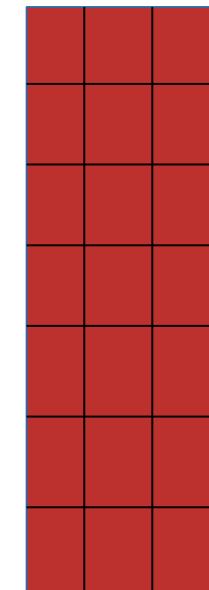


Single Cells Make Big Data



Genomics / WGS: Little Data

$$N_1 \sim 10^3$$

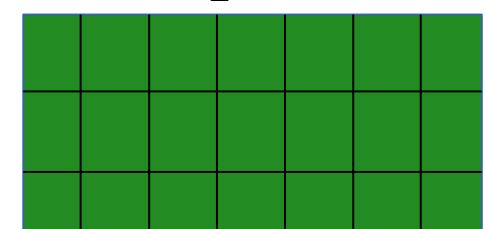


$$P_1 \sim 10^6$$

$$N_1 * P_1 = N_2 * P_2 = 10^9$$

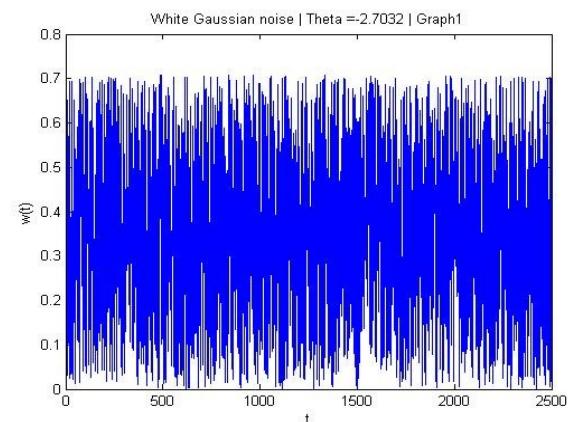
scRNAseq : Big Data

$$N_2 \sim 10^6$$

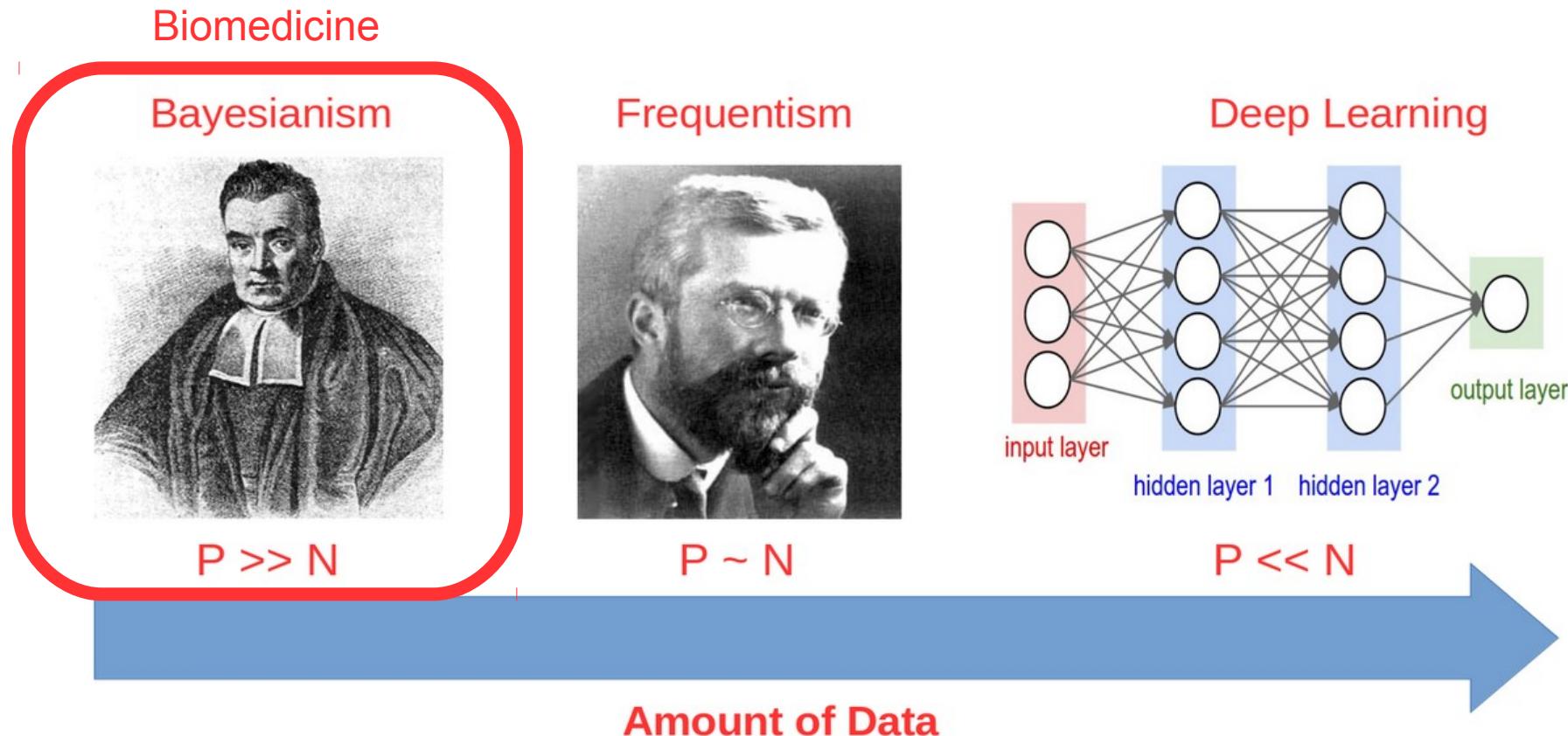


$$P_2 \sim 10^3$$

A file with **White Noise** can also take a lot of disk space



P is the number of features (genes, proteins, genetic variants etc.)
N is the number of observations (samples, cells, nucleotides etc.)

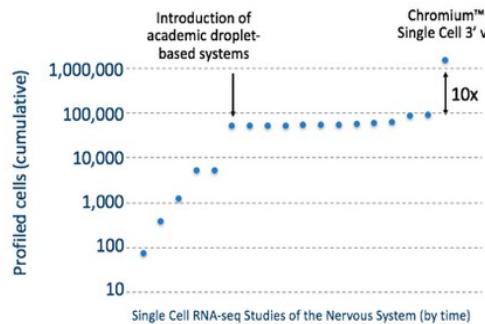


CAREERS BLOG 10X UNIVERSITY

10X GENOMICS SOLUTIONS & PRODUCTS RESEARCH & APPLICATIONS EDUCATION & RESOURCES

< Back to Blog

< Newer Article Older Article >



Our 1.3 million single cell dataset is ready 0 KUDOS



POSTED BY: grace-10x, on Feb 21, 2017 at 2:28 PM

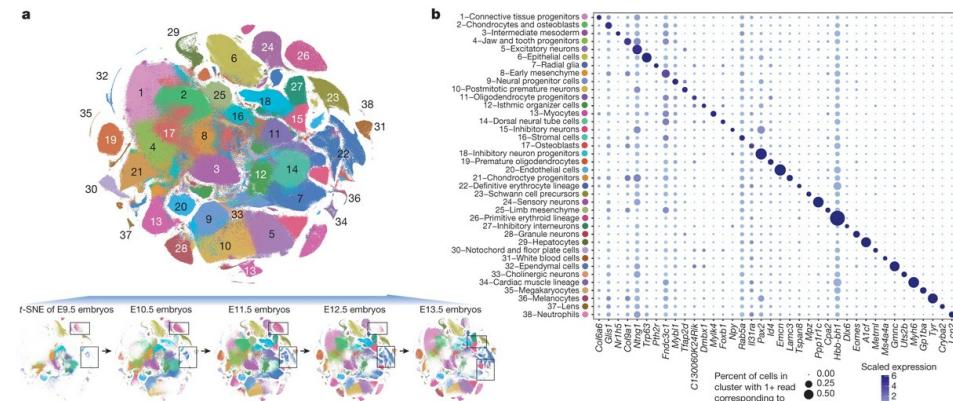
At ASHG last year, we announced our 1.3 Million Brain Cell Dataset, which is, to date, the largest dataset published in the single cell RNA-sequencing (scRNA-seq) field. Using the Chromium™ Single Cell 3' Solution (v2 Chemistry), we were able to sequence and profile 1,308,421 individual cells from embryonic mice brains. Read more in our application note [Transcriptional Profiling of 1.3 Million Brain Cells with the Chromium™ Single Cell 3' Solution](#).

**Watch out Underfitting!
Paradise for Deep Learning!**

MENU nature

Fig. 2: Identifying the major cell types of mouse organogenesis.

From: [The single-cell transcriptional landscape of mammalian organogenesis](#)



a, t-SNE visualization of 2,026,641 mouse embryo cells (after removing a putative doublet cluster), coloured by cluster identity (ID) from Louvain clustering (in **b**), and annotated on the basis of marker genes. The same t-SNE is plotted below, showing only cells from each stage (cell numbers from left to right: n = 151,000 for E9.5; 370,279 for E10.5; 602,784 for E11.5; 468,088 for E12.5; 434,490 for E13.5). Primitive erythroid (transient) and definitive erythroid (expanding) clusters are boxed. **b**, Dot plot showing expression of one selected marker gene per cell type. The size of the dot encodes the percentage of cells within a cell type in

BioTuring™ Solutions Resources

Explore **4,000,000 CELLS** at ease with **Bioturing Browser**
A next-generation platform to re-analyze published single-cell sequencing data
[EXPLORER NOW](#)

Single Cell Analysis

5,500,000 cells will be indexed into BioTuring Single-cell Data Repository this September

by biomembers • August 30, 2019

Human Cell Atlas, single-cell data

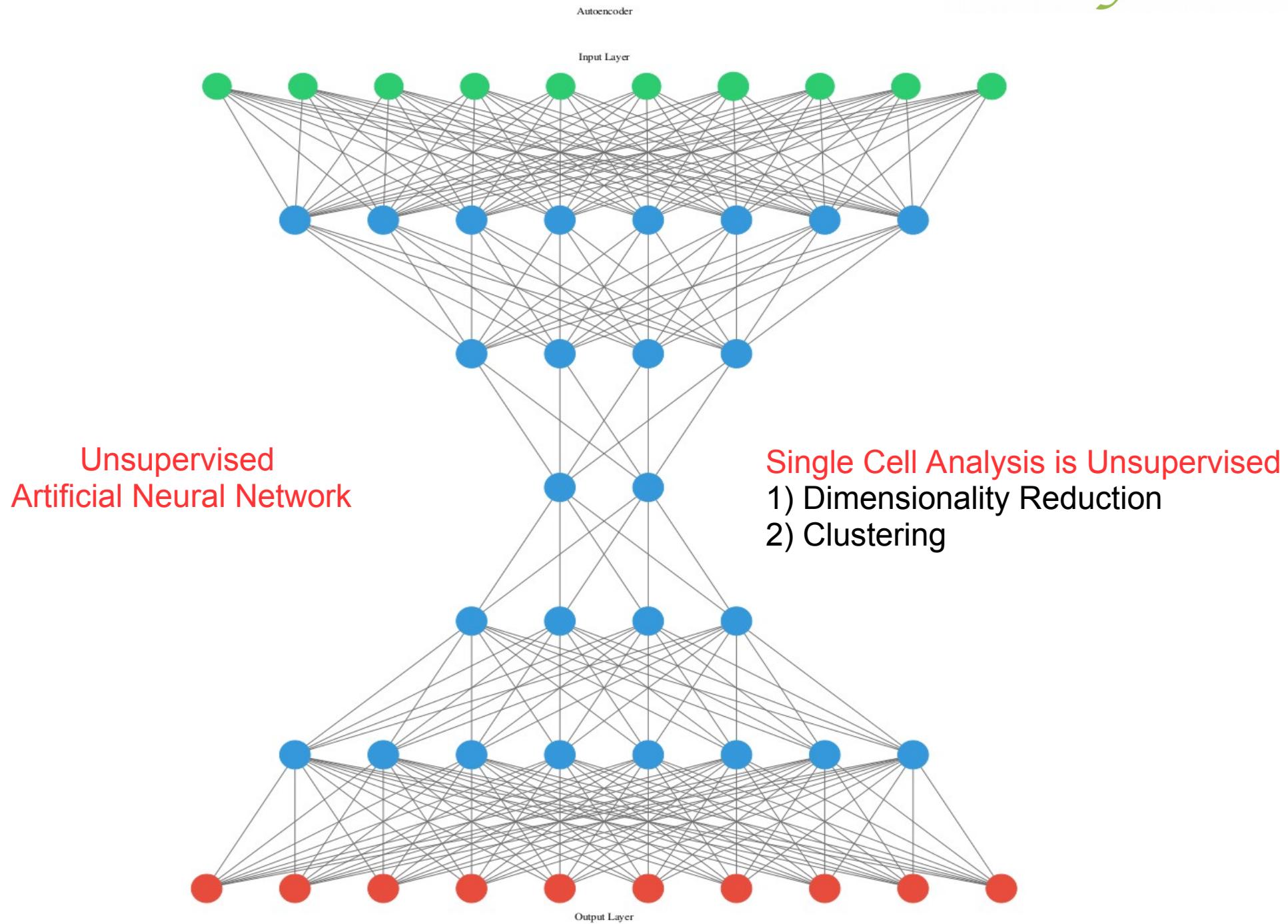
We are glad to announce that we will upsize the current single-cell database in **BioTuring Single-cell Browser** to 5,500,000 cells this September. With this release, we will double the current number of publications indexed in BioTuring Single-cell Browser, and cross the number of cells hosted on available public single-cell data repositories like [Human Cell Atlas \(HCA\)](#) and [Broad Institute's Single-cell Portal](#).

Search

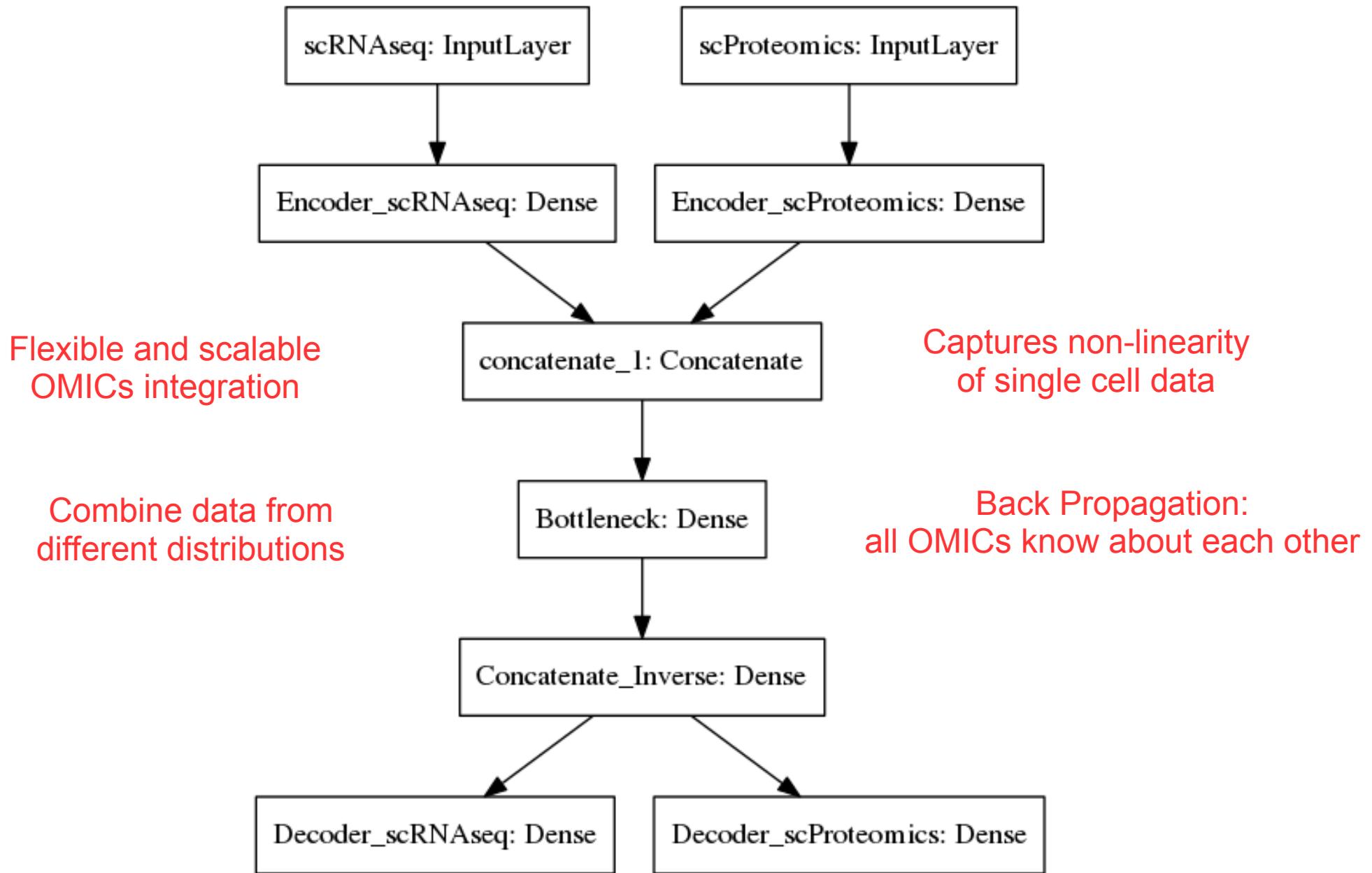
RECENT POSTS

A new tool to interactively visualize single-cell objects (Seurat, Scanpy, SingleCellExperiments, ...)
September 26, 2019

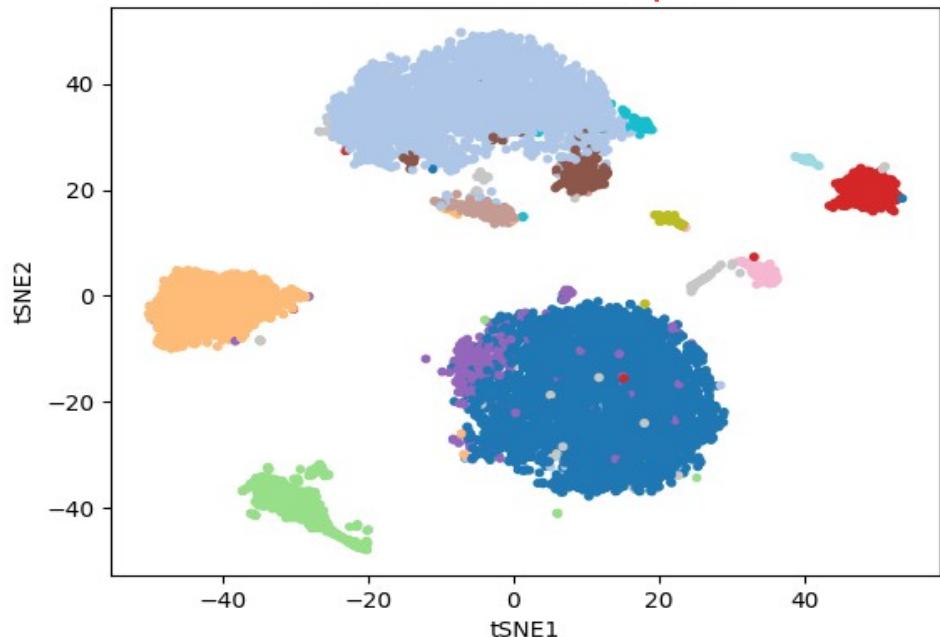
5,500,000 cells will be indexed into BioTuring Single-cell Data Repository this September
August 30, 2019



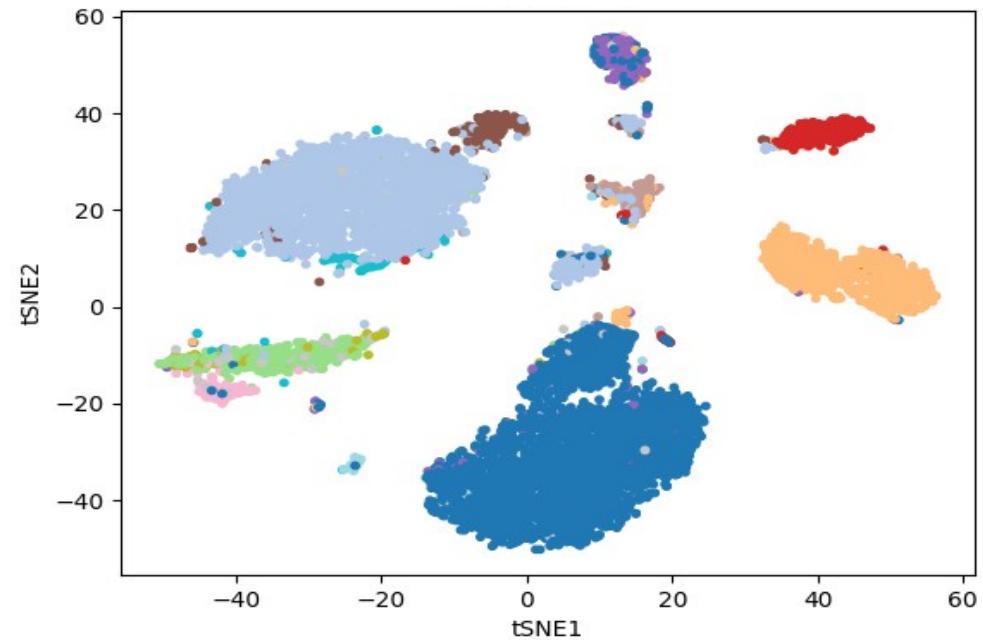
CITE-seq: Data Integration Autoencoder



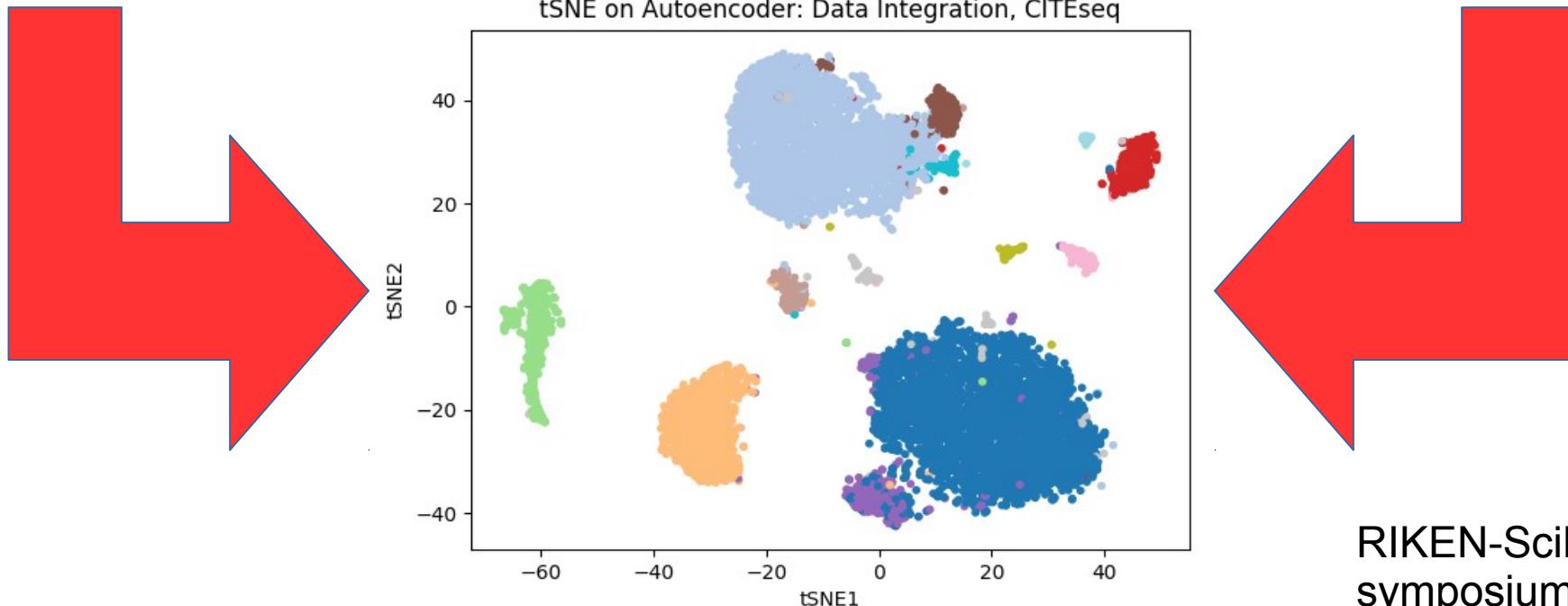
scRNAseq



scProteomics



tSNE on Autoencoder: Data Integration, CITEseq



Seurat Confirms CD8+ T Cluster Formation

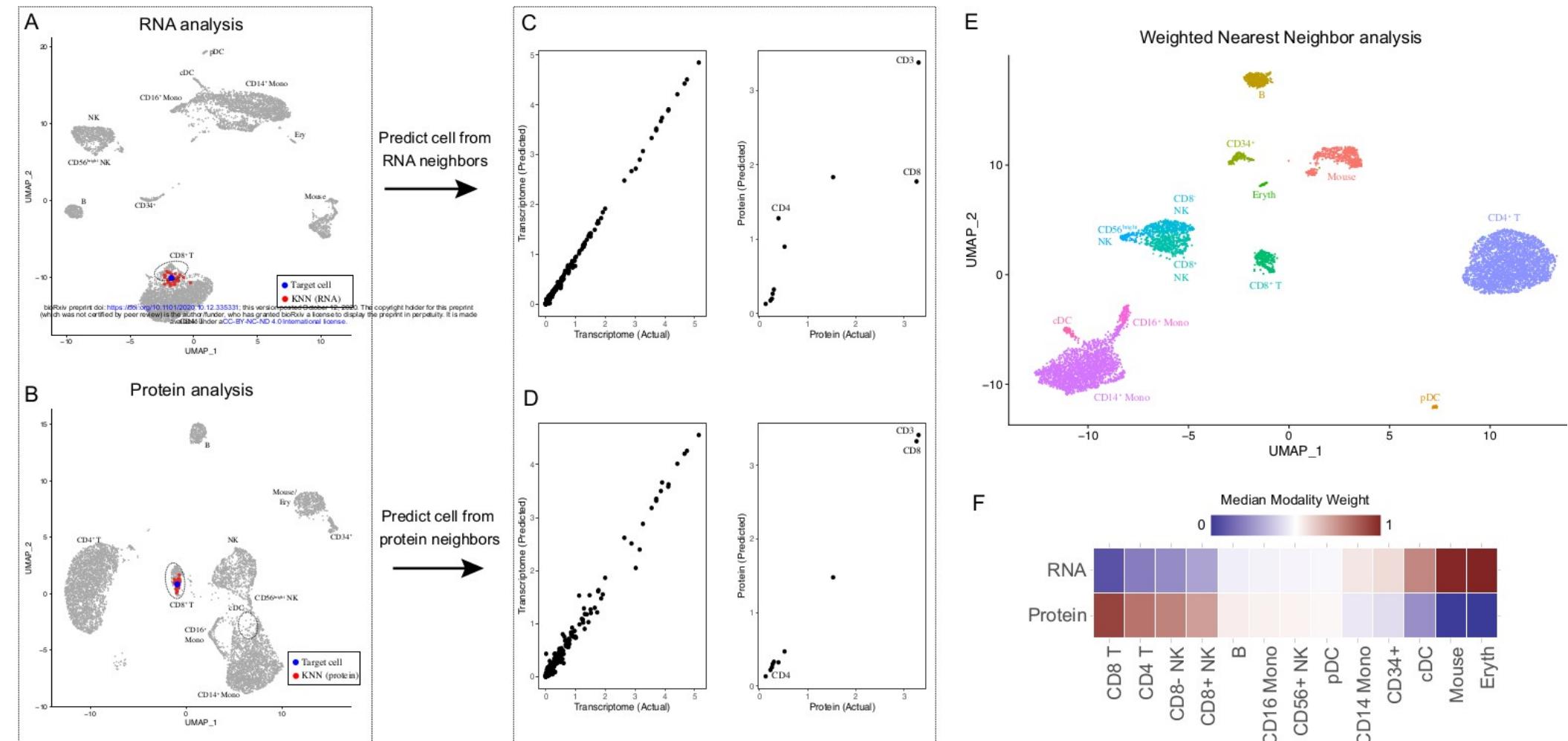
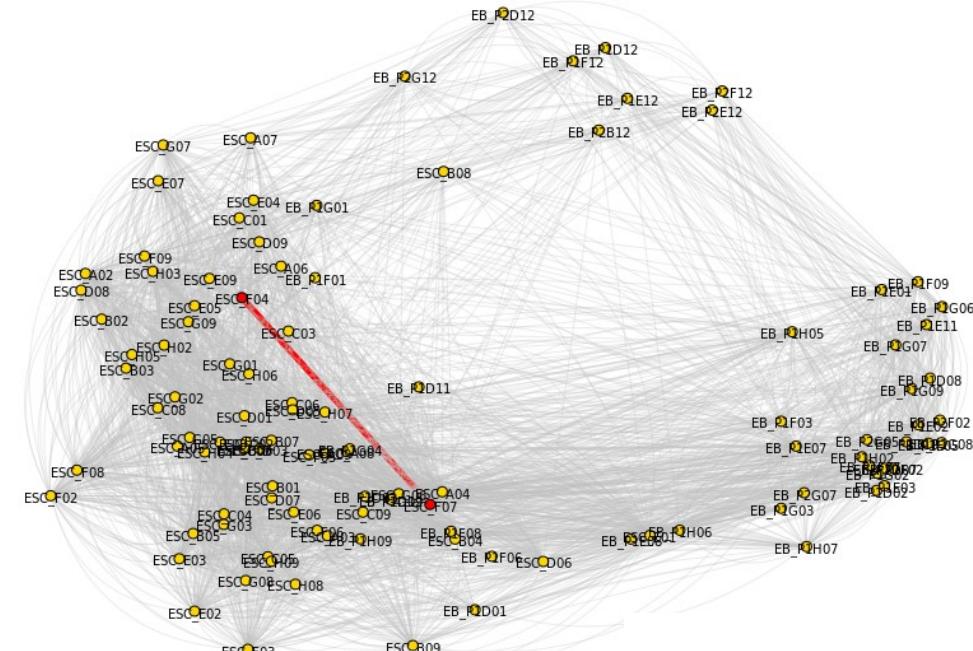


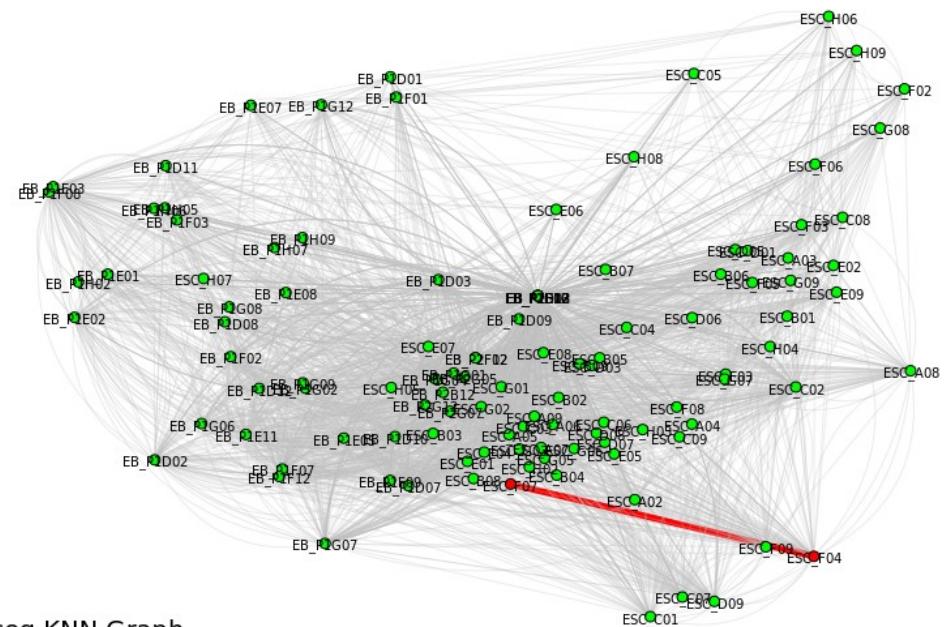
Figure 1: Schematic overview of multimodal integration using Weighted Nearest Neighbor analysis

Graph-Based scOMICs Integration

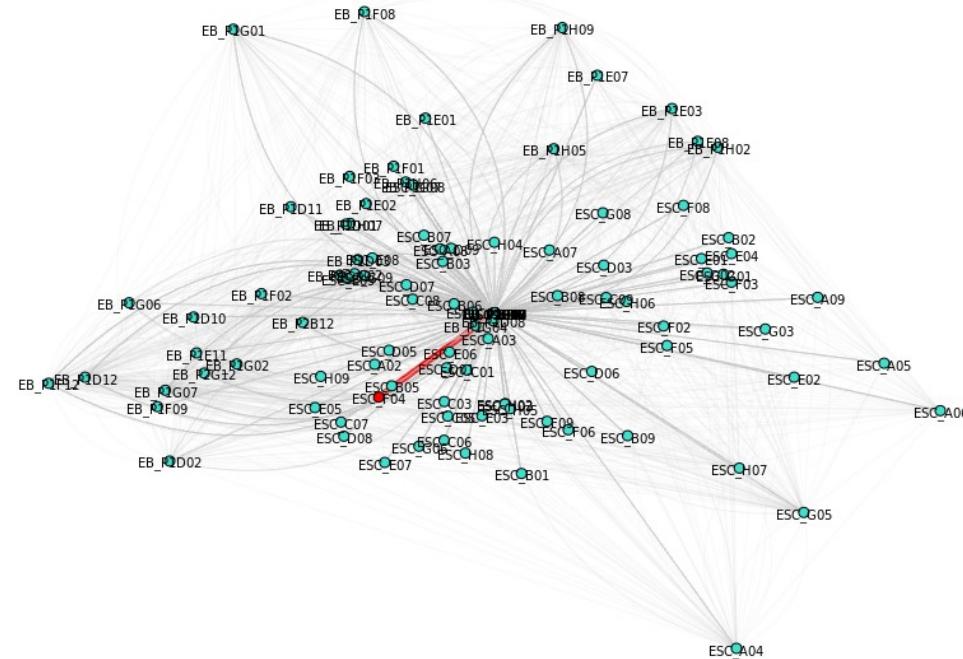
scRNAseq KNN Graph



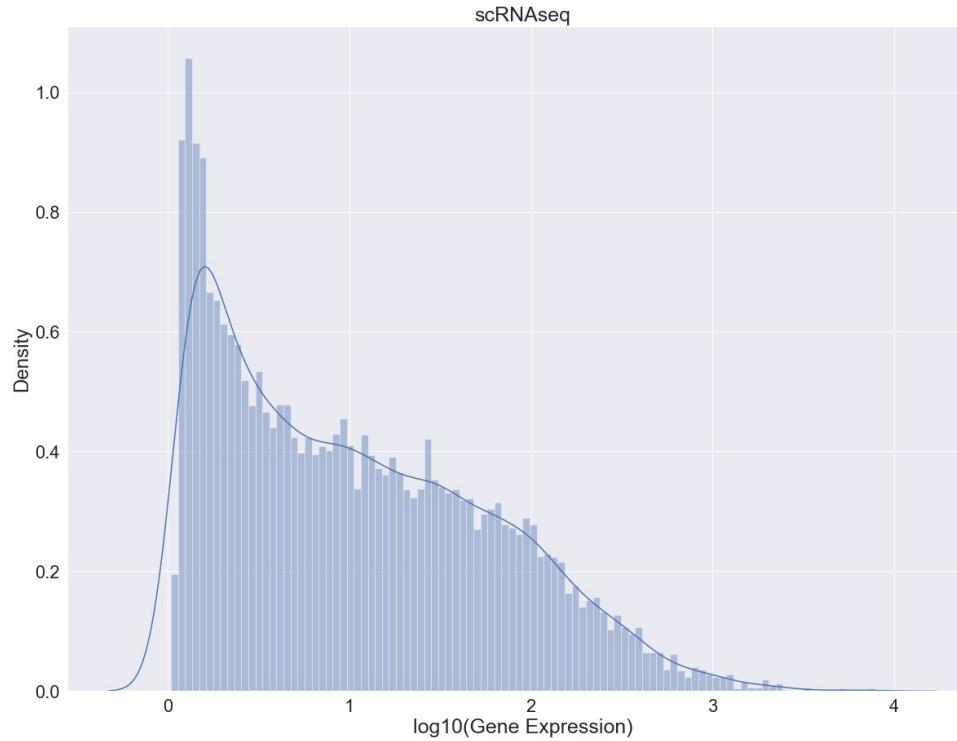
scBSseq KNN Graph



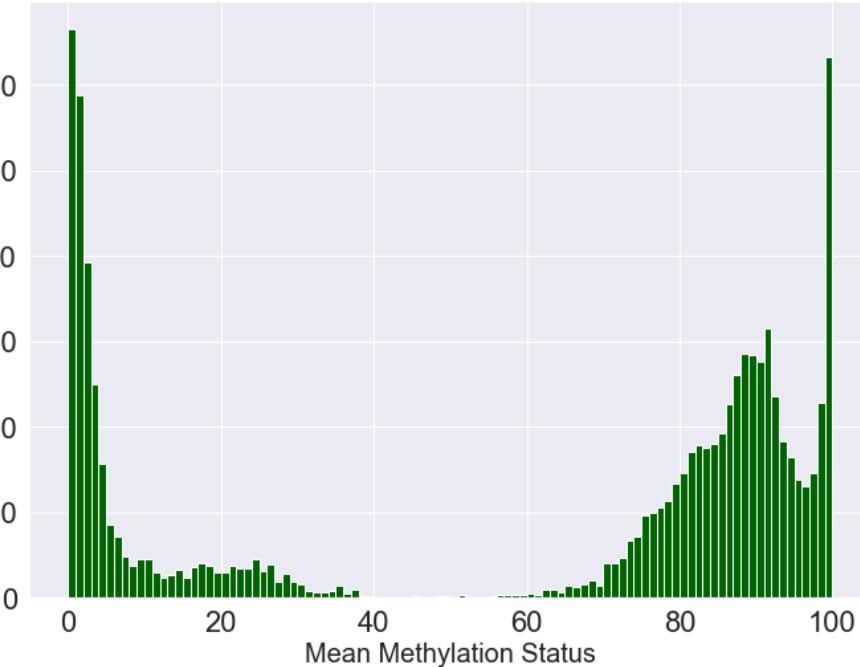
scATACseq KNN Graph



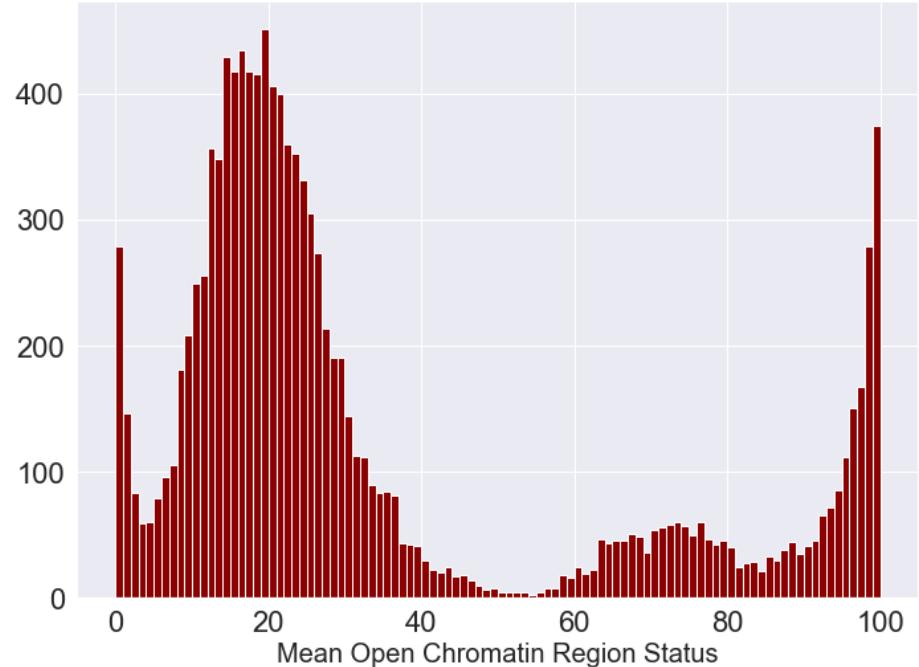
Graph-Based Integration



scBSseq



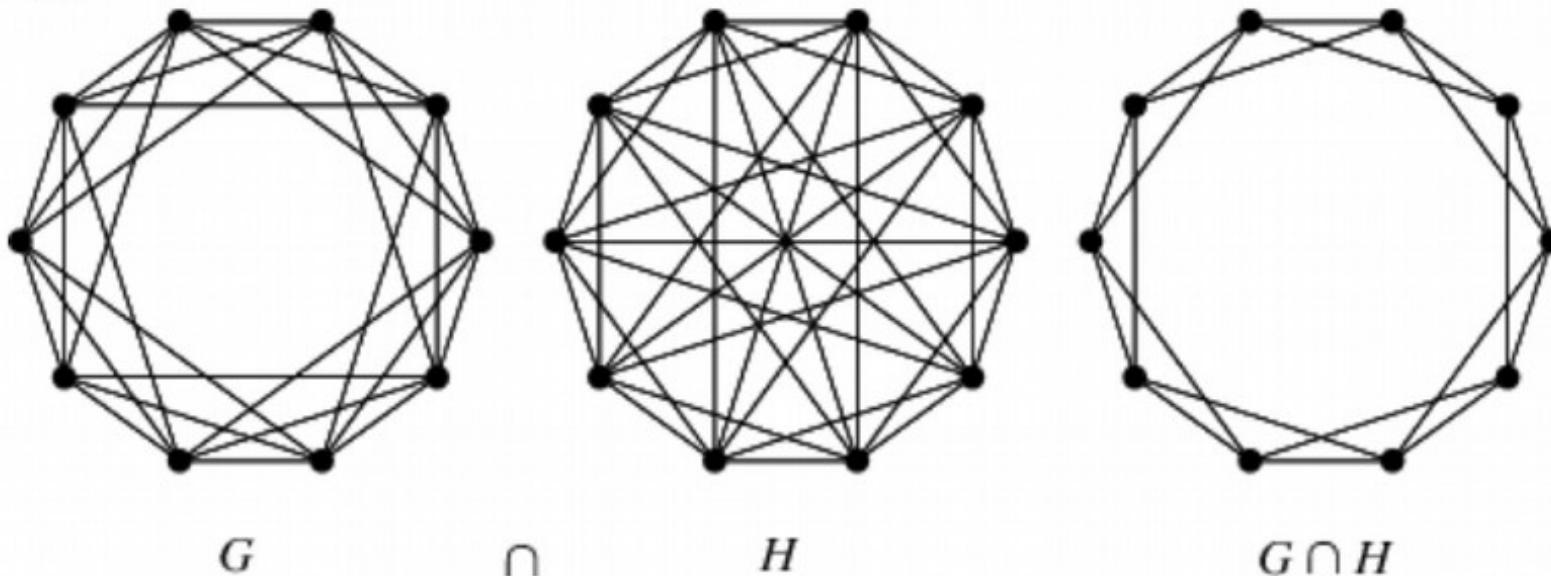
scATACseq



Graph Intersection

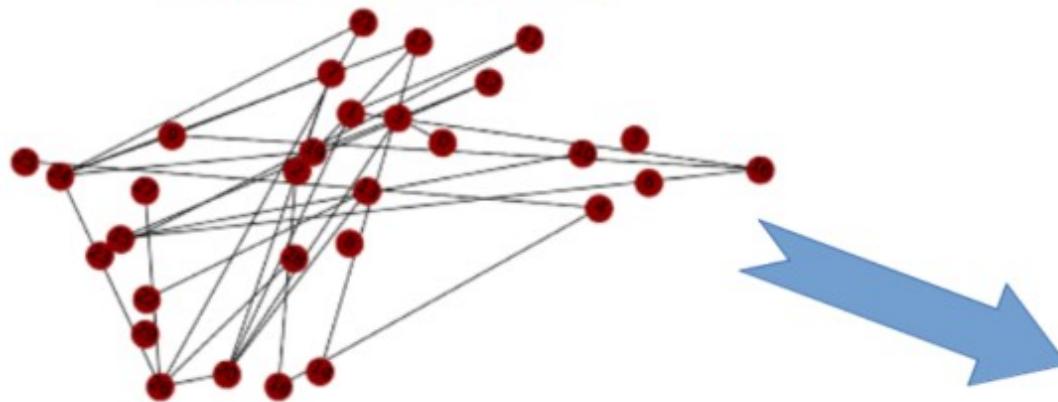


DOWNLOAD
Wolfram Notebook

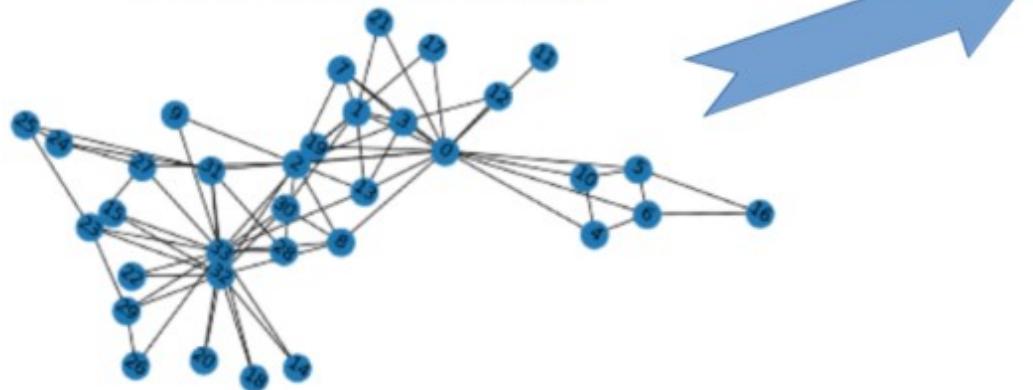


Let S be a set and $F = \{S_1, \dots, S_p\}$ a nonempty family of distinct nonempty subsets of S whose union is $\bigcup_{i=1}^p S_i = S$. The intersection graph of F is denoted $\Omega(F)$ and defined by $V(\Omega(F)) = F$, with S_i and S_j adjacent whenever $i \neq j$ and $S_i \cap S_j \neq \emptyset$. Then a graph G is an intersection graph on S if there exists a family F of subsets for which G and $\Omega(F)$ are isomorphic graphs (Harary 1994, p. 19). Graph intersections can be computed in the Wolfram Language using `GraphIntersection[g, h]`.

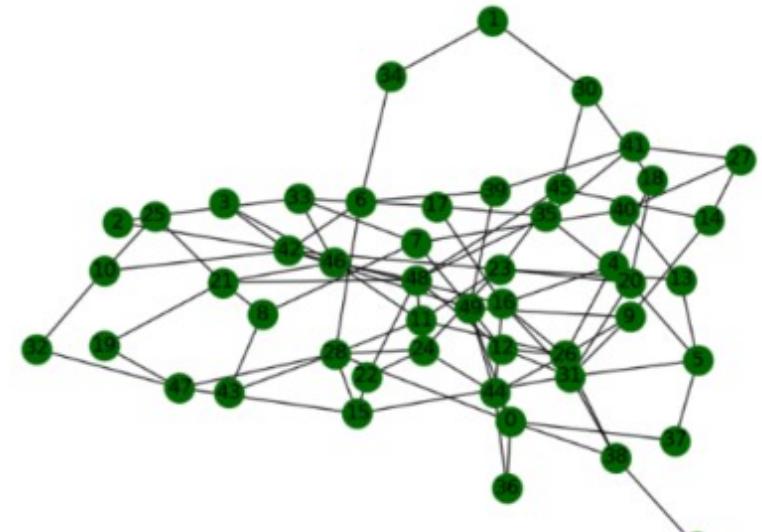
**Binary Data Graph:
Hamming Similarity**



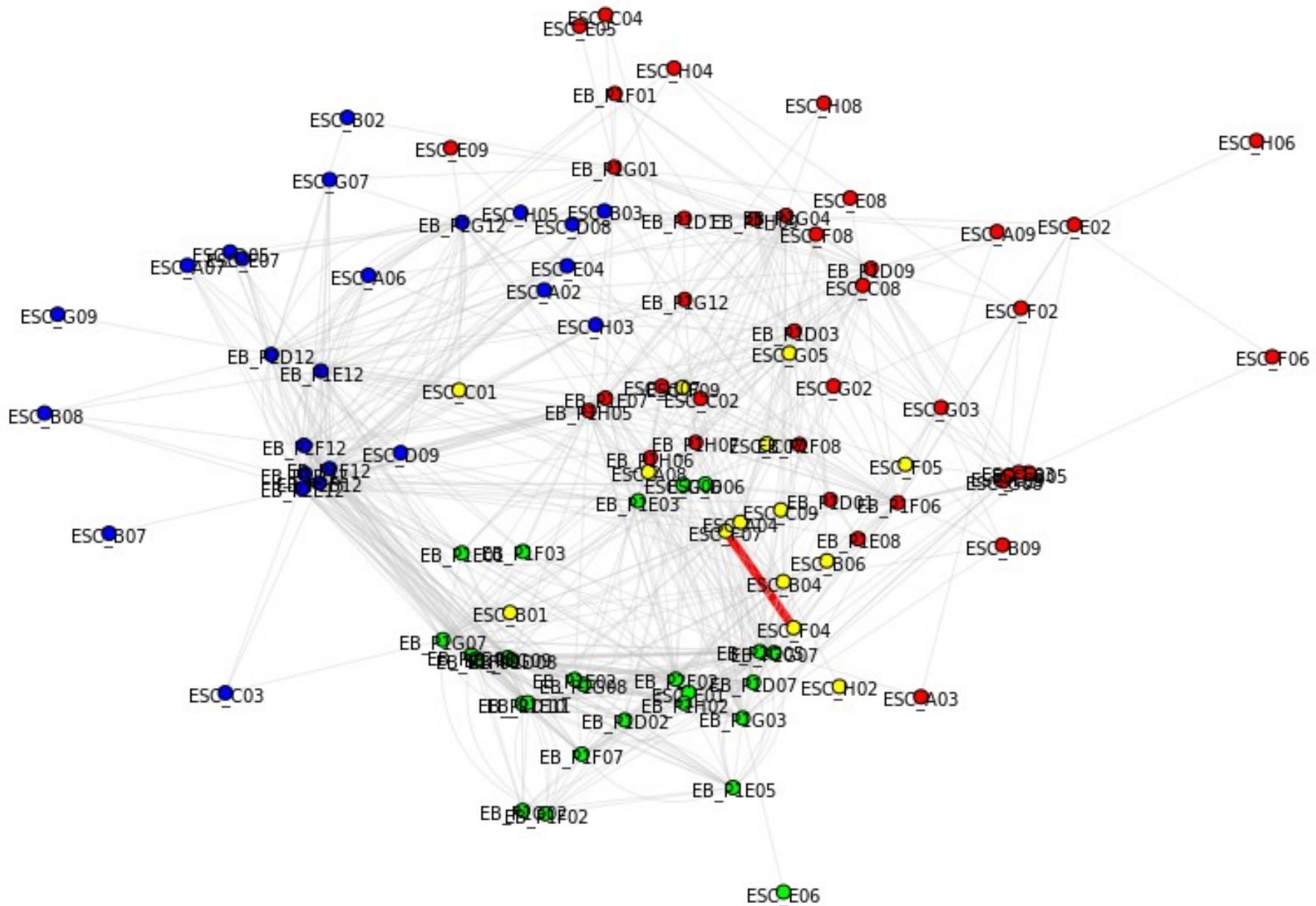
**Continuous Data Graph:
Euclidean Similarity**

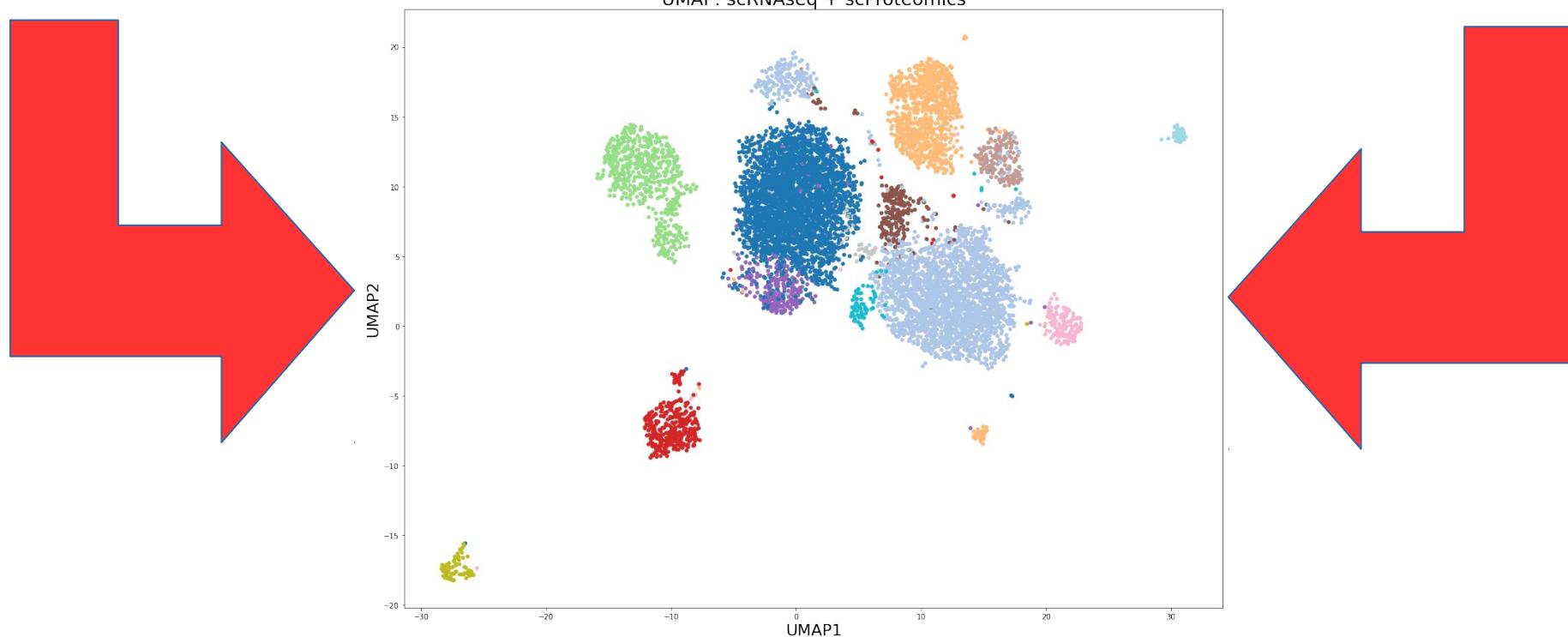
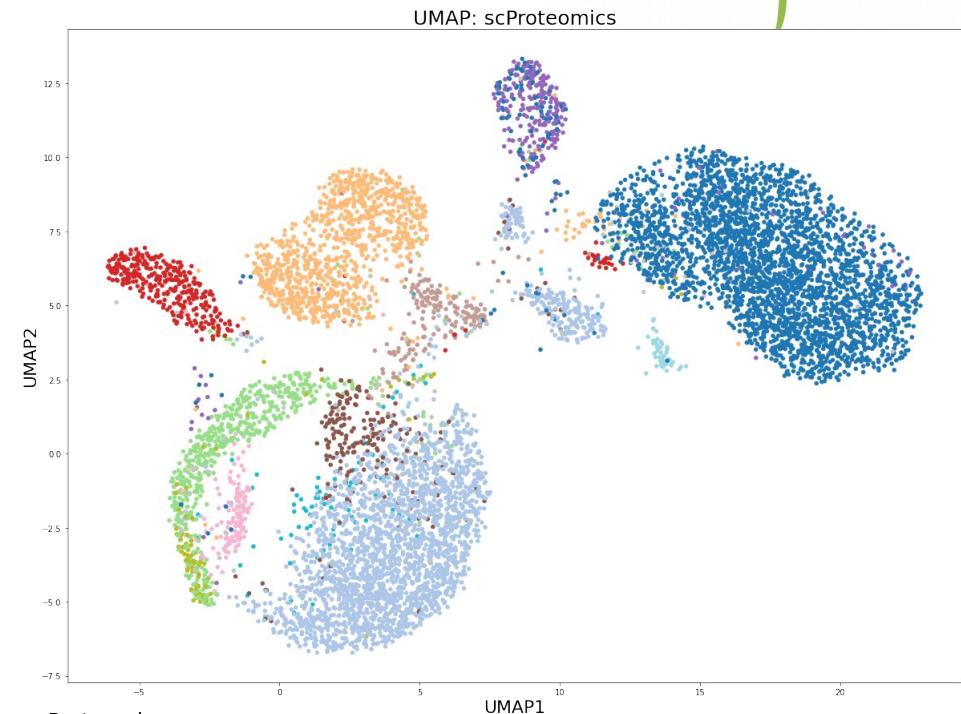
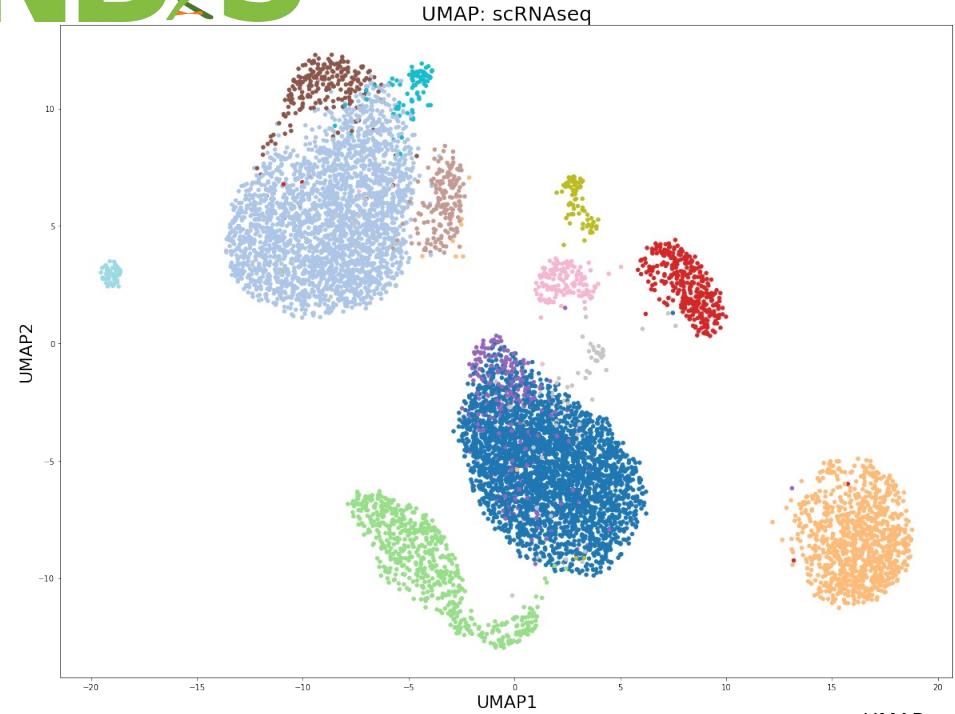


**Consensus Graph:
Keep Consistent Edges**



Consensus Graph





Recent Works

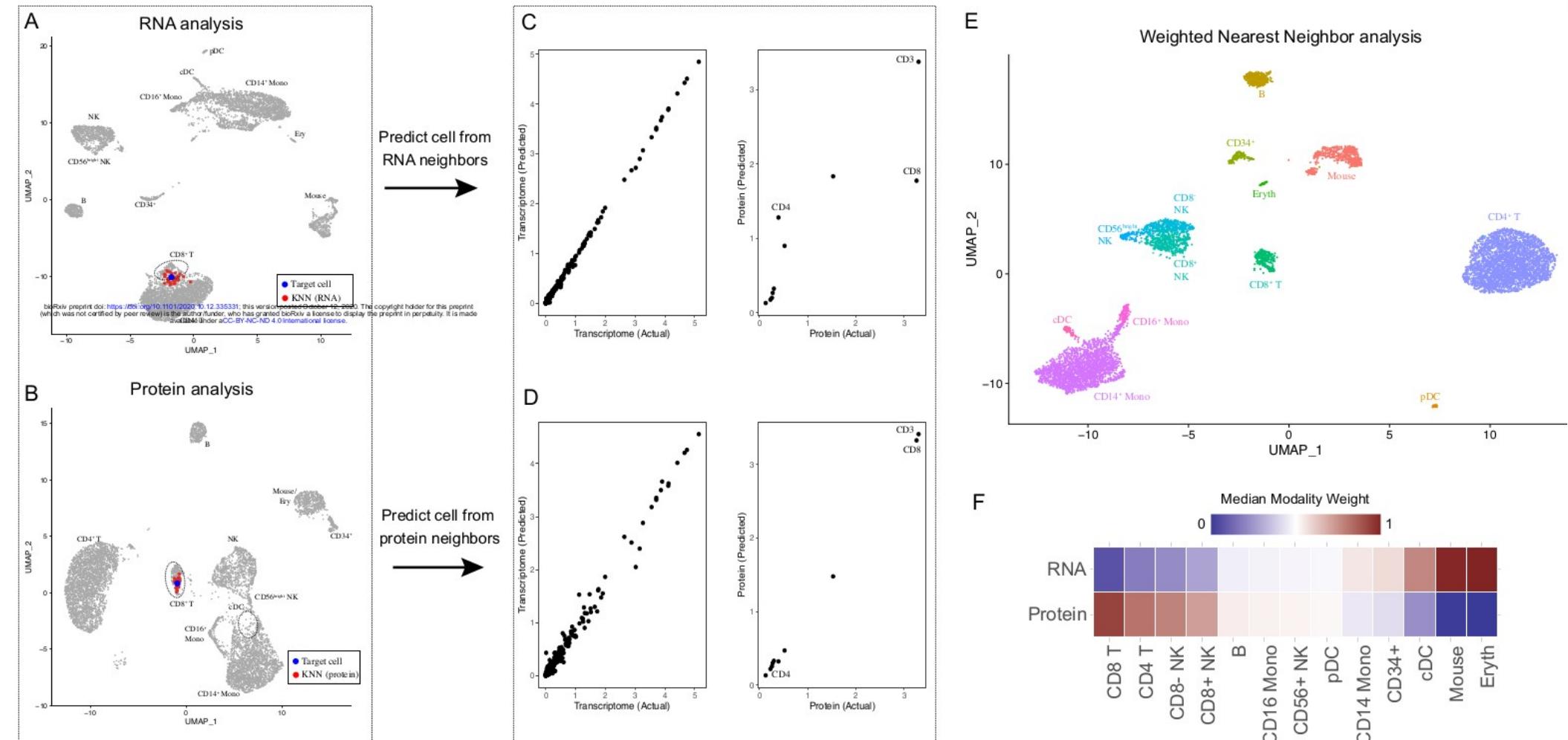
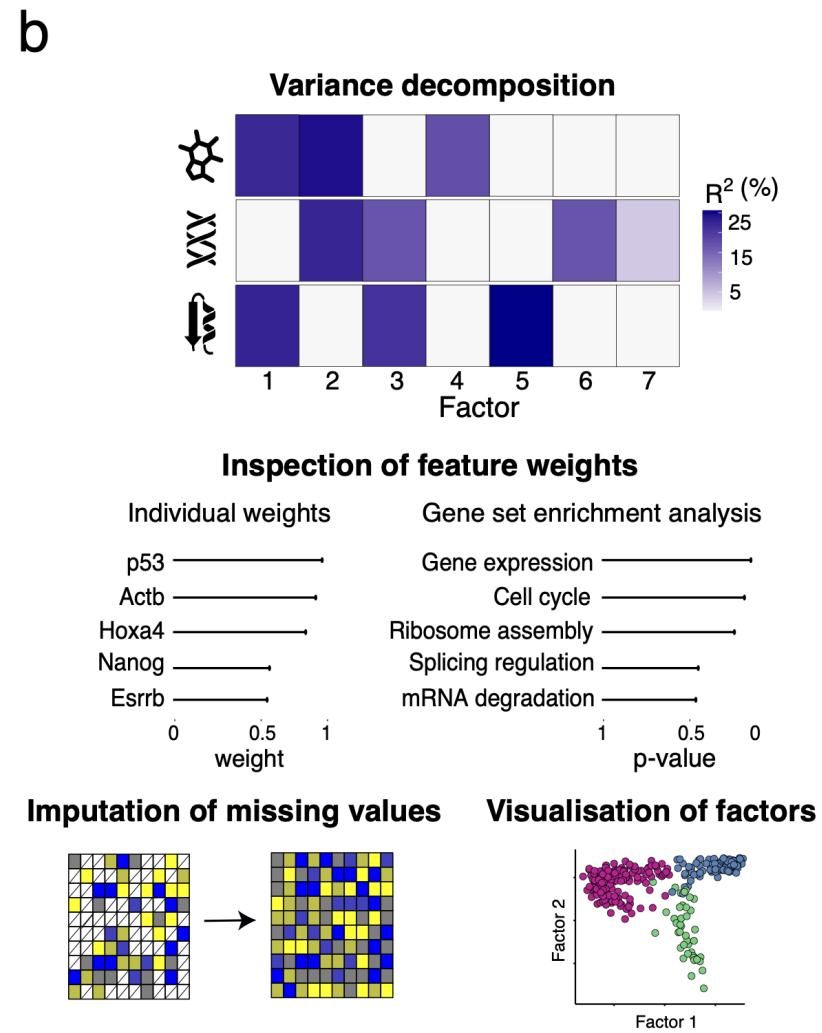
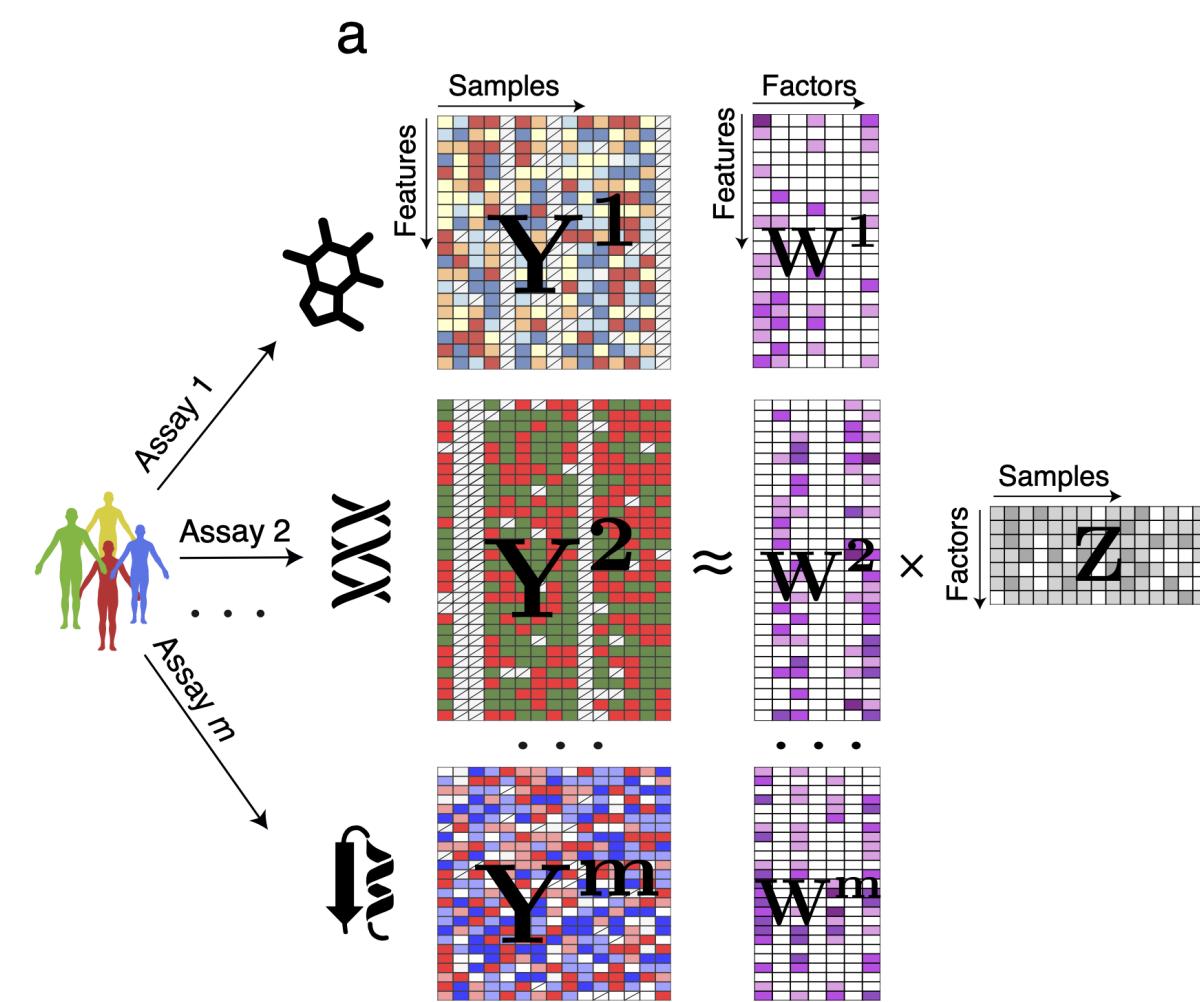


Figure 1: Schematic overview of multimodal integration using Weighted Nearest Neighbor analysis

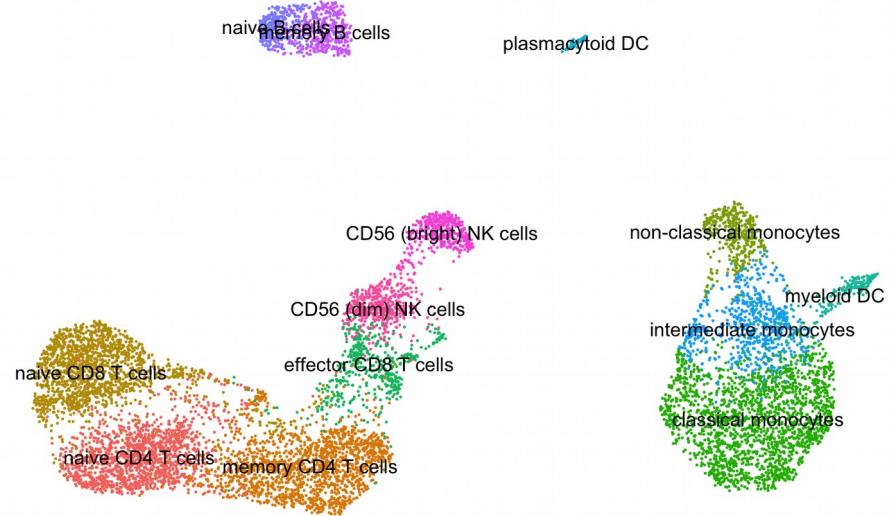
Hao et al., bioRxiv 2020.10.12.335331; doi: <https://doi.org/10.1101/2020.10.12.335331>

For each cell, we began by independently calculating sets of $k=20$ nearest neighbors for each modality

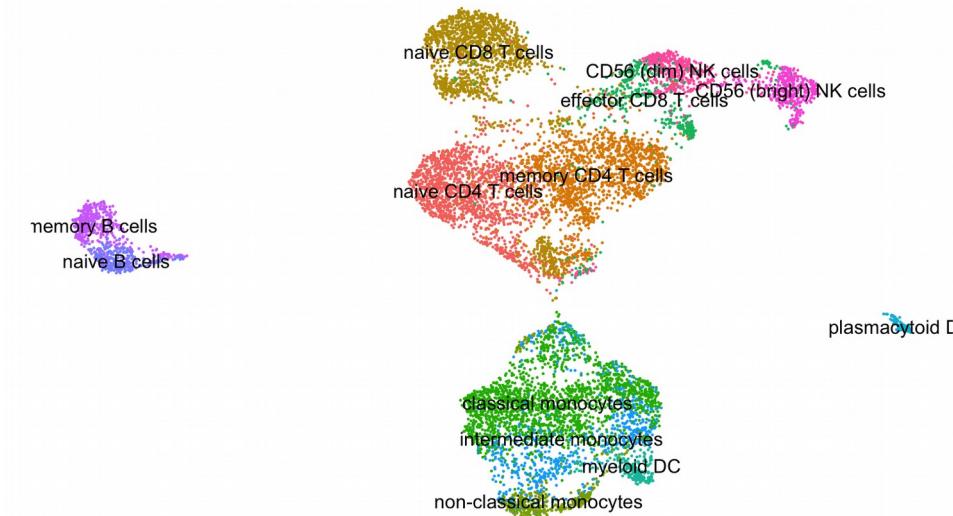
For each cell, we calculate its closest neighbors in the dataset based on a weighted averaging of RNA and protein similarities.



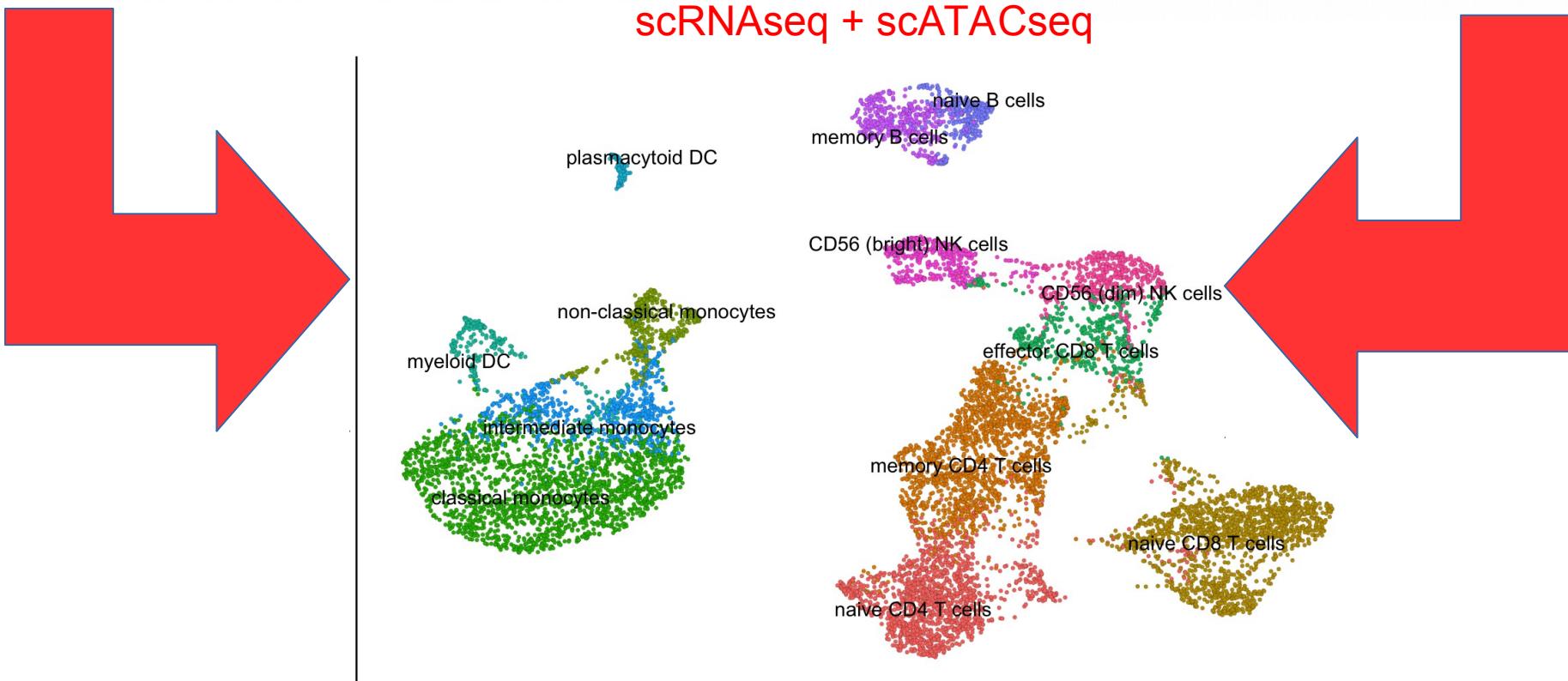
scRNAseq

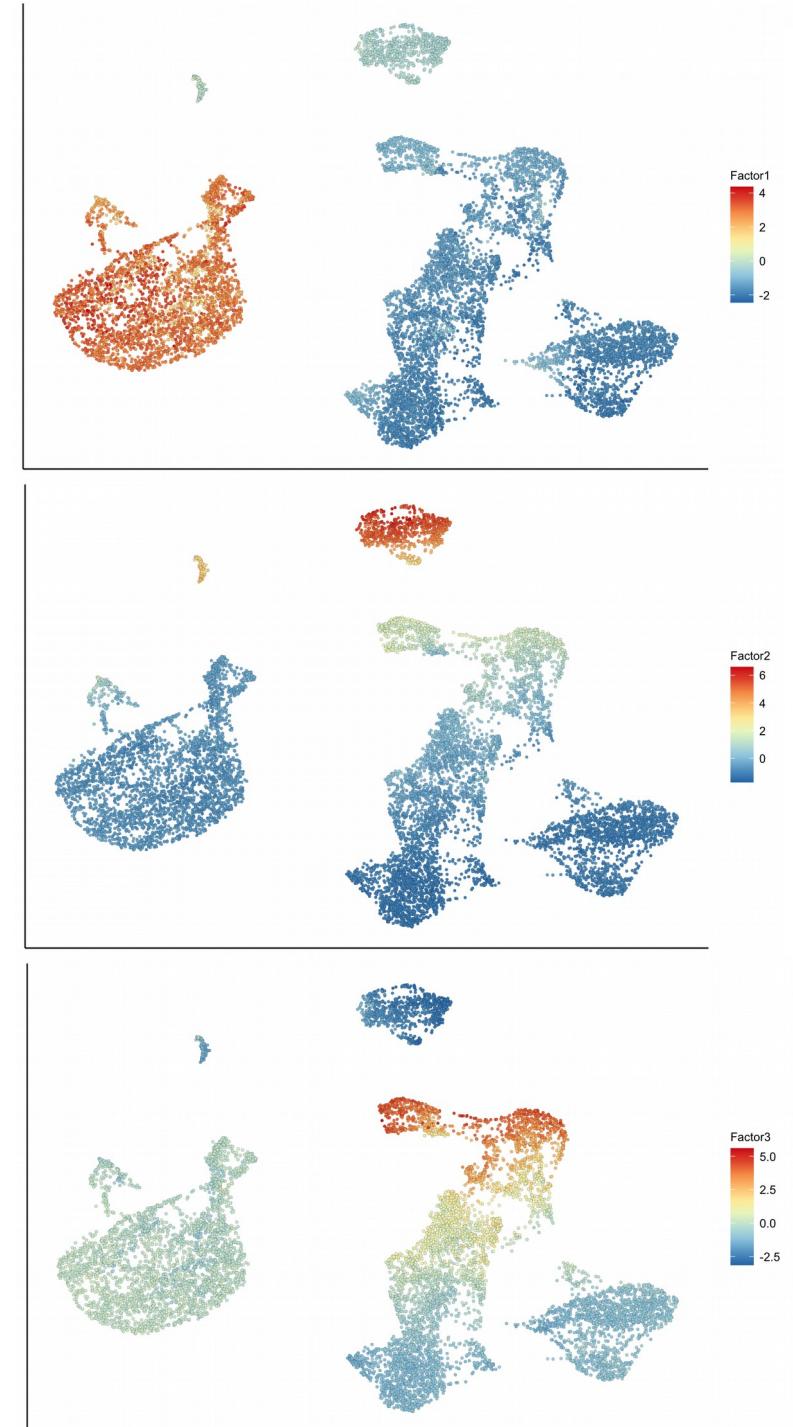
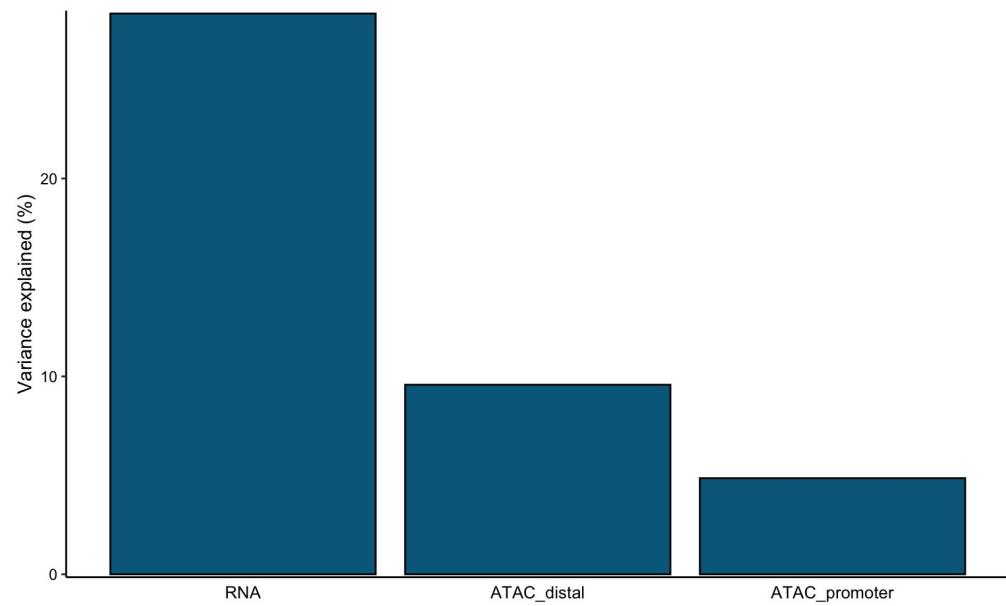


scATACseq

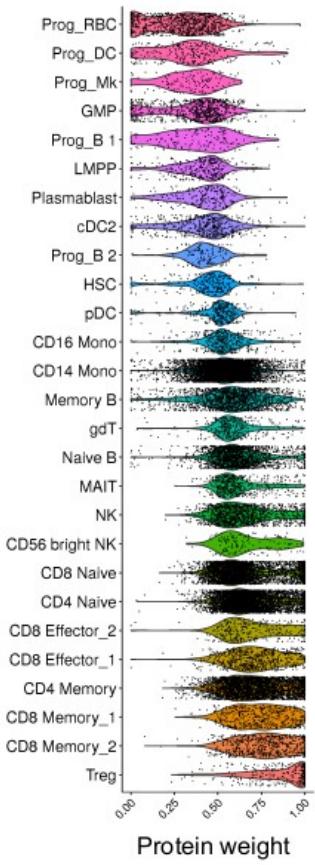


scRNAseq + scATACseq

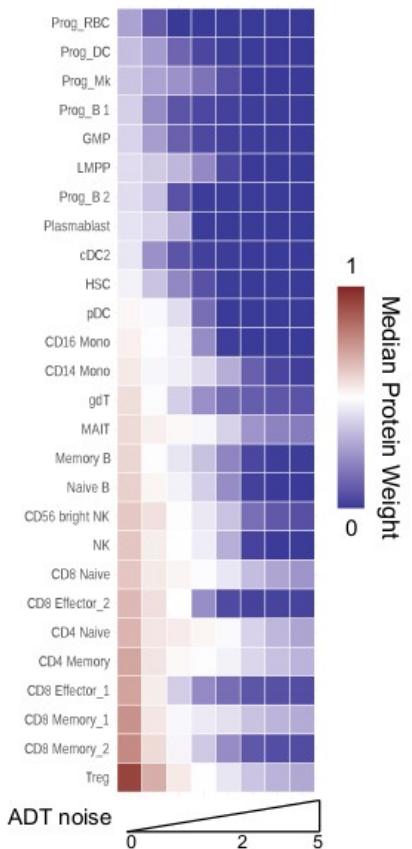




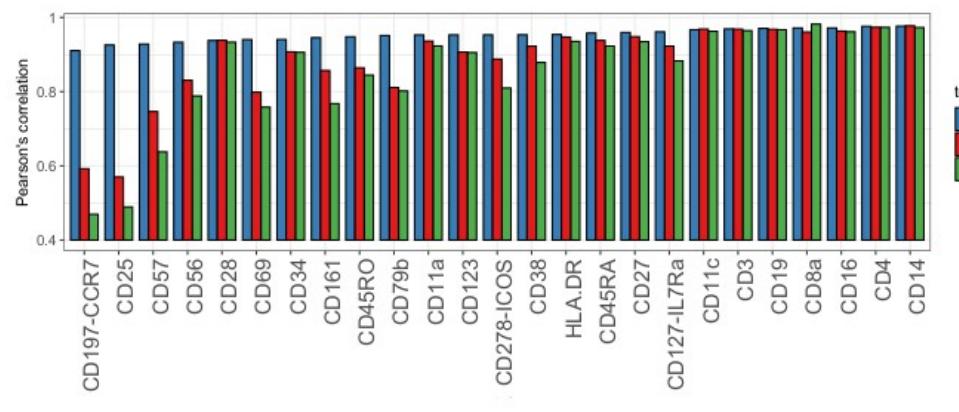
B



C



D



E

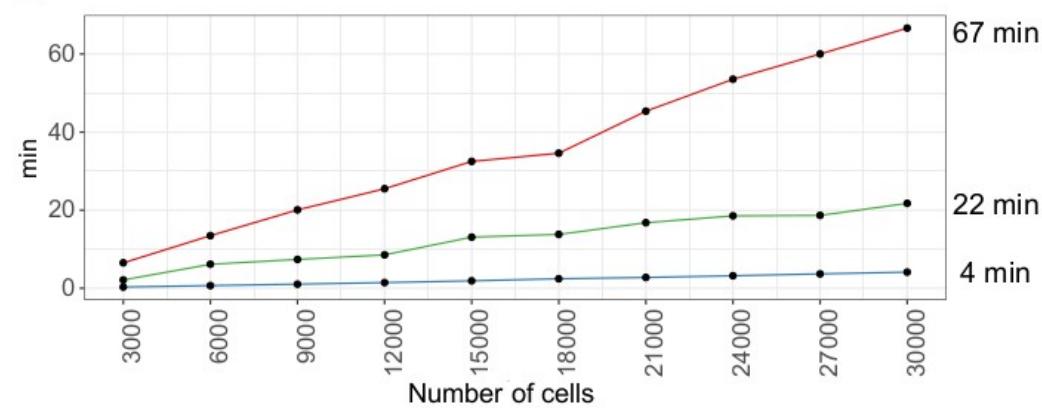


Figure 2: Benchmarking and robustness analysis for WNN integration.



National Bioinformatics
Infrastructure Sweden (NBIS)

SciLifeLab



*Knut och Alice
Wallenbergs
Stiftelse*



LUNDS
UNIVERSITET