Frederick National Laboratory
for Cancer Research


# Applications of Single Cell Sequencing: CITE-Seq

**Nathan Wong**

**CCBR**

**2019 October 10**

DEPARTMENT OF HEALTH AND HUMAN SERVICES • National Institutes of Health • National Cancer Institute
The Frederick National Laboratory is a Federally Funded Research and Development Center operated by Leidos Biomedical Research, Inc., for the National Cancer Institute

# What is CITE-Seq?



- "Cellular Indexing of Transcriptomes and Epitopes by Sequencing"
- Short definition: Including surface protein antibodies with single cell sequencing
- Purpose: Show that the antibody cell sorting often used prior to single cell sequencing is replicated after sequencing
- Need came up because the single cell transcript data does not always resemble the expected surface protein profile

# CITE-Seq, courtesy of the Satija lab

Frederick National Laboratory for Cancer Research

nature methods

Brief Communication | Published: 31 July 2017

# Simultaneous epitope and transcriptome measurement in single cells

Marlon Stoeckius ✉, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija & Peter Smibert

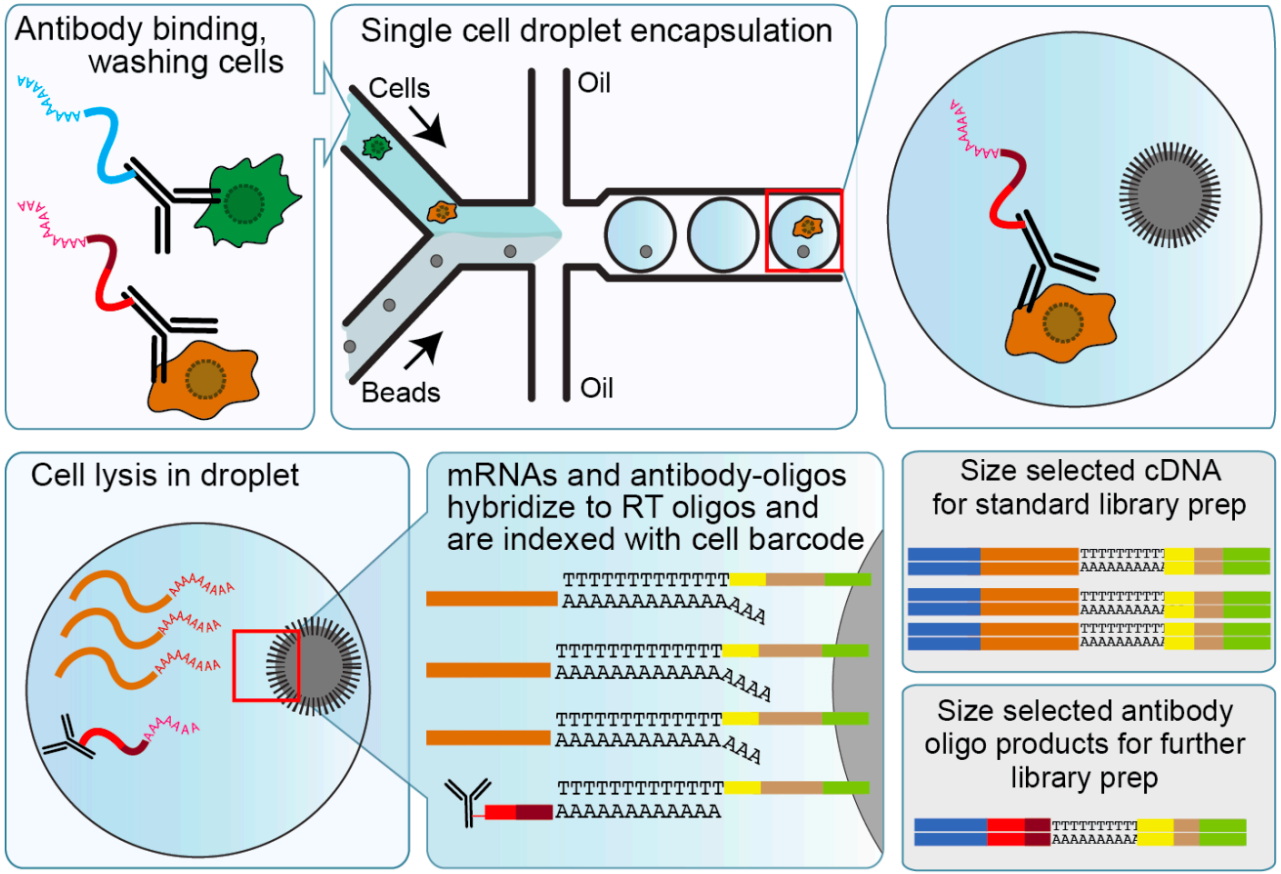
*Nature Methods* **14**, 865–868 (2017) | Download Citation ↓

# Methodology



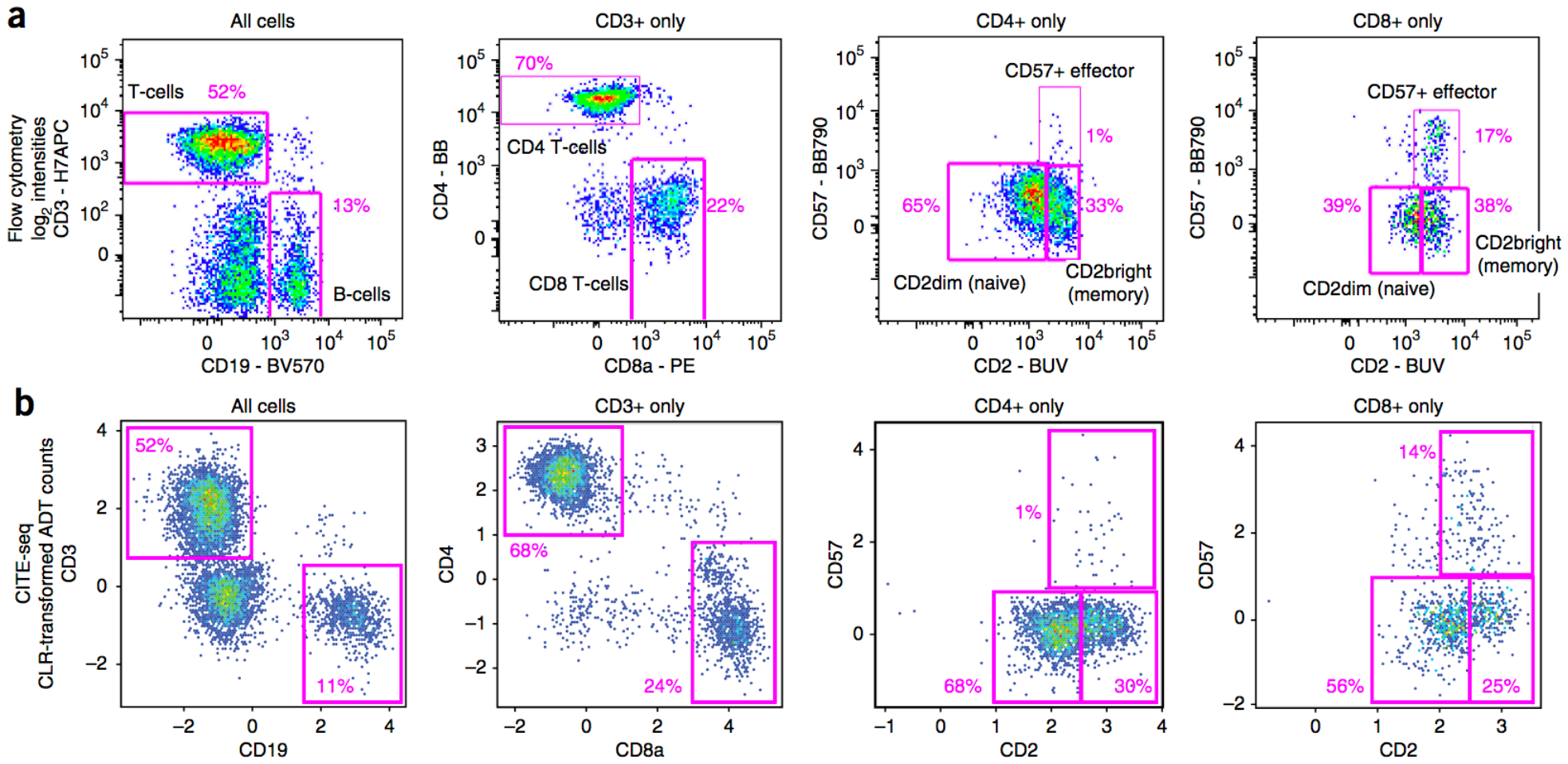
- **Specific antibodies with sequence tags are bound to cells prior to single cell encapsulation**


- **Each droplet contains both the cell and the bound antibody tag**
- **mRNA and antibody oligos hybridize to RT oligos and are indexed with cell barcode**
- **Standard scRNASeq is performed for mRNAs**
- **Antibody tags are isolated by size selection**



# CITE-Seq is comparable to initial flow cytometry

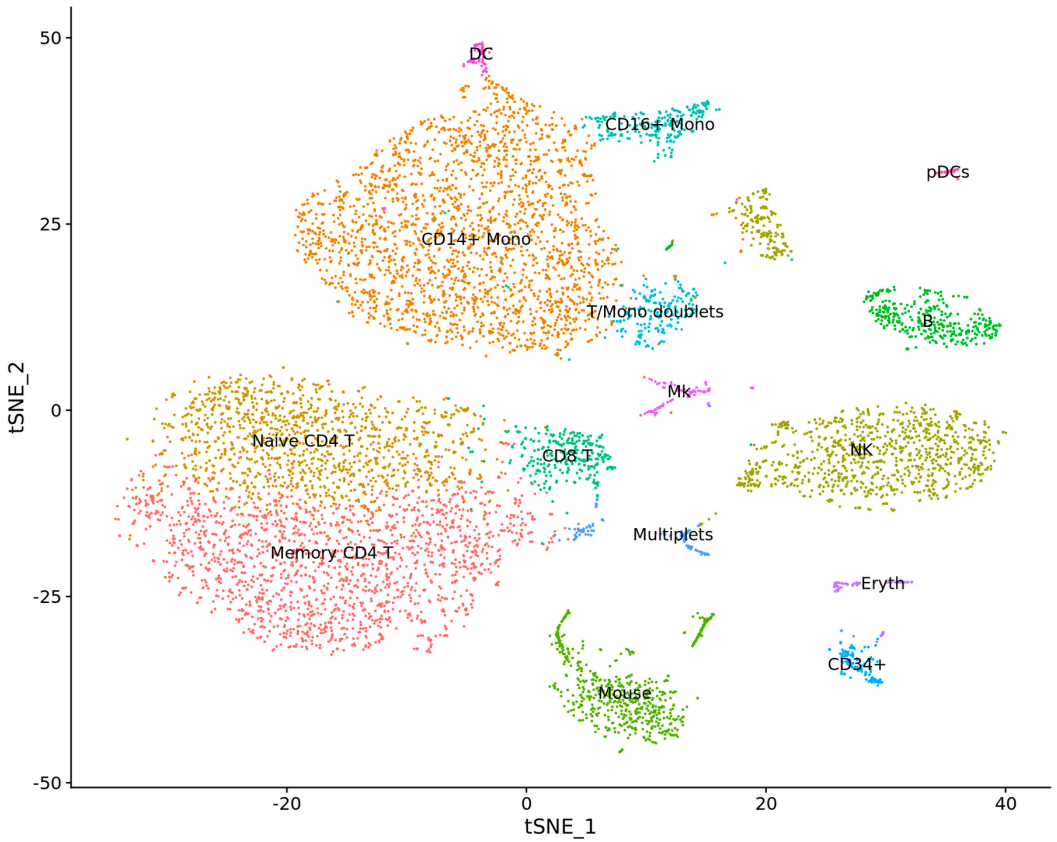
Frederick National Laboratory for Cancer Research



# Antibody-based identification can be projected onto RNA-Seq based clustering



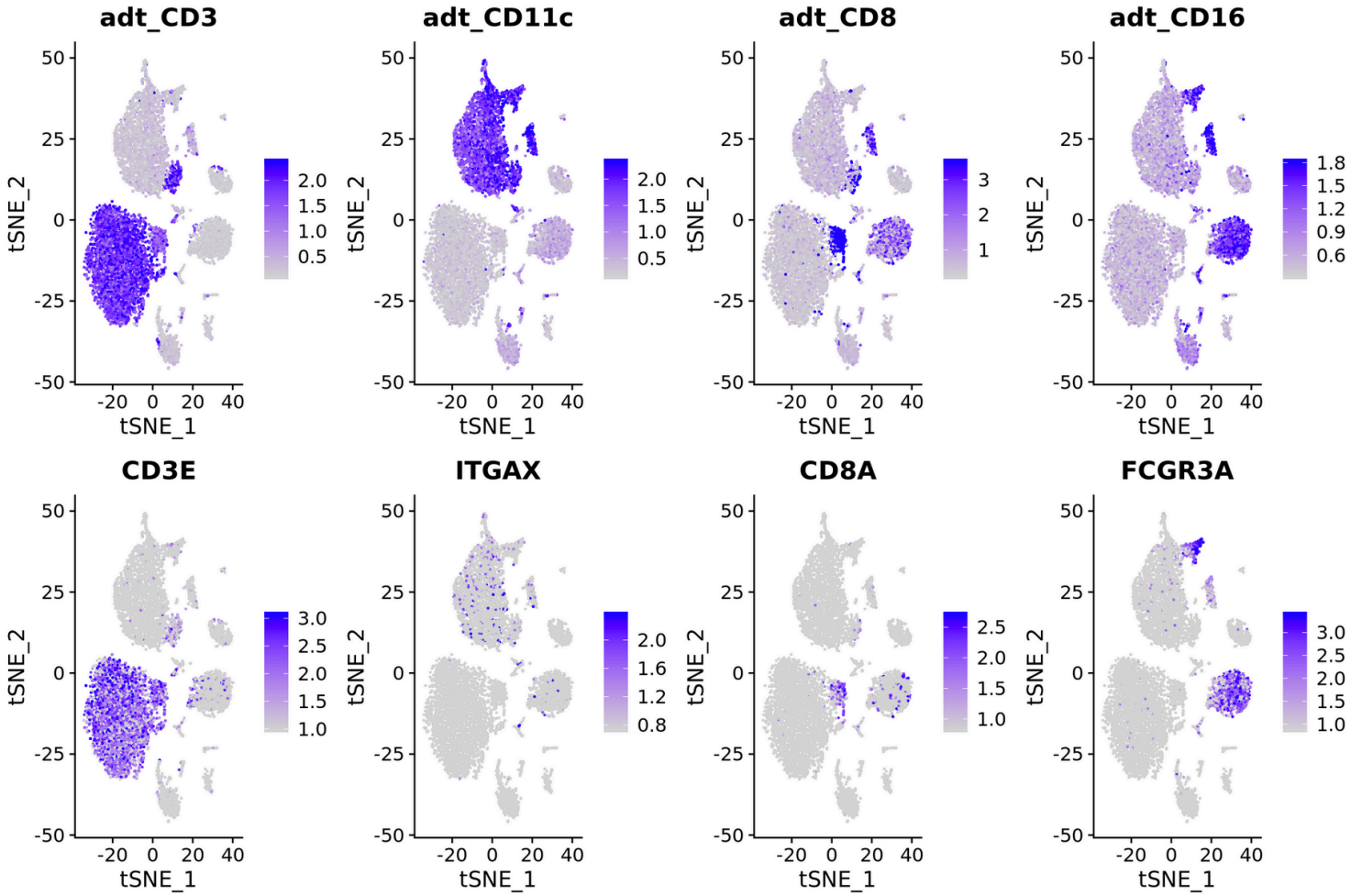
- **Initial reference image is processed using “standard” scRNA-Seq protocol**
- **Labels for this image were assigned by the Satija group based on prior knowledge of the dataset**
- **CITE-Seq protein markers for this dataset include:**
  - CD3, CD4, CD11c, CD14, CD16, CD19 CD34, CD45RA, and CD56



# Antibody-based identification can be projected onto RNA-Seq based clustering

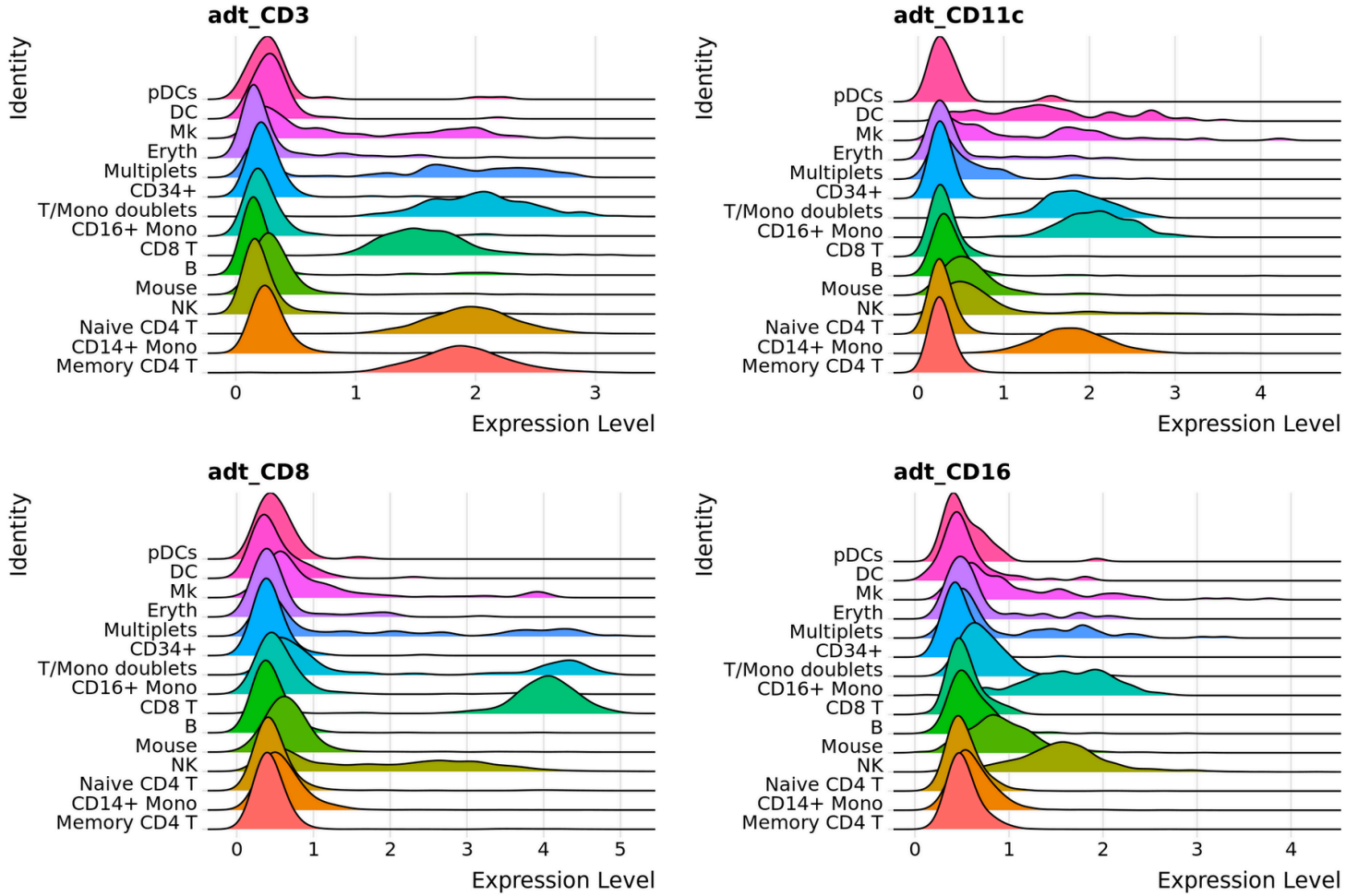


- Top: Protein markers
- Bottom: Associated gene IDs
- Using combined biological knowledge and protein expression allows for the confirmation of individual cell types in scRNA-Seq
- Cell surface markers can confirm the RNA-Seq-based cell typing



# Visualizing protein surface marker binding can confirm cell identities

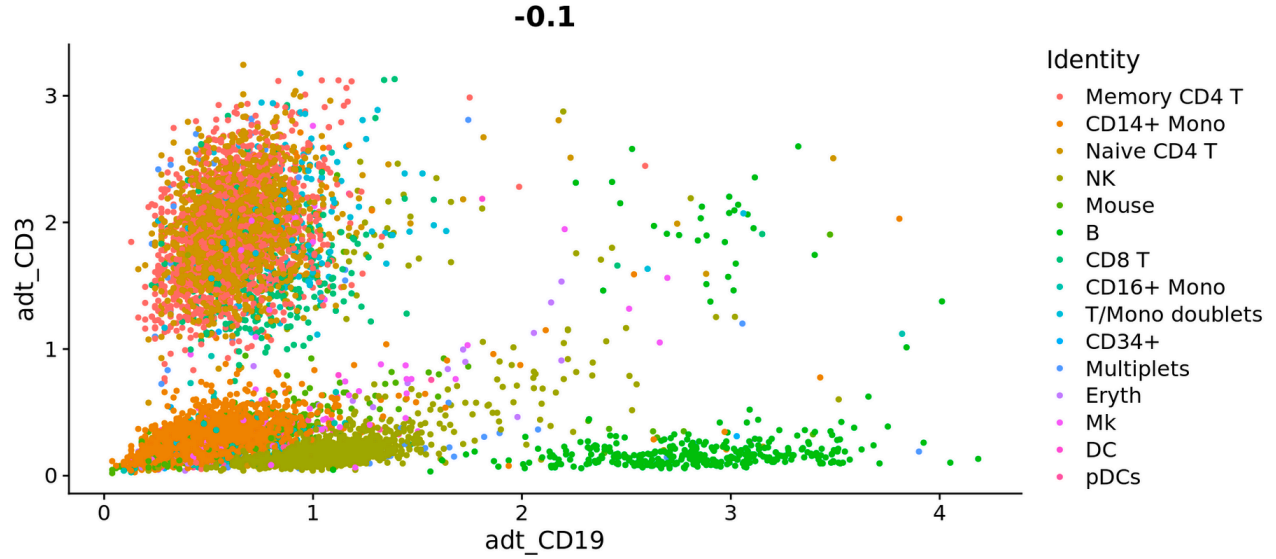
Frederick National Laboratory for Cancer Research



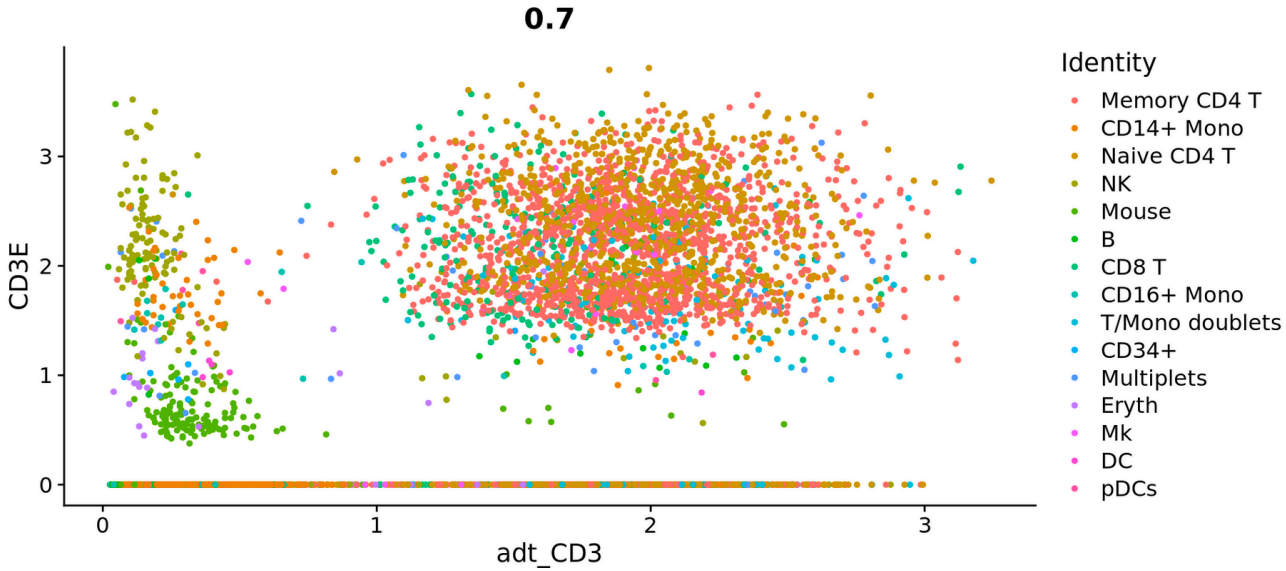
# Scatter plots and correlation analysis can be used to show “protein vs. protein” and “protein vs. RNA”

Frederick National Laboratory for Cancer Research

- **Interest in surface protein expression can show positive, negative, or null correlation**



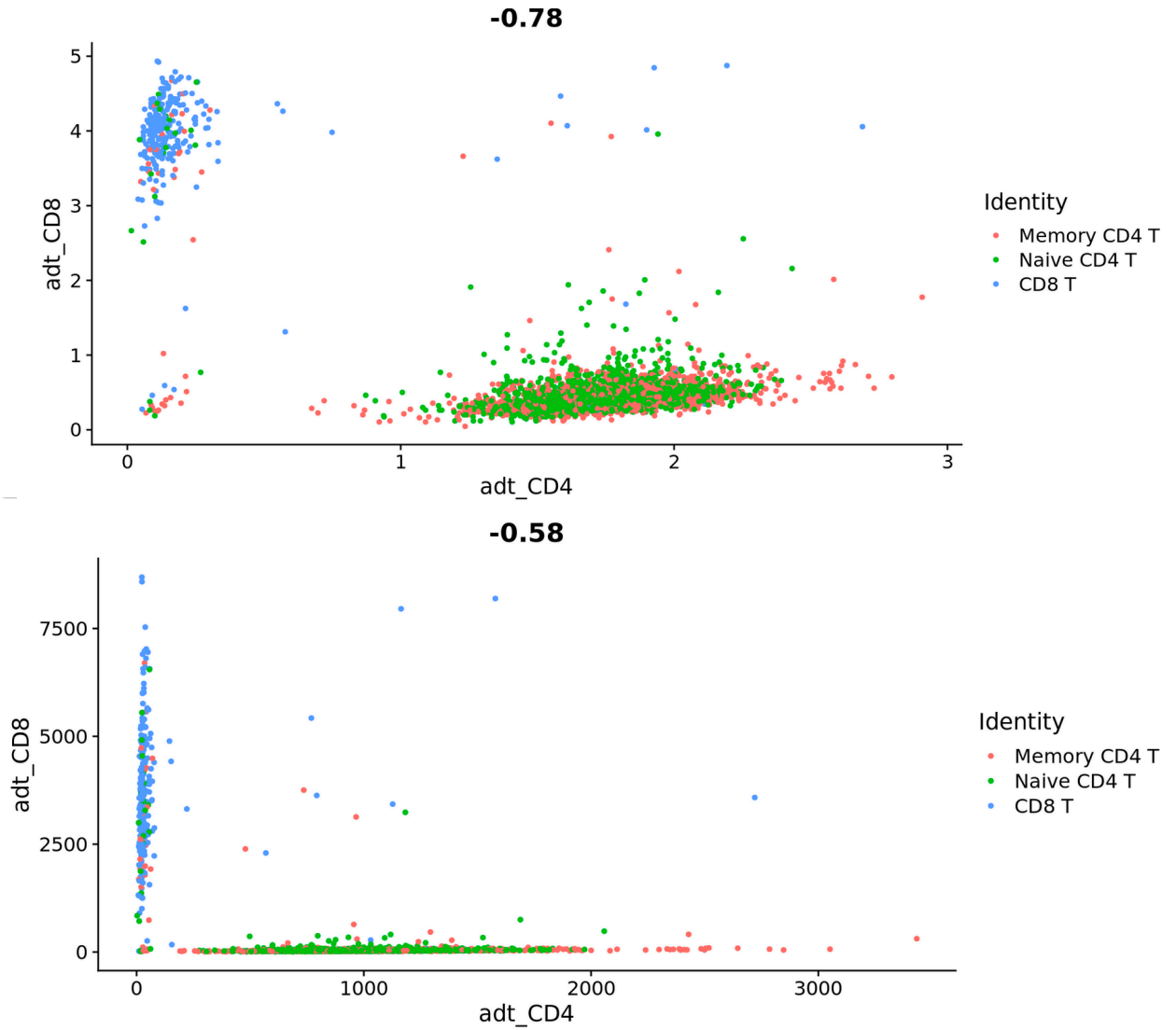

- **The previous image for surface protein vs. transcript expression can be indicated with scatter plots as well**
  - Note that the gene expression has more zero values than the protein marker



# Protein markers can improve cell typing

Frederick National Laboratory for Cancer Research

- CD4 and CD8 markers are projected for T-cells
- Some cells identified transcriptomically as CD4+ T-cells are being shown in the CD8+ protein marker population
- Biologically, this is not unexpected, because naïve CD4+ T-cells and CD8+ T-cells have similar transcriptomes
- This could have larger effects on downstream analysis without CITE-Seq reinforcement

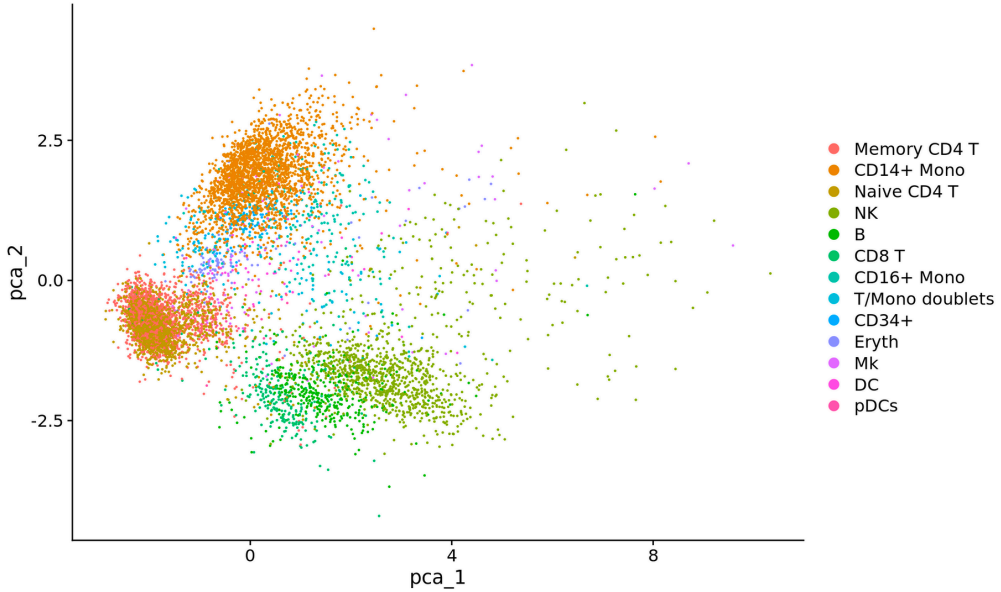


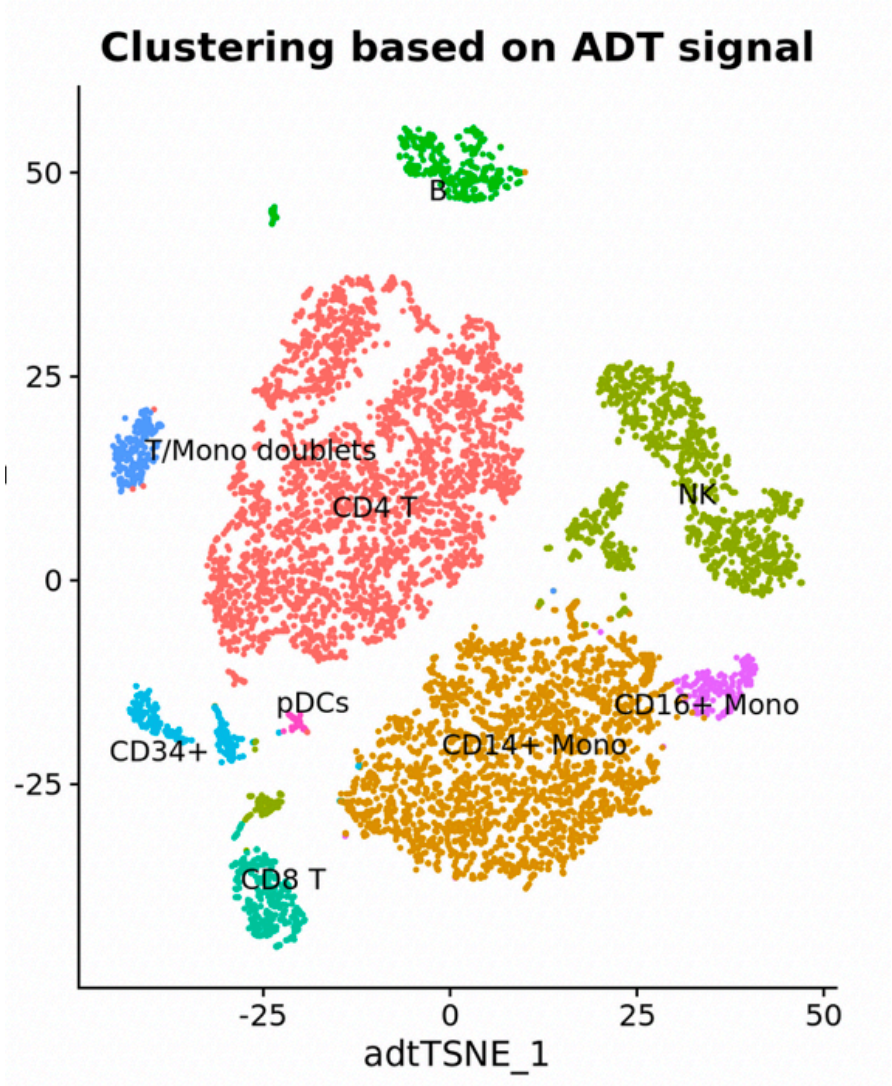
# Clustering can be performed directly on the protein expression profile



- **Many of the analyses performed based on scRNASeq (transcriptome) can be performed based on CITESeq**

- **The more markers that are used, the better the profiling/clustering that can be extracted**

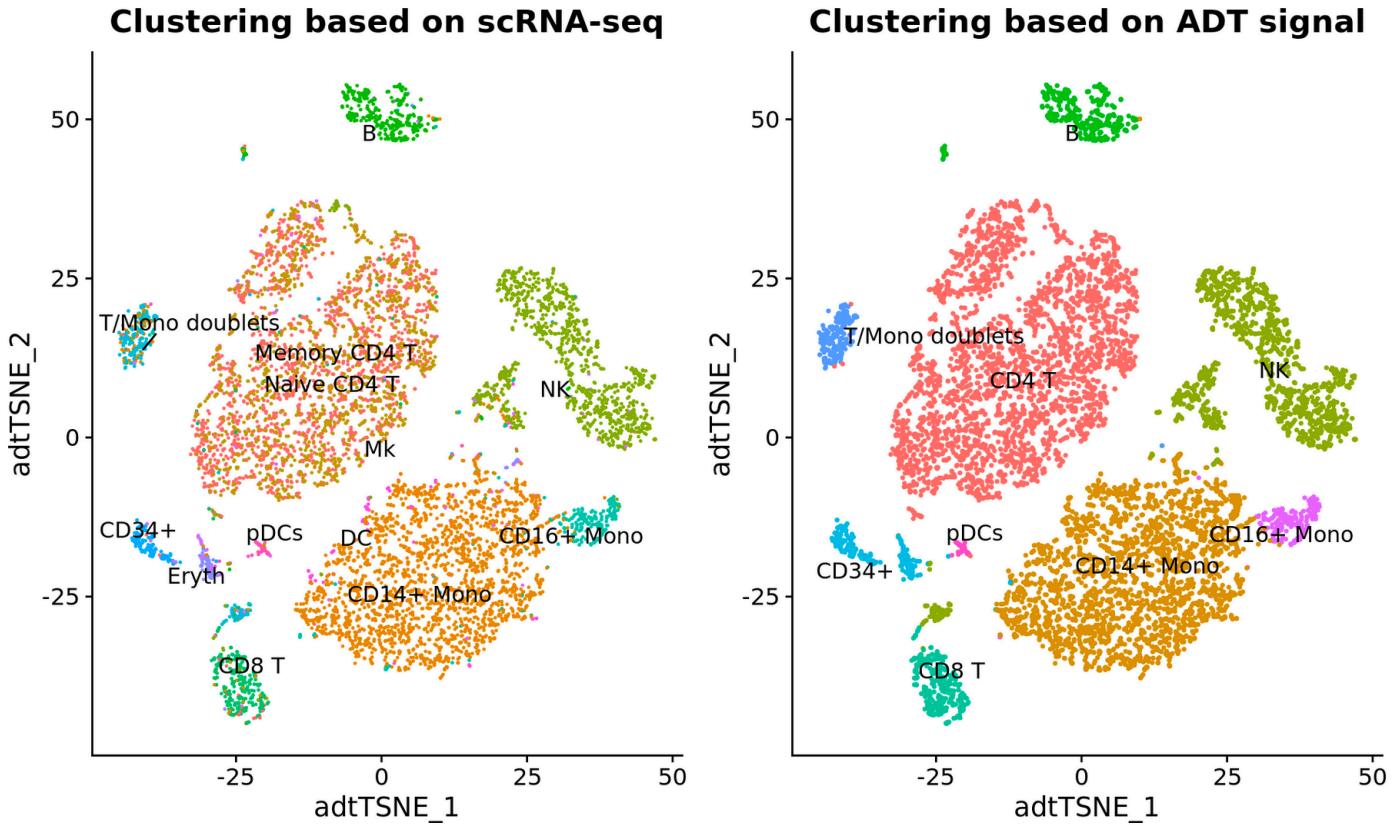


# Clustering between scRNA-Seq and CITE-Seq/protein markers generate similar results

Frederick National Laboratory for Cancer Research

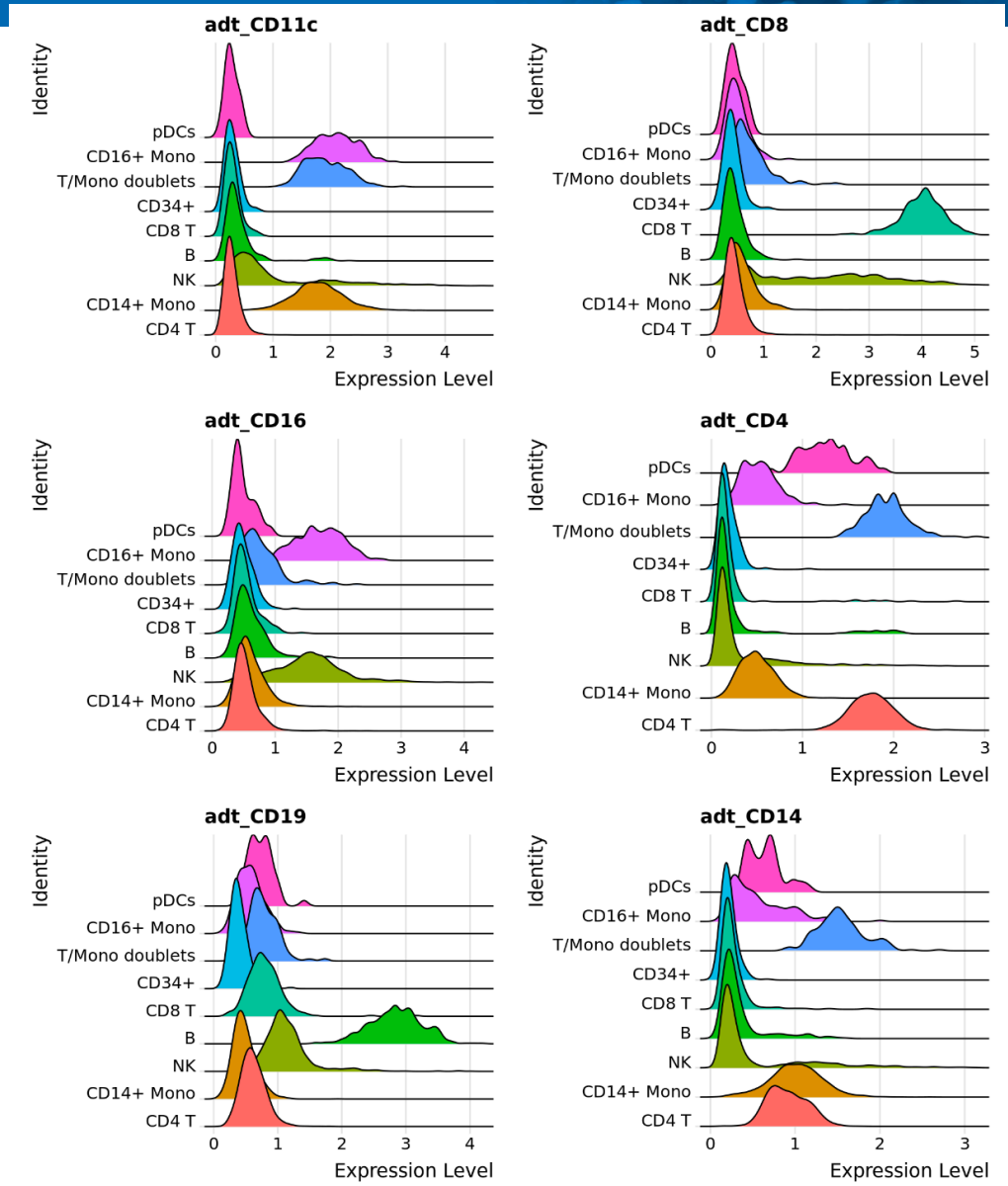
- Cluster positions in this image are based on the protein signal
- The labels on the left are based on the scRNA signal, the labels on the right based on protein signal
- Advantage: The CD4+ and CD8+ T-cells are more clearly defined on the protein signal
- Disadvantage: Some cell types with poor/absent markers (e.g. erythrocytes) are lost



# Surface marker profiles are representative of the cell types



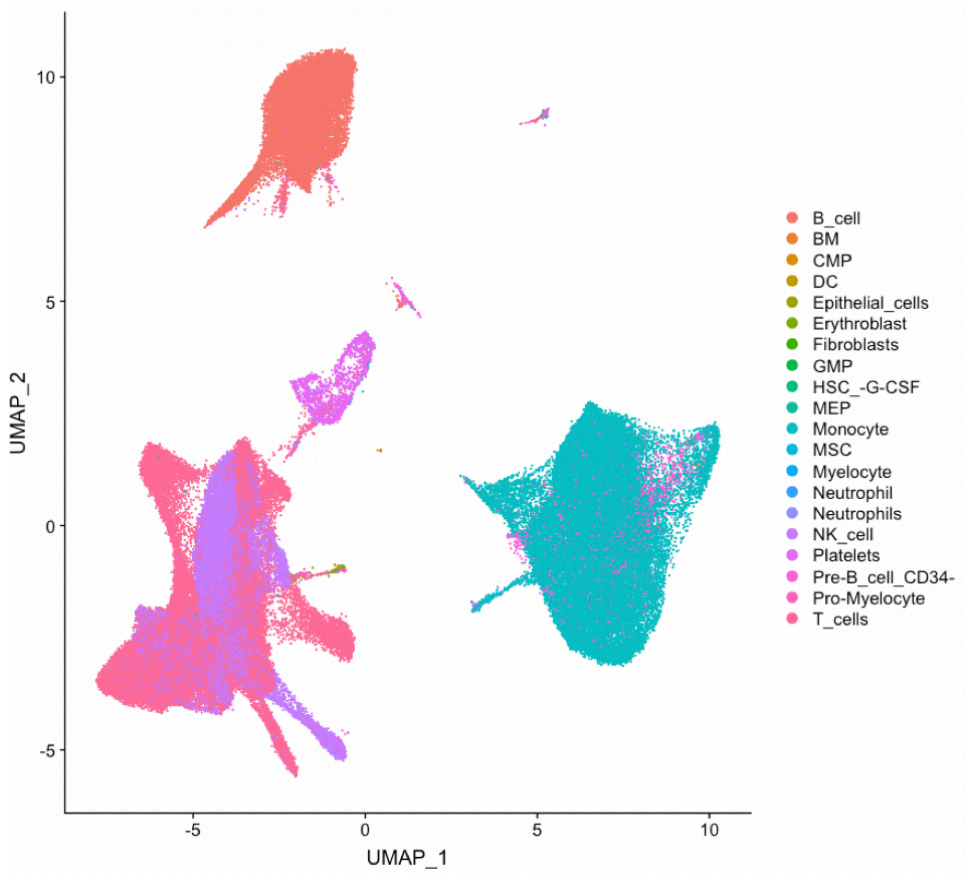
- The ridge plot can show the relative expression of the proteins for each cluster ID
- Based on these results, it would be similar to using FACS in determining cell typing



# “Live” application of CITE-Seq

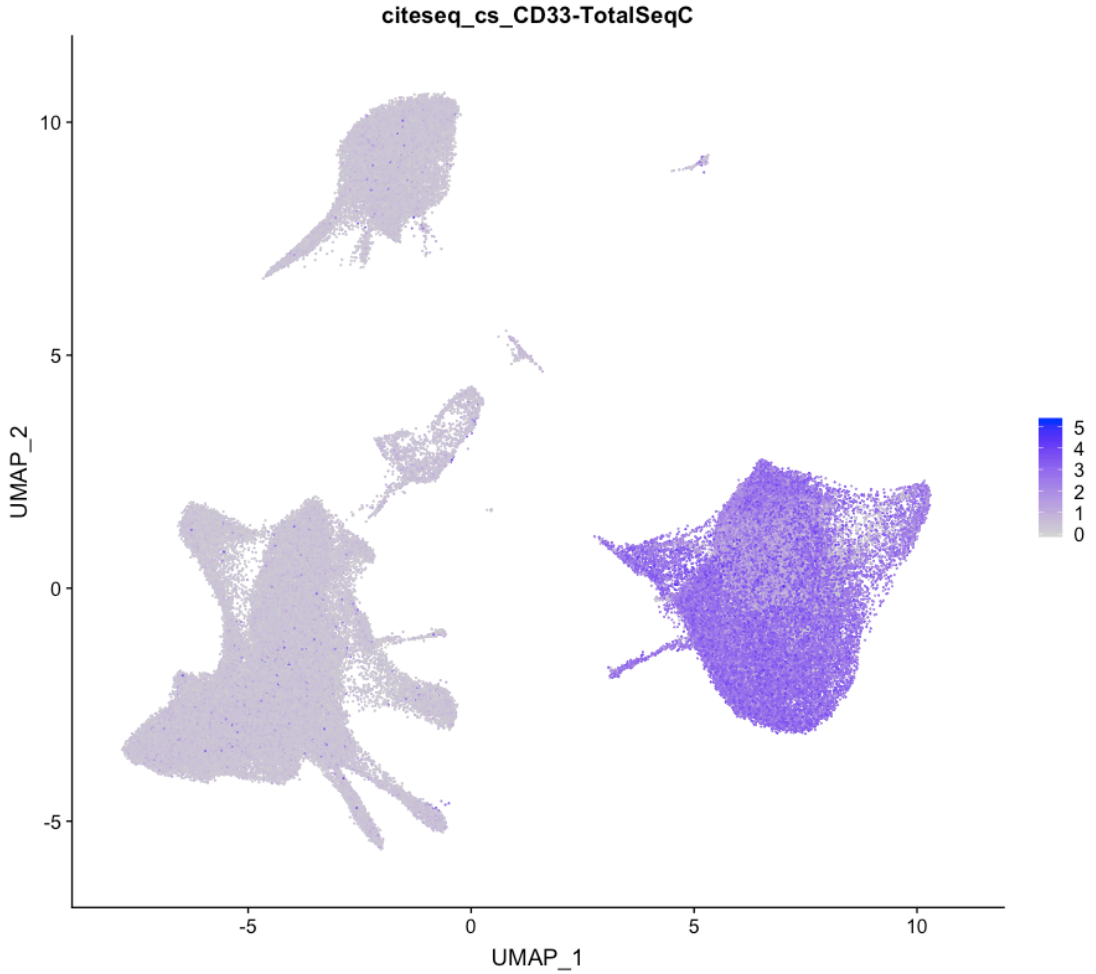


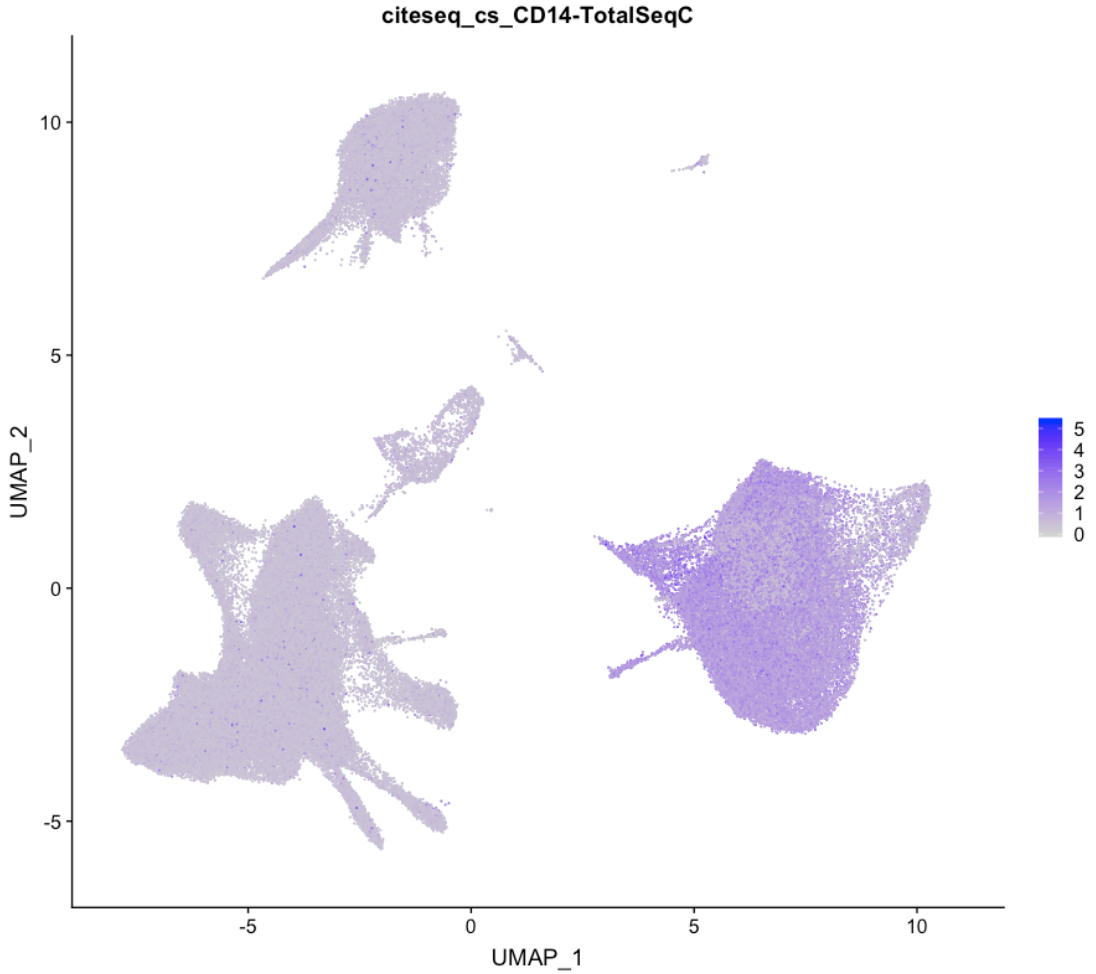
- **Short background: Human PBMCs from 11 different samples**
- **Interest in CD33 and CD14 as potential surface markers**
- **The initial transcript based cell-clustering and cell typing with SingleR showed the majority of cells as:**
  - B-cells (top)
  - T-cells and NK-cells (bottom left)
  - Monocytes (bottom right)



# CD33 and CD14 are strong correlative markers for monocytes

Frederick National Laboratory for Cancer Research

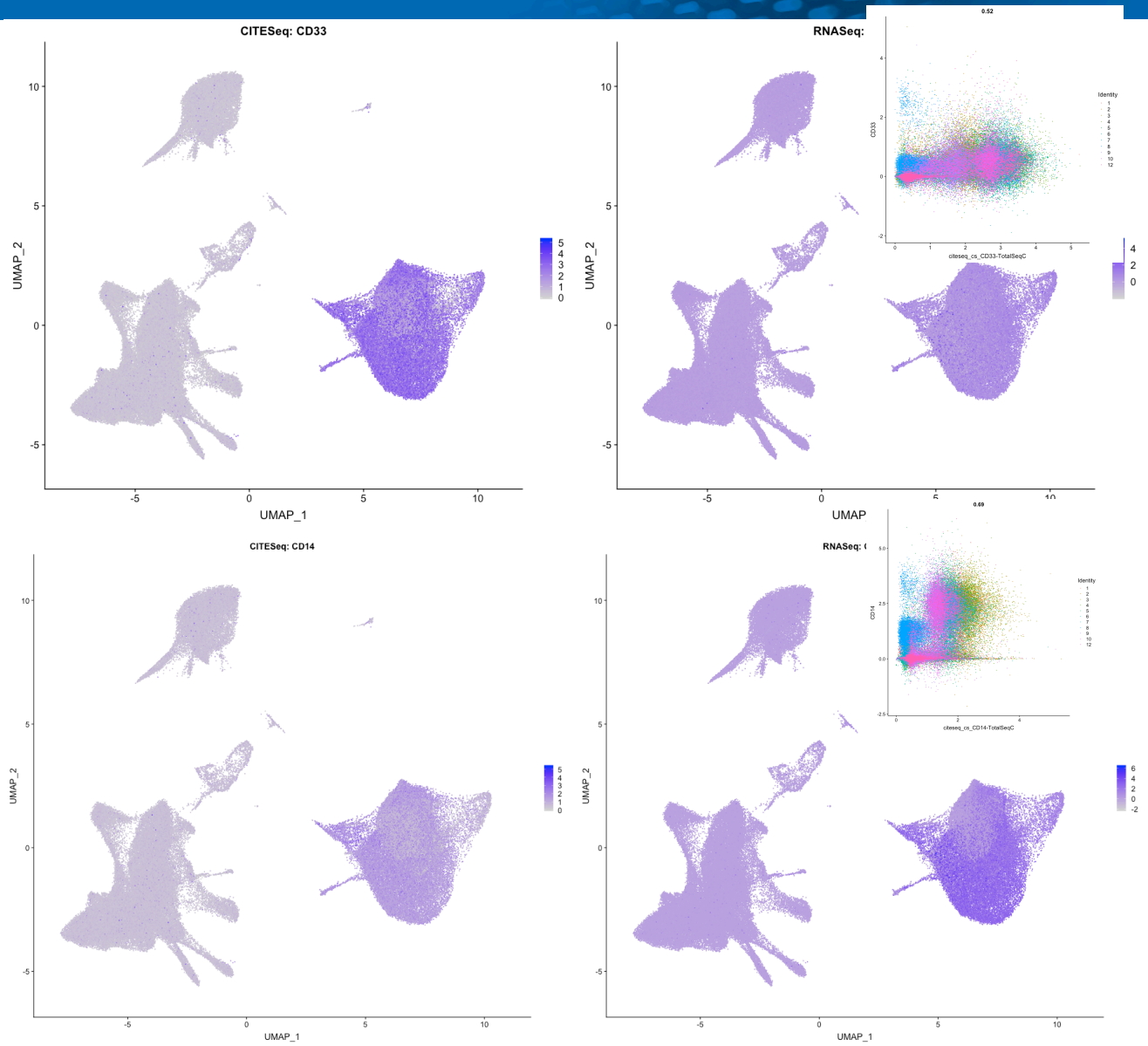


# CITESeq indicates a discrepancy between transcription and surface protein accessibility

Frederick National Laboratory for Cancer Research

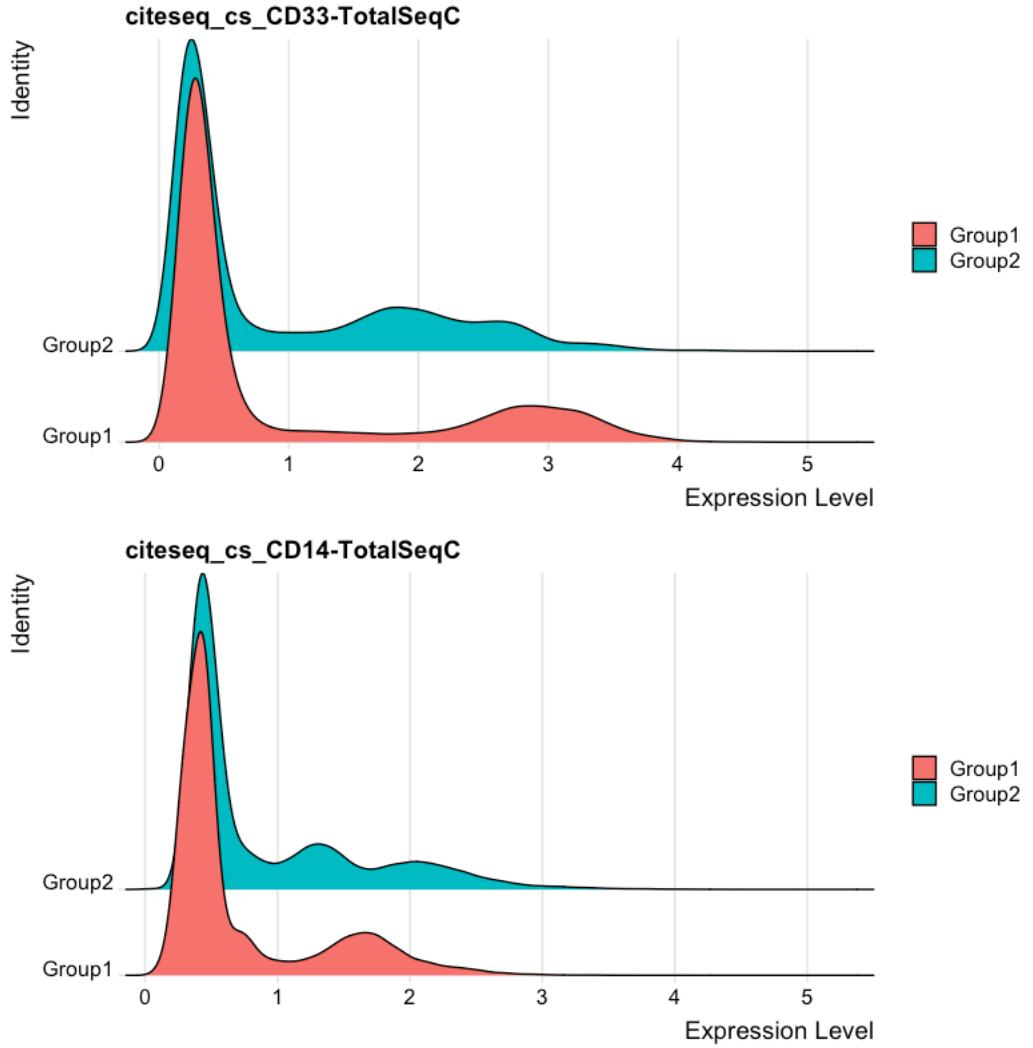
- Protein profiles (left) showed more noticeably “antibody responsive” markers in monocytes
- The difference is less pronounced at the transcription level (right), especially for CD33
- This is even more apparent when observing the scatter plots (insets)



# CITESeq for identifying sample groups



- **Samples were split into two categories.**
- **Group 1 showed more definite CD33 surface protein**
- **CD14 was confirmed as a less readily usable marker in defining the two groups**



# Take home message



- **CITE-Seq is an excellent tool for integrating traditional antibody cell sorting with scRNA-Seq**
- **This can help synchronize the FACS results with scRNA-Seq**
- **Word of warning: Understand the underlying biology**
- **Per the Hitchhiker's Guide to the Galaxy: Don't Panic**

# References


- Seurat Guided Analyses. Satija R, et al. https://satijalab.org/seurat/vignettes.html
- Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Swerdlow H, Satija R, Smibert P. 2017. Simultaneous measurement of epitopes and transcriptomes in single cells. *Nature Methods*.
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*.
- Stuart T, Satija R. 2019. Integrative single-cell analysis. *Nature Reviews Genetics.*
- Stuart T*, Butler A*, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell*.